ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ ΤΜΗΜΑ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ & ΠΑΡΑΓΩΓΗΣ

Γεννητικά Ανταγωνιστικά Δίκτυα και εφαρμογές τους

Διπλωματική Εργασία

Γεροντόπουλος Ιωάννης
Α.Μ. 222017011

Επιβλέπων Καθηγητής: Νικολάου Γρηγόριος
Λέκτορας ΠΑ.Δ.Α.

Αθήνα, 2023

Η παρούσα διπλωματική εργασία εγκρίθηκε ομόφωνα από την τριμελή εξεταστική επιτροπή, η οποία ορίστηκε από την Γ.Σ του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής του Πανεπιστήμιου Δυτικής Αττικής, σύμφωνα με τον νομό και τον εγκεκριμένο Οδηγό Σπουδών του τμήματος.

Επιβλέπων: Νικολάου Γρηγόριος

Λέκτορας

Επιτροπή Αξιολόγησης:

GRIGORIOS NIKOLAOU Digitally signed by GRIGORIOS NIKOLAOU Date: 2023.03.31 15:04:34 +03'00'

Νικολάου Γρηγόριος

Soultana Vasileiadou Digitally signed by Soultana Vasileiadou Date: 2023.04.04 21:32:58 +03'00'

Βασιλειάδου Σουλτάνα

Christos Drosos Digitally signed by Christos Drosos Date: 2023.03.31 16:41:17 +03'00'

Δρόσος Χρήστος

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Ιωάννης Γεροντόπουλος του Ηλία , με αριθμό μητρώου 222017011 φοιτητής Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Βιομηχανικής Σχεδίασης και Παραγωγής , δηλώνω υπεύθυνα ότι: «Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου»

Ο Δηλών

Yannis Geronotpoulos

Περίληψη

Τα τελευταία χρόνια, η Τεχνητή Νοημοσύνη (AI) έχει γίνει αντικείμενο έντονων αντιπαραθέσεων. Η μηχανική μάθηση, η βαθιά μάθηση και η τεχνητή νοημοσύνη εμφανίζονται σε αμέτρητα άρθρα, ακόμη και σε δημοσιεύσεις που δεν σχετίζονται με την τεχνολογία. Μας έχουν υποσχεθεί ένα μέλλον έξυπνων συσκευών, ένα μέλλον άλλοτε ζωγραφισμένο στην πραγματικότητα και άλλοτε ουτοπικό, όπου οι ανθρώπινες θέσεις εργασίας θα είναι σπάνιες και οι περισσότερες οικονομικές δραστηριότητες θα τις χειρίζονται ρομπότ ή πράκτορες τεχνητής νοημοσύνης. Για έναν μελλοντικό ή σημερινό επαγγελματία της μηχανικής μάθησης, είναι σημαντικό να μπορεί να αναγνωρίζει το ουσιώδες, μέσα από τις πρωτοποριακές εξελίξεις σε σχέση με την παραπληροφόρηση που δέχεται.

Ένα από αυτά τα σημαντικά επιτεύγματα είναι η ραγδαία ανάπτυξη στις τεχνικές αναπαραγωγής συνθετικών δεδομένων. Μια από αυτές τις τεχνικές η οποία ξεχωρίζει είναι τα Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks, GANs). Είναι μία οικογένεια μοντέλων, ικανών να παράγουν ρεαλιστικά, συνθετικά δεδομένα που δεν ξεχωρίζουν από την πραγματικότητα. Ένα GAN αποτελείται από δύο δίκτυα, με το ένα να προσπαθεί να παράξει όσο το δυνατόν πιο ρεαλιστικά δεδομένα (δημιουργός), και το άλλο να προσπαθεί να ξεχωρίσει όσο το δυνατόν καλύτερα τα τεχνητά από τα αληθινά (διευκρινιστής). Μέσω του συνεχούς ανταγωνισμού μεταξύ τους, τα δίκτυα βελτιώνονται μέχρι να επέλθει η ισορροπία, όπου τα αληθινά και τα τεχνητά δεδομένα φαίνονται πανομοιότυπα στον διευκρινιστή. Ο δημιουργός τότε έχει φτάσει στο βέλτιστο σημείο να παράγει αληθοφανή δεδομένα.

Τα GAN έχουν βρει εφαρμογή σε πολλά πεδία της τεχνολογίας και όχι μόνο. Παράλληλα, με έμπνευση το αρχικό μοντέλο του GAN, έχουν εφευρεθεί διάφορες χρήσιμες παραλλαγές οι οποίες χρησιμοποιούνται σε πολλές καινοτόμες εφαρμογές. Μερικές από τις πιο αξιοσημείωτες επιτυχίες τους, είναι η χρήση τους στην ιατρική, στην κυβερνοασφάλεια και στις τέχνες. Υπάρχουν πολλά που θα πρέπει να γίνουν, ώστε το μέλλον φαίνεται αρκετά ελπιδοφόρο.

Στο πλαίσιο της παρούσας εργασίας γίνεται μια εισαγωγή στη γεννητική μοντελοποίηση και στην άνοδο της. Στην συνέχεια, προσδιορίζονται οι όροι της τεχνητής νοημοσύνης, της μηχανικής μάθησης και της βαθιάς μάθησης, εστιάζοντας στην βαθιά μάθηση και τον τρόπο εκμάθησης των δεδομένων της. Ακολούθως, επισημαίνεται η λειτουργία των βαθιών νευρωνικών δικτύων μαζί με όλες τις μαθηματικές και αλγοριθμικές επεξηγήσεις. Τα παραπάνω κεφάλαια είναι σημαντικά για την σωστή κατανόηση των GANs, διότι χρησιμοποιούν όλα όσα αναφέρονται. Ύστερα, παρουσιάζονται τα VAEs τα οποία αποτέλεσαν κίνητρο για την δημιουργία των GANs. Στην συνέχεια, προσδιορίζονται λεπτομερώς τα GANs, μαζί με όλες τις δυσκολίες εκπαίδευσης τους. Τέλος, αναφέρονται οι καινοτόμες εφαρμογές και παραλλαγές, μαζί με τους κινδύνους και τις προοπτικές της χρήσης των GANs.

Λέξεις Κλειδιά: Γεννητικά Ανταγωνιστικά Δίκτυα (GANs), Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Βαθιά Μάθηση, Βαθιά Νευρωνικά Δίκτυα, Εκπαίδευση Δεδομένων

Abstract

In recent years, Artificial Intelligence (AI) has been the subject of intense publicity and debate. Machine learning, deep learning, and artificial intelligence appear in countless articles, even in non-tech publications. We have been promised a future of smart devices, a future sometimes painted in reality and sometimes utopian, where human jobs will be rare and most economic activities will be handled by robots or artificial intelligence agents. For a future or current machine learning professional, it is important to be able to identify the noise, through cutting-edge developments, in relation to the misinformation we receive.

One of these major achievements is the rapid development in synthetic data reproduction techniques. One of these techniques that stands out is Generative Adversarial Networks (GANs). It is a family of models, capable of producing realistic, synthetic data that is indistinguishable from reality. A GAN consists of two networks, one trying to produce as realistic data as possible (generator), and the other trying to distinguish artificial from real as best as possible (discriminator). Through constant competition with each other, the networks improve until equilibrium is reached, where real and artificial data appear identical to the discriminator. The generator has then reached the optimal point of producing plausible data.

GANs have found application in many fields of technology and beyond. At the same time, inspired by the original GAN model, various useful variants have been invented which are used in many innovative applications. Some of their most notable successes are their use in medicine, cyber security, and the arts. There is still much progress to be made, but the future looks quite promising.

In the context of this work, an introduction to genetic modeling and its rise is made. Next, the terms artificial intelligence, machine learning, and deep learning are defined, focusing on deep learning and how it learns from data. Next, the operation of deep neural networks is highlighted along with all the mathematical and algorithmic explanations. The above chapters are important for a proper understanding of GANs, because they use everything mentioned. Then, VAEs are presented which motivated the creation of GANs. Next, GANs are specified in detail, along with all their training difficulties. Finally, the innovative applications and variants are mentioned, along with the risks and prospects of using GANs.

Key Words: Generative Adversarial Networks (GANs), Artificial Intelligence, Machine Learning, Deep Learning, Deep Neural Networks, Data Training

Περιεχόμενα

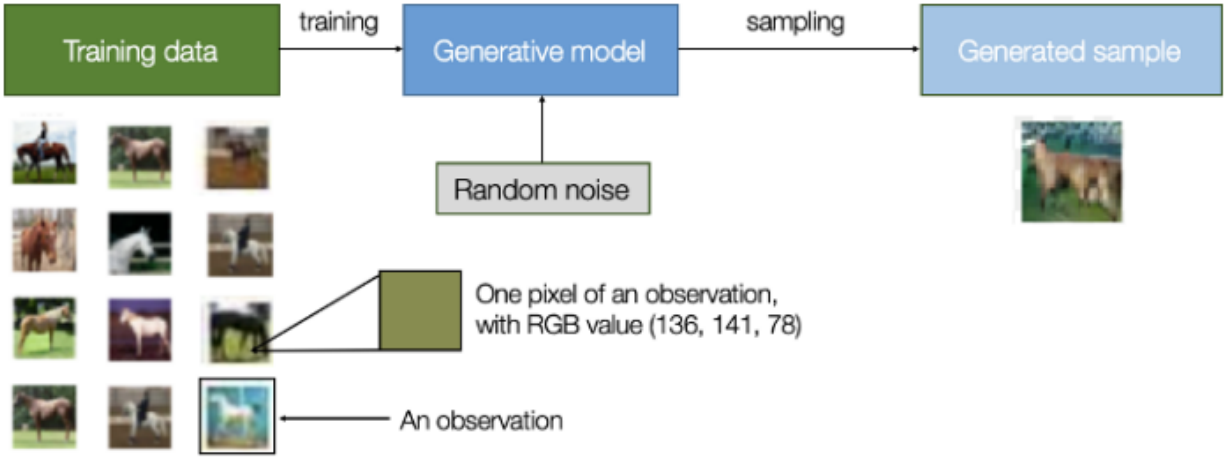
1. Εισαγωγή

Κανείς δεν μπορεί να αμφισβητήσει την επιρροή της μηχανικής μάθησης, και γενικότερα της τεχνητής νοημοσύνης στις ζωές μας τα τελευταία χρόνια. Μια από τις τεχνικές που ξεχωρίζει είναι η γεννητική μοντελοποίηση και ειδικότερα τα GAN. Για να φτάσουμε όμως στο σημείο να καταλάβουμε τα GAN και τις λειτουργίες του, χρειάζεται πρώτα να γίνει μια εισήγηση στο πως φθάσαμε σε αυτό το σημείο και ποιες τεχνολογίες χρησιμοποιούνται για την δημιουργία των μοντέλων GAN.

1.1 Τι είναι η Γεννητική Μοντελοποίηση;

Ένα γεννητικό μοντέλο μπορεί να οριστεί ευρέως ως εξής: Ένα γεννητικό μοντέλο περιγράφει πώς δημιουργείται ένα σύνολο δεδομένων (dataset), από την άποψη ενός πιθανοτικού μοντέλου (probabilistic model). Με δειγματοληψία από αυτό το μοντέλο, είμαστε σε θέση να δημιουργήσουμε νέα δεδομένα.

Ας υποθέσουμε ότι έχουμε ένα σύνολο δεδομένων που περιέχει εικόνες αλόγων. Μπορεί να θέλουμε να φτιάξουμε ένα μοντέλο που μπορεί να δημιουργήσει μια νέα εικόνα ενός αλόγου που δεν υπήρξε ποτέ αλλά εξακολουθεί να φαίνεται αληθινό, επειδή το μοντέλο έχει μάθει τους γενικούς κανόνες που διέπουν την εμφάνιση ενός αλόγου, μπορεί να το καταφέρει αυτό. Αυτό είναι το είδος του προβλήματος που μπορεί να λυθεί χρησιμοποιώντας γεννητική μοντελοποίηση.



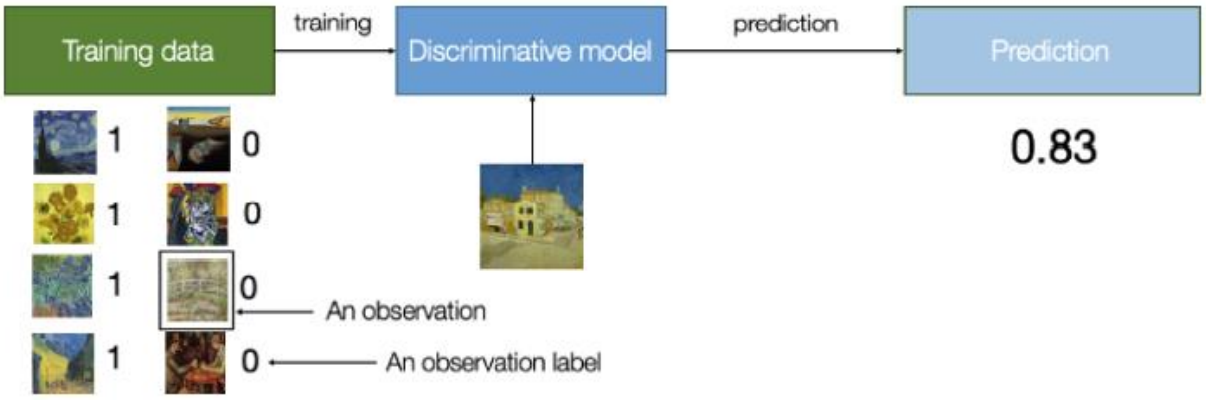
Σχήμα 1-1. Μια σύνοψη μιας τυπικής διαδικασίας γεννητικής μοντελοποίησης
(Πηγή: Generative Deep Learning, Teaching Machines to Paint, Write, Compose and Play, 2019)

Αρχικά, χρειαζόμαστε ένα σύνολο δεδομένων (dataset) που αποτελείται από πολλά παραδείγματα της οντότητας που προσπαθούμε να δημιουργήσουμε. Ο όρος αυτός είναι γνωστός ως δεδομένα εκπαίδευσης (training data) και ένα συγκεκριμένο σημείο δεδομένων ονομάζεται παρατήρηση (observation). Κάθε παρατήρηση αποτελείται από πολλά χαρακτηριστικά, για ένα πρόβλημα δημιουργίας εικόνας για παράδειγμα, τα χαρακτηριστικά είναι συνήθως οι μεμονωμένες τιμές pixel. Στόχος μας είναι να δημιουργήσουμε ένα μοντέλο που μπορεί να δημιουργήσει νέα σύνολα χαρακτηριστικών που φαίνονται σαν να έχουν δημιουργηθεί χρησιμοποιώντας τους ίδιους κανόνες με τα αρχικά δεδομένα. Εννοιολογικά, για τη δημιουργία εικόνων, αυτό είναι ένα απίστευτα δύσκολο έργο, λαμβάνοντας υπόψη τον τεράστιο αριθμό τρόπων με τους οποίους μπορούν να εκχωρηθούν οι μεμονωμένες τιμές

pixel και τον σχετικά μικρό αριθμό τέτοιων διατάξεων που αποτελούν μια ρεαλιστική εικόνα της οντότητας που προσπαθούμε να προσομοιώσουμε. Ένα γεννητικό μοντέλο πρέπει επίσης να είναι πιθανολογικό[1] παρά ντετερμινιστικό[2]. Εάν το μοντέλο μας είναι απλώς ένας σταθερός υπολογισμός, όπως η λήψη της μέσης τιμής κάθε pixel από το σύνολο δεδομένων, δεν είναι γεννητικό διότι το μοντέλο παράγει το ίδιο αποτέλεσμα κάθε φορά. Το μοντέλο πρέπει να περιλαμβάνει ένα στοχαστικό (τυχαίο) στοιχείο που επηρεάζει τα μεμονωμένα δείγματα που παράγονται από το μοντέλο. Είναι δουλειά μας να δημιουργήσουμε ένα μοντέλο που να μιμείται αυτήν την κατανομή όσο το δυνατόν πιο κοντά και στη συνέχεια να κάνουμε δειγματοληψία από αυτήν για να δημιουργήσουμε νέες, διακριτές παρατηρήσεις που φαίνονται σαν να μπορούσαν να είχαν συμπεριληφθεί στο αρχικό σετ δεδομένων εκπαίδευσης.

1.2 Γεννητικής έναντι Διακριτικής Μοντελοποίησης

Προκειμένου να κατανοήσουμε πραγματικά τι στοχεύει να επιτύχει η γεννητική μοντελοποίηση και γιατί αυτό είναι σημαντικό, είναι χρήσιμο να το συγκρίνουμε με το αντίστοιχο της, την διακριτική μοντελοποίηση. Στον τομέα της μηχανικής μάθησης, τα περισσότερα προβλήματα που θα αντιμετωπίσουμε πιθανότατα θα είναι διακριτικής φύσης. Η διακριτική μοντελοποίηση είναι συνώνυμη με την εποπτευόμενη μάθηση (supervised learning) ή με άλλα λόγια, η εκμάθηση μιας συνάρτησης που αντιστοιχίζει μια είσοδο σε μια έξοδο χρησιμοποιώντας ένα επισημασμένο σύνολο δεδομένων (labeled dataset).



Σχήμα 1-2. Παράδειγμα διακριτικής μοντελοποίησης
(Πηγή: Generative Deep Learning, Teaching Machines to Paint, Write, Compose and Play, 2019)

Η γεννητική μοντελοποίηση εκτελείται συνήθως με ένα σύνολο δεδομένων χωρίς επισήμανση (unlabeled dataset, δηλαδή, ως μια μορφή μάθησης χωρίς επίβλεψη, unsupervised learning), αν και μπορεί επίσης να εφαρμοστεί σε ένα επισημασμένο σύνολο δεδομένων, το οποίο θα αναφέρουμε αργότερα.

1 Πιθανολογικό - Στοχαστικό μοντέλο: Το φαινόμενο μπορεί να προσεγγιστεί μέσω κάποιας στατιστικής διαδικασίας.
2 Ντετερμινιστικό μοντέλο: Το φαινόμενο μπορεί να προσεγγιστεί μέσω αναλυτικής συνάρτησης, η οποία δεν περιέχει καθόλου τυχαιότητα.

Με άλλα λόγια, η διακριτική μοντελοποίηση επιχειρεί να εκτιμήσει την πιθανότητα μια παρατήρηση x να ανήκει στην κατηγορία y. Η γεννητική μοντελοποίηση δεν ενδιαφέρεται για την επισήμανση παρατηρήσεων. Αντίθετα, προσπαθεί να εκτιμήσει την πιθανότητα χωρίς καθόλου να δει την παρατήρηση. Το βασικό σημείο είναι ότι ακόμα κι αν μπορούσαμε να φτιάξουμε ένα τέλειο διακριτό μοντέλο για να αναγνωρίσουμε τους πίνακες του Βαν Γκογκ, δεν θα είχαμε καθόλου ιδέα πώς να δημιουργήσουμε έναν πίνακα που μοιάζει με του Βαν Γκογκ. Μπορεί να εξάγει μόνο πιθανότητες έναντι υπαρχουσών εικόνων, καθώς αυτό έχει εκπαιδευτεί να κάνει. Αντίθετα, θα χρειαστεί να εκπαιδεύσουμε ένα γεννητικό μοντέλο, το οποίο μπορεί να παράγει σύνολα pixel που έχουν μεγάλες πιθανότητες να ανήκουν στο αρχικό σύνολο δεδομένων εκπαίδευσης.

1.3 Η Άνοδος της Γεννητικής Μοντελοποίησης

Υπάρχουν τρεις βαθύτεροι λόγοι για τους οποίους η γεννητική μοντελοποίηση μπορεί να θεωρηθεί το κλειδί για το ξεκλείδωμα μιας πολύ πιο εξελιγμένης μορφής τεχνητής νοημοσύνης, που ξεπερνά αυτό που μπορεί να επιτύχει μόνη της η διακριτική μοντελοποίηση.

Σχήμα 1-3. Η πρόοδος των μοντέλων γεννητικής μοντελοποίησης
(Πηγή: Generative Deep Learning, Teaching Machines to Paint, Write, Compose and Play, 2019)

Πρώτον, καθαρά από πνευματική σκοπιά, δεν θα πρέπει να αρκεστούμε στο να μπορούμε να διαπρέπουμε μόνο στην κατηγοριοποίηση των δεδομένων, αλλά θα πρέπει επίσης να επιδιώκουμε μια πληρέστερη κατανόηση του τρόπου με τον οποίο δημιουργήθηκαν τα δεδομένα εξαρχής. Αυτό είναι αναμφίβολα ένα πιο δύσκολο πρόβλημα στην επίλυση, λόγω της μεγάλης διάστασης του χώρου των εφικτών εξόδων και του σχετικά μικρού αριθμού εισόδων που ανήκουν στο σύνολο δεδομένων. Ωστόσο, όπως θα δούμε, πολλές από τις ίδιες τεχνικές που οδήγησαν την ανάπτυξη στη διακριτική μοντελοποίηση, όπως η βαθιά μάθηση, μπορούν να χρησιμοποιηθούν και από τα γεννητικά μοντέλα.

Δεύτερον, είναι πολύ πιθανό ότι η γεννητική μοντελοποίηση θα είναι σημαντική για την καθοδήγηση μελλοντικών εξελίξεων σε άλλους τομείς της μηχανικής μάθησης, όπως η ενισχυτική μάθηση (η μελέτη των πρακτόρων διδασκαλίας, agents) για τη βελτιστοποίηση ενός στόχου σε ένα περιβάλλον μέσω δοκιμής και λάθους. Για παράδειγμα, θα μπορούσαμε να χρησιμοποιήσουμε την ενισχυτική μάθηση (reinforcement learning) για να εκπαιδεύσουμε ένα ρομπότ να περπατά σε ένα δεδομένο έδαφος. Η γενική προσέγγιση θα ήταν η κατασκευή μιας προσομοίωσης του εδάφους σε υπολογιστή και στη συνέχεια η εκτέλεση πολλών πειραμάτων όπου ο πράκτορας δοκιμάζει διαφορετικές στρατηγικές. Με την πάροδο του χρόνου ο πράκτορας θα μάθαινε ποιες στρατηγικές είναι πιο επιτυχημένες από άλλες και επομένως σταδιακά θα βελτιωνόταν. Ένα τυπικό πρόβλημα με αυτήν την προσέγγιση είναι ότι η φυσική του περιβάλλοντος είναι συχνά πολύ πιο περίπλοκη και θα πρέπει να

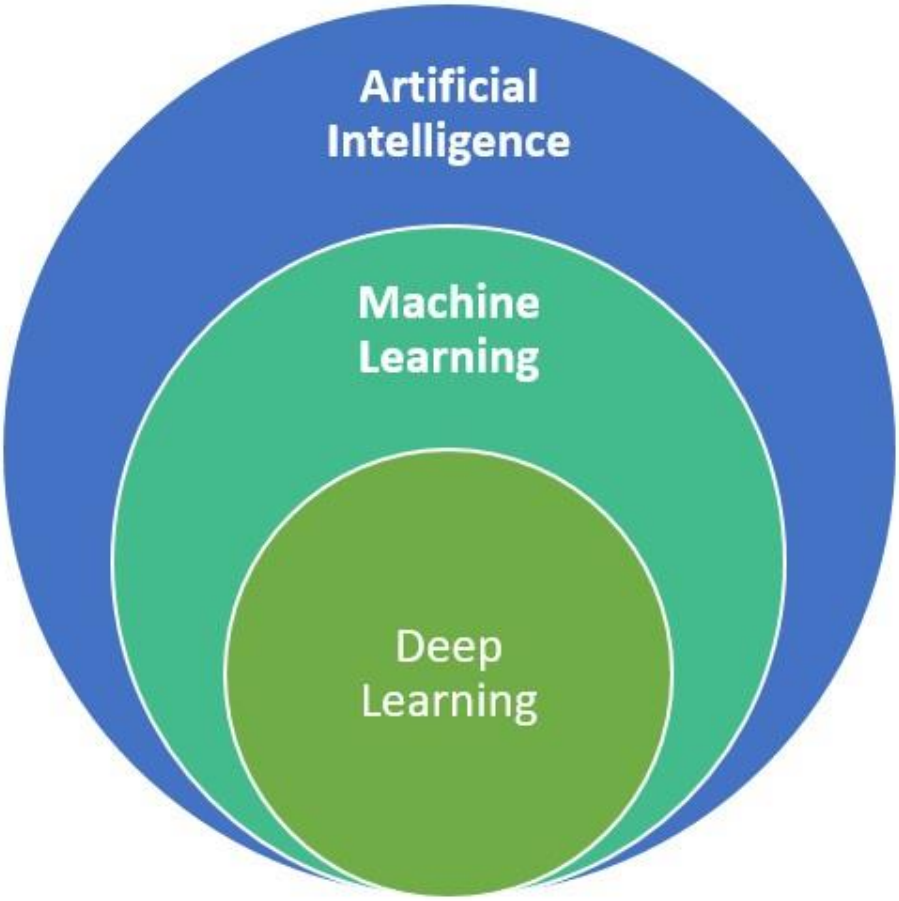
υπολογίζεται σε κάθε χρονικό βήμα, προκειμένου να τροφοδοτήσει τις πληροφορίες πίσω στον πράκτορα για να αποφασίσει την επόμενη κίνησή του. Ωστόσο, εάν ο πράκτορας ήταν σε θέση να προσομοιώσει το περιβάλλον του μέσω ενός γεννητικού μοντέλου, δεν θα χρειαζόταν να χρησιμοποιηθεί η στρατηγική στην προσομοίωση υπολογιστή ή στον πραγματικό κόσμο, αλλά θα μπορούσε να μάθει στο δικό του φανταστικό περιβάλλον.

Τέλος, αν θέλουμε να πούμε ειλικρινά ότι κατασκευάσαμε μια μηχανή που έχει αποκτήσει μια μορφή νοημοσύνης που είναι συγκρίσιμη με αυτή ενός ανθρώπου, η γεννητική μοντελοποίηση είναι σίγουρα ένα μέρος της λύσης. Ένα από τα καλύτερα παραδείγματα ενός γεννητικού μοντέλου στον φυσικό κόσμο είμαστε εμείς. Ας αφιερώσουμε λίγο χρόνο για να σκεφτούμε τι απίστευτα μοντέλα γεννητικής είμαστε. Μπορούμε να κλείσουμε τα μάτια μας και να φανταστούμε πώς θα έμοιαζε ένας σκύλος από οποιαδήποτε πιθανή γωνία. Μπορούμε να φανταστούμε έναν αριθμό εύλογων διαφορετικών καταλήξεων για την αγαπημένη μας τηλεοπτική εκπομπή και μπορούμε να σχεδιάσουμε την εβδομάδα μας μπροστά, δουλεύοντας διάφορα μελλοντικά σενάρια στο μυαλό μας και αναλαμβάνοντας δράση ανάλογα. Η τρέχουσα νευροεπιστημονική θεωρία προτείνει ότι η αντίληψή μας για την πραγματικότητα δεν είναι ένα εξαιρετικά περίπλοκο διακριτικό μοντέλο που λειτουργεί με την αισθητηριακή μας εισροή για να παράγει προβλέψεις για το τι βιώνουμε, αλλά είναι ένα γεννητικό μοντέλο που εκπαιδεύεται από τη γέννηση μας για να παράγει προσομοιώσεις του περιβάλλοντός μας που ταιριάζουν με ακρίβεια με το μέλλον. Σαφώς, η βαθιά κατανόηση του τρόπου με τον οποίο μπορούμε να κατασκευάσουμε μηχανές για να αποκτήσουμε αυτή την ικανότητα θα είναι κεντρικής σημασίας για τη συνεχή κατανόηση της λειτουργίας του εγκεφάλου και της γενικής τεχνητής νοημοσύνης.

Για να κατανοήσουμε καλύτερα την λειτουργία των GANs, πρέπει πρώτα να κατανοήσουμε τον σκοπό και το νόημα της βαθιάς μάθησης.

2. Εισαγωγή στην Βαθιά Μάθηση (Deep Learning)

Ξεκινώντας με έναν απλό ορισμό της βαθιάς μάθησης: Η βαθιά μάθηση είναι μια κατηγορία αλγορίθμων μηχανικής μάθησης που χρησιμοποιούν πολλαπλά στοιβαγμένα επίπεδα μονάδων επεξεργασίας για να μάθουν αναπαραστάσεις υψηλού επιπέδου από μη δομημένα δεδομένα.



Σχήμα 2-1. Η Βαθιά Μάθηση κατηγοριοποιημένη στην Τεχνητή Νοημοσύνη
(Πηγή ιστοσελίδα: https://master-iesc-angers.com/artificial-intelligence-machine-learning-and-deep-learning-same-context-different-concepts/)

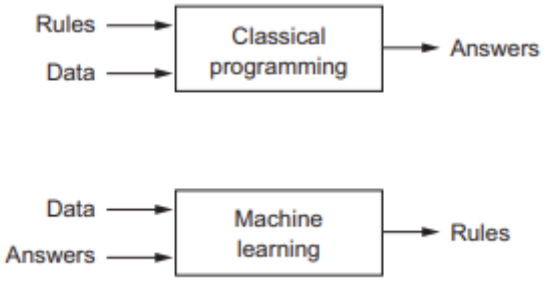
2.1 Τεχνητή νοημοσύνη

Η τεχνητή νοημοσύνη γεννήθηκε τη δεκαετία του 1950, όταν κάποιοι πρωτοπόροι από τον χώρο της επιστήμης των υπολογιστών άρχισαν να ρωτούν εάν οι υπολογιστές θα μπορούσαν να «σκέφτονται», ένα ερώτημα του οποίου τις προεκτάσεις εξακολουθούμε να διερευνούμε σήμερα. Ένας συνοπτικός ορισμός του πεδίου θα ήταν ο εξής: η προσπάθεια αυτοματοποίησης των πνευματικών εργασιών που συνήθως εκτελούνται από ανθρώπους. Ως εκ τούτου, η τεχνητή νοημοσύνη είναι ένα γενικό πεδίο που περιλαμβάνει τη μηχανική μάθηση και τη βαθιά μάθηση, αλλά περιλαμβάνει επίσης πολλούς περισσότερους τομείς, που δεν περιλαμβάνουν καμία μάθηση. Τα πρώτα προγράμματα σκακιού, για παράδειγμα, περιλάμβαναν μόνο κωδικοποιημένους κανόνες που είχαν δημιουργηθεί από προγραμματιστές και δεν πληρούσαν τις προϋποθέσεις για μηχανική μάθηση. Για αρκετά μεγάλο χρονικό διάστημα, πολλοί ειδικοί πίστευαν ότι η τεχνητή νοημοσύνη σε ανθρώπινο

επίπεδο θα μπορούσε να επιτευχθεί αν οι προγραμματιστές δημιουργήσουν ένα αρκετά μεγάλο σύνολο σαφών κανόνων για τον χειρισμό της γνώσης. Αυτή η προσέγγιση είναι γνωστή ως συμβολική τεχνητή νοημοσύνη (symbolic AI) και ήταν το κυρίαρχο παράδειγμα στην τεχνητή νοημοσύνη από τη δεκαετία του 1950 έως τα τέλη της δεκαετίας του 1980. Αν και η συμβολική τεχνητή νοημοσύνη αποδείχθηκε κατάλληλη για την επίλυση σαφώς καθορισμένων, λογικών προβλημάτων, όπως για το σκάκι, αποδείχθηκε ότι ήταν δύσκολο να βρεθούν σαφείς κανόνες για την επίλυση πιο πολύπλοκων, ασαφών προβλημάτων, όπως η ταξινόμηση εικόνων, η αναγνώριση ομιλίας και η μετάφραση γλώσσας. Μια νέα προσέγγιση προέκυψε για να πάρει τη θέση της συμβολικής τεχνητής νοημοσύνης: η μηχανική μάθηση.

2.2 Μηχανική μάθηση

Η μηχανική μάθηση προκύπτει από αυτό το ερώτημα: θα μπορούσε ένας υπολογιστής να προχωρήσει πέρα από τις οδηγίες που του βάζουμε να εκτελεί, να μπορέσει να μάθει μόνος του πώς να εκτελεί μια καθορισμένη εργασία; Αντί οι προγραμματιστές να δημιουργούν κανόνες επεξεργασίας δεδομένων με το χέρι, θα μπορούσε ένας υπολογιστής να μάθει αυτόματα αυτούς τους κανόνες κοιτάζοντας δεδομένα; Αυτή η ερώτηση ανοίγει την πόρτα σε ένα νέο πρότυπο προγραμματισμού. Στον κλασικό προγραμματισμό, το παράδειγμα της συμβολικής τεχνητής νοημοσύνης, οι άνθρωποι εισάγουν κανόνες (ένα πρόγραμμα) και δεδομένα προς επεξεργασία σύμφωνα με αυτούς τους κανόνες, βγαίνουν απαντήσεις (βλ. σχήμα 2-2.). Με τη μηχανική μάθηση, οι άνθρωποι εισάγουν δεδομένα καθώς και τις αναμενόμενες απαντήσεις από τα δεδομένα και βγαίνουν οι κανόνες. Αυτοί οι κανόνες μπορούν στη συνέχεια να εφαρμοστούν σε νέα δεδομένα για την παραγωγή νέων απαντήσεων.



Σχήμα 2-2. Η διαφορά μεταξύ κλασικού προγραμματισμού και μηχανικής μάθησης
(Πηγή: Deep Learning with Python, 2018)

Ένα σύστημα μηχανικής μάθησης εκπαιδεύεται και όχι ρητά προγραμματίζεται. Παρουσιάζεται με πολλά παραδείγματα σχετικά με μια εργασία και βρίσκει στατιστική δομή σε αυτά τα παραδείγματα που τελικά επιτρέπει στο σύστημα να βρει κανόνες για την αυτοματοποίηση της εργασίας. Για παράδειγμα, εάν θέλουμε να αυτοματοποιήσουμε την προσθήκη ετικετών στις φωτογραφίες των διακοπών μας, θα μπορούσαμε να παρουσιάσουμε ένα σύστημα μηχανικής μάθησης με πολλά παραδείγματα εικόνων που έχουν ήδη επισημανθεί από ανθρώπους και το σύστημα θα μάθει στατιστικούς κανόνες για τη συσχέτιση συγκεκριμένων εικόνων με συγκεκριμένες ετικέτες. Αν και η μηχανική μάθηση άρχισε να ανθίζει μόλις τη δεκαετία του 1990, έγινε γρήγορα το πιο δημοφιλές και πιο επιτυχημένο υποπεδίο της τεχνητής νοημοσύνης, μια τάση που καθοδηγείται από τη διαθεσιμότητα ταχύτερου υλικού (hardware) και μεγαλύτερων συνόλων δεδομένων (datasets). Η μηχανική μάθηση σχετίζεται στενά με τις μαθηματικές στατιστικές, αλλά διαφέρει από τις στατιστικές με πολλούς σημαντικούς τρόπους. Σε αντίθεση με τις

στατιστικές, η μηχανική μάθηση τείνει να ασχολείται με μεγάλα, πολύπλοκα σύνολα δεδομένων (όπως ένα σύνολο δεδομένων εκατομμυρίων εικόνων, το καθένα από τα οποία αποτελείται από δεκάδες χιλιάδες pixel) για τα οποία η κλασική στατιστική ανάλυση δεν θα ήταν πρακτική. Ως αποτέλεσμα, η μηχανική μάθηση, και ιδιαίτερα η βαθιά μάθηση, παρουσιάζει σχετικά λίγη μαθηματική θεωρία και είναι προσανατολισμένη στη μηχανική. Είναι μια πρακτική διαδικασία στην οποία οι ιδέες αποδεικνύονται εμπειρικά πιο συχνά, παρά θεωρητικά.

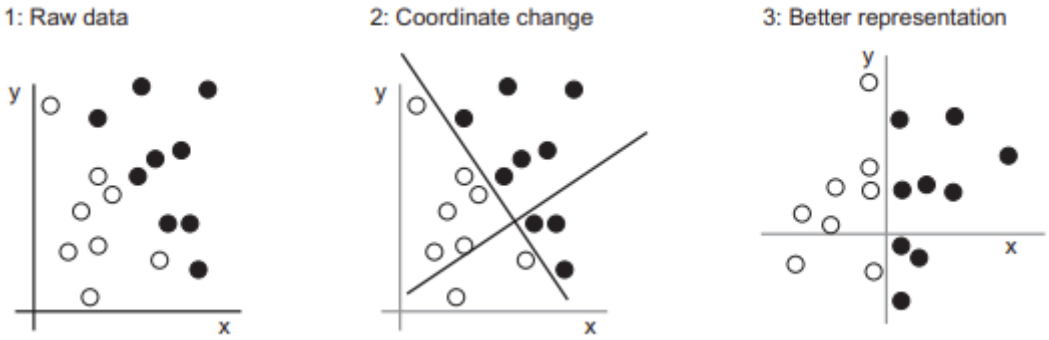
2.3 Εκμάθηση αναπαραστάσεων από δεδομένα

Για να ορίσουμε τη βαθιά μάθηση και να κατανοήσουμε τη διαφορά μεταξύ της βαθιάς μάθησης και άλλων προσεγγίσεων μηχανικής μάθησης, χρειαζόμαστε πρώτα μια ιδέα για το τι κάνουν οι αλγόριθμοι μηχανικής μάθησης. Προηγουμένως αναφέραμε ότι η μηχανική μάθηση ανακαλύπτει κανόνες για την εκτέλεση μιας εργασίας επεξεργασίας δεδομένων, δίνοντας παραδείγματα του αναμενόμενου. Έτσι, για να κάνουμε μηχανική μάθηση, χρειαζόμαστε τρία πράγματα:

- Σημεία δεδομένων εισόδου (input data points): Για παράδειγμα, εάν η εργασία είναι η αναγνώριση ομιλίας, αυτά τα σημεία δεδομένων θα μπορούσαν να είναι αρχεία ήχου ατόμων που μιλούν. Εάν η εργασία είναι η προσθήκη ετικετών σε εικόνες, θα μπορούσαν να είναι εικόνες.
- Παραδείγματα της αναμενόμενης εξόδου (examples of the expected output): Σε μια εργασία αναγνώρισης ομιλίας, αυτά θα μπορούσαν να είναι οι συχνότητες μεταγλωττισμένες. Σε μια εργασία εικόνας, τα αναμενόμενα αποτελέσματα θα μπορούσαν να είναι ετικέτες όπως "σκύλος", "γάτα" και ούτω καθεξής.
- Ένας τρόπος για να μετρήσουμε εάν ο αλγόριθμος κάνει καλή δουλειά (metric): Αυτό είναι απαραίτητο για να προσδιοριστεί η απόσταση μεταξύ της τρέχουσας εξόδου του αλγορίθμου και της αναμενόμενης εξόδου του. Η μέτρηση χρησιμοποιείται ως σήμα ανάδρασης για την προσαρμογή του τρόπου λειτουργίας του αλγόριθμου. Αυτό το βήμα προσαρμογής είναι αυτό που ονομάζουμε μάθηση.

Ένα μοντέλο μηχανικής μάθησης μετατρέπει τα δεδομένα εισόδου του σε ουσιαστικά αποτελέσματα, μια διαδικασία που «μαθαίνεται» από την έκθεση σε γνωστά παραδείγματα εισόδων και εξόδων. Επομένως, το κεντρικό πρόβλημα στη μηχανική μάθηση και τη βαθιά μάθηση είναι ο ουσιαστικός μετασχηματισμός των δεδομένων, με άλλα λόγια, η εκμάθηση χρήσιμων αναπαραστάσεων των δεδομένων εισόδου, αναπαραστάσεις που μας φέρνουν πιο κοντά στο αναμενόμενο αποτέλεσμα. Η αναπαράσταση ειδικότερα, είναι ένας διαφορετικός τρόπος να βλέπουμε ή να κωδικοποιούνται δεδομένα. Για παράδειγμα, μια έγχρωμη εικόνα μπορεί να κωδικοποιηθεί σε μορφή RGB (κόκκινο-πράσινο-μπλε) ή σε μορφή HSV (απόχρωση-κορεσμός-τιμή). Πρόκειται για δύο διαφορετικές αναπαραστάσεις των ίδιων δεδομένων. Ορισμένες εργασίες που μπορεί να είναι δύσκολες με μια αναπαράσταση μπορούν να γίνουν εύκολες με μια άλλη. Για παράδειγμα, η εργασία "επιλογή όλων των κόκκινων εικονοστοιχείων (pixel) στην εικόνα" είναι απλούστερη στη μορφή RGB, ενώ η "κάντε την εικόνα λιγότερο κορεσμένη" είναι απλούστερη στη μορφή HSV. Τα μοντέλα μηχανικής μάθησης έχουν να κάνουν με την εύρεση κατάλληλων αναπαραστάσεων για τα δεδομένα εισόδου τους, τους μετασχηματισμούς των δεδομένων που τα καθιστούν πιο επιδεκτικά στην εκάστοτε εργασία. Πιο συγκεκριμένα, ας θεωρήσουμε έναν άξονα x, έναν άξονα y και μερικά σημεία που αντιπροσωπεύονται από τις συντεταγμένες τους στο σύστημα (x, y), όπως φαίνεται στο σχήμα 2-3. Όπως μπορούμε να δούμε, έχουμε μερικά λευκά σημεία

και μερικά μαύρα σημεία. Ας υποθέσουμε ότι θέλουμε να αναπτύξουμε έναν αλγόριθμο που μπορεί να πάρει τις συντεταγμένες (x, y) ενός σημείου και να δώσει έξοδο αν αυτό το σημείο είναι πιθανό να είναι μαύρο ή λευκό. Σε αυτήν την περίπτωση, οι είσοδοι είναι οι συντεταγμένες των σημείων μας. Τα αναμενόμενα αποτελέσματα είναι τα χρώματα των σημείων μας. Ένας τρόπος για να μετρήσουμε εάν ο αλγόριθμός μας κάνει καλή δουλειά θα μπορούσε να είναι, για παράδειγμα, το ποσοστό των σημείων που ταξινομούνται σωστά. Αυτό που χρειαζόμαστε εδώ είναι μια νέα αναπαράσταση των δεδομένων μας που διαχωρίζει καθαρά τα λευκά σημεία από τα μαύρα σημεία. Ένας μετασχηματισμός που θα μπορούσαμε να χρησιμοποιήσουμε, μεταξύ πολλών άλλων δυνατοτήτων, θα ήταν μια αλλαγή συντεταγμένων, που απεικονίζεται στο σχήμα 2-3.

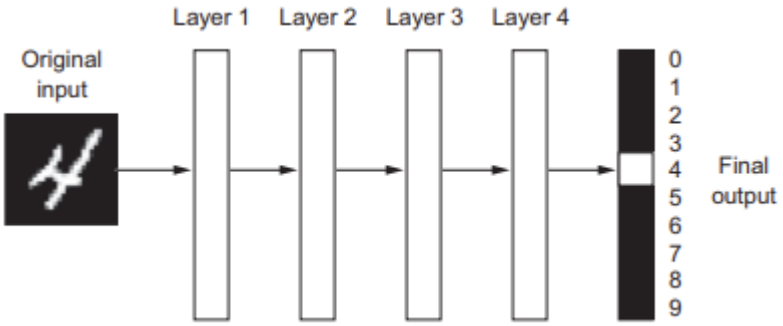


Σχήμα 2-3. 1) Αναπαράσταση δεδομένων στον άξονα (x,y) 2) Κατασκευή μιας καλύτερης αναπαράστασης 3) Τελική χρήσιμη αναπαράσταση
(Πηγή: Deep Learning with Python, 2018)

Σε αυτό το νέο σύστημα συντεταγμένων, οι συντεταγμένες των σημείων μας μπορούμε να πούμε ότι είναι μια νέα αναπαράσταση των δεδομένων μας. Με αυτήν την αναπαράσταση, το πρόβλημα ταξινόμησης μαύρου/λευκού μπορεί να εκφραστεί ως ένας απλός κανόνας: "Τα μαύρα σημεία είναι τέτοια ώστε x > 0" ή "Τα λευκά σημεία είναι τέτοια ώστε x < 0". Αυτή η νέα αναπαράσταση λύνει το πρόβλημα της ταξινόμησης. Σε αυτήν την περίπτωση, ορίσαμε την αλλαγή συντεταγμένων με το χέρι. Η μάθηση, στο πλαίσιο της μηχανικής μάθησης, περιγράφει μια διαδικασία αυτόματης αναζήτησης για καλύτερες αναπαραστάσεις. Όλοι οι αλγόριθμοι μηχανικής μάθησης συνίστανται στην αυτόματη εύρεση τέτοιων μετασχηματισμών που μετατρέπουν τα δεδομένα σε πιο χρήσιμες αναπαραστάσεις για μια δεδομένη εργασία. Οι αλγόριθμοι μηχανικής μάθησης δεν είναι συνήθως δημιουργικοί στην εύρεση αυτών των μετασχηματισμών, απλώς αναζητούν μέσα από ένα προκαθορισμένο σύνολο πράξεων, που ονομάζεται χώρος υποθέσεων. Αυτό λοιπόν είναι, η μηχανική μάθηση, τεχνικά: η αναζήτηση χρήσιμων αναπαραστάσεων ορισμένων δεδομένων εισόδου, μέσα σε έναν προκαθορισμένο χώρο δυνατοτήτων, χρησιμοποιώντας καθοδήγηση από ένα σήμα ανάδρασης. Αυτή η απλή ιδέα επιτρέπει την επίλυση ενός εξαιρετικά μεγάλου φάσματος πνευματικών εργασιών.

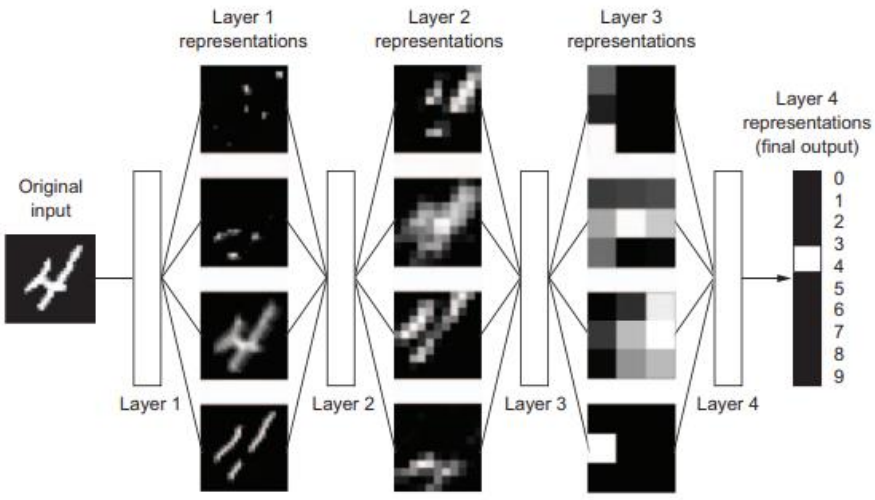
2.4 Το «βαθύ» στη βαθιά μάθηση

Η βαθιά μάθηση είναι ένα συγκεκριμένο υποπεδίο της μηχανικής μάθησης, μια νέα προσέγγιση στις αναπαραστάσεις εκμάθησης από δεδομένα που δίνουν σημασία στην εκμάθηση διαδοχικών στρωμάτων ολοένα και πιο ουσιαστικών αναπαραστάσεων. Η βαθιά μάθηση δεν είναι η αναφορά σε οποιοδήποτε είδος βαθύτερης κατανόησης που επιτυγχάνεται με την προσέγγιση. Αντίθετα, αντιπροσωπεύει την ιδέα των διαδοχικών στρωμάτων αναπαραστάσεων. Τα επίπεδα που συμβάλλουν σε ένα μοντέλο δεδομένων, χαρακτηρίζεται ως βάθος του μοντέλου. Η σύγχρονη βαθιά μάθηση περιλαμβάνει συχνά δεκάδες ή και εκατοντάδες διαδοχικά επίπεδα αναπαραστάσεων και μαθαίνονται όλα αυτόματα από την έκθεση σε δεδομένα εκπαίδευσης. Στη βαθιά μάθηση, αυτές οι πολύ-επίπεδες αναπαραστάσεις μαθαίνονται (σχεδόν πάντα) μέσω μοντέλων που ονομάζονται νευρωνικά δίκτυα. Ο όρος νευρωνικό δίκτυο είναι μια αναφορά στη νευροβιολογία, αλλά αν και ορισμένες από τις κεντρικές έννοιες στη βαθιά μάθηση εμπνεύστηκαν από την αντίληψη μας από τον εγκέφαλο, τα μοντέλα βαθιάς μάθησης δεν είναι μοντέλα του εγκεφάλου. Δεν υπάρχουν τεκμηριώσεις ότι ο εγκέφαλος λειτουργεί με παρόμοιο τρόπο με τους μηχανισμούς μάθησης που χρησιμοποιούνται στα σύγχρονα μοντέλα βαθιάς μάθησης. Για τους σκοπούς μας, η βαθιά μάθηση είναι ένα μαθηματικό πλαίσιο για την εκμάθηση αναπαραστάσεων από δεδομένα. Πώς μοιάζουν οι αναπαραστάσεις που μαθαίνονται από έναν αλγόριθμο βαθιάς μάθησης;



Σχήμα 2-4. Ένα βαθύ νευρωνικό δίκτυο για ταξινόμηση ψηφίων
(Πηγή: Deep Learning with Python, 2018)

Όπως μπορούμε να δούμε στο σχήμα 2-5, το δίκτυο τροποποιεί την ψηφιακή εικόνα σε αναπαραστάσεις που διαφέρουν όλο και περισσότερο από την αρχική εικόνα και ολοένα και πιο λεπτομερείς για το τελικό αποτέλεσμα. Μπορούμε να σκεφτούμε ένα βαθύ δίκτυο ως μια λειτουργία πολλαπλών σταδίων πληροφοριών-απόσταξης, όπου οι πληροφορίες διέρχονται από διαδοχικά φίλτρα και βγαίνουν όλο και πιο καθαρές (δηλαδή χρήσιμες σε σχέση με κάποια εργασία).

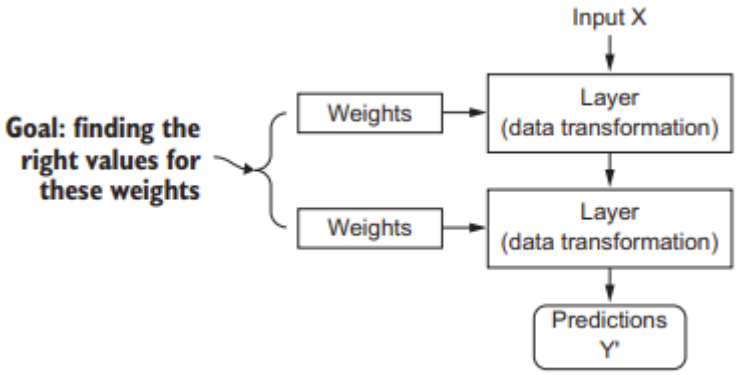


Σχήμα 2-5. Βαθιές αναπαραστάσεις που μαθαίνονται από ένα μοντέλο ψηφιακής ταξινόμησης
(Πηγή: Deep Learning with Python, 2018)

Αυτό είναι, λοιπόν, η βαθιά μάθηση, τεχνικά, ένας πολυσταδιακός τρόπος εκμάθησης αναπαραστάσεων δεδομένων. Είναι μια απλή ιδέα, αλλά, όπως αποδεικνύεται, πολύ απλοί μηχανισμοί, επαρκώς κλιμακωμένοι, μπορεί να καταλήξουν να μοιάζουν μαγικοί.

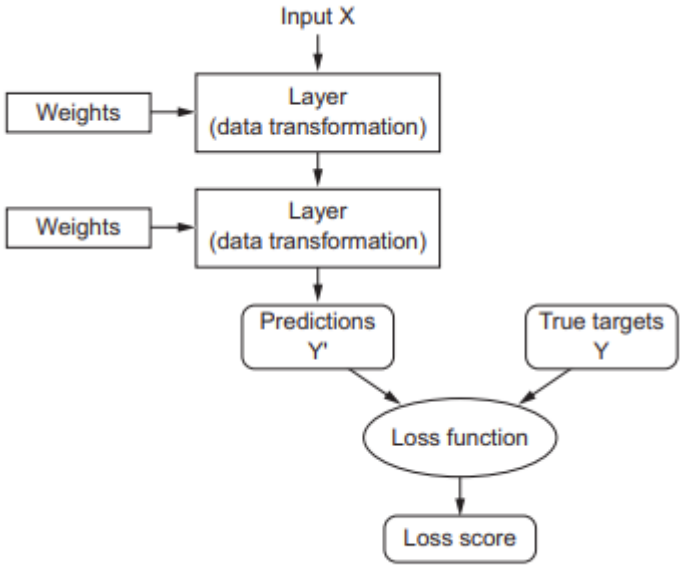
2.5 Κατανόηση του τρόπου λειτουργίας της βαθιάς μάθησης, σε τρία σχήματα

Συγκεκριμένα, η προδιαγραφή του τι κάνει ένα επίπεδο στα δεδομένα εισόδου του, αποθηκεύεται στα βάρη του στρώματος, τα οποία είναι μια δέσμη αριθμών. Σε τεχνικούς όρους, θα λέγαμε ότι ο μετασχηματισμός που προκύπτει από ένα επίπεδο παραμετροποιείται από τα βάρη (weights) του (βλ. σχήμα 2-6). (Τα βάρη ονομάζονται μερικές φορές και παράμετροι ενός επιπέδου). Σε αυτό το πλαίσιο, μάθηση σημαίνει η ανίχνευση ενός συνόλου τιμών για τα βάρη (weights) όλων των επιπέδων σε ένα δίκτυο, έτσι ώστε το δίκτυο να αντιστοιχίζει σωστά τις εισόδους παραδειγμάτων στους ανάλογους στόχους τους. Αλλά εδώ χρειάζεται να εστιάσουμε, διότι ένα βαθύ νευρωνικό δίκτυο μπορεί να αποτελείται από δεκάδες εκατομμύρια παραμέτρους. Η ανίχνευση της σωστής τιμής για όλες μπορεί να φαίνεται ακατόρθωτη εργασία, καθώς η τροποποίηση της τιμής μιας παραμέτρου θα επηρεάσει τη συμπεριφορά όλων των άλλων.



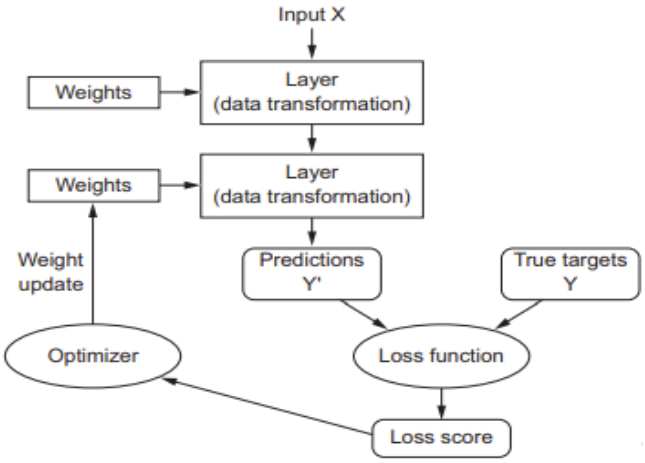
Σχήμα 2-6. Παραμετροποίηση του δικτύου από τα βάρη
(Πηγή: Deep Learning with Python, 2018)

Για να εξετάσουμε κάτι, πρέπει πρώτα να είμαστε σε θέση να το παρατηρήσουμε. Για να εξετάσουμε την έξοδο ενός νευρωνικού δικτύου, πρέπει να είμαστε σε θέση να υπολογίσουμε πόσο απέχει αυτή η έξοδος από αυτό που αναμένουμε. Αυτός είναι ο ρόλος της συνάρτησης απώλειας (loss function) του δικτύου. Η συνάρτηση απώλειας λαμβάνει τις προβλέψεις του δικτύου και τον αληθινό στόχο (αυτό που θέλουμε να δώσει το δίκτυο) και υπολογίζει μια βαθμολογία απόστασης, καταγράφοντας πόσο καλά τα πήγε το δίκτυο σε αυτό το συγκεκριμένο παράδειγμα (βλ. εικόνα 2-7).



Σχήμα 2-7. Η συνάρτηση απώλειας υπολογίζει την ποιότητα της εξόδου του δικτύου
(Πηγή: Deep Learning with Python, 2018)

Το θεμελιώδες τέχνασμα στη βαθιά μάθηση είναι να χρησιμοποιήσουμε αυτή τη βαθμολογία ως σήμα ανάδρασης για να προσαρμόσουμε λίγο την τιμή των βαρών (weights), προς μια κατεύθυνση που θα μειώσει τη βαθμολογία απώλειας (loss score) για το συγκεκριμένο παράδειγμα (βλ. εικόνα 2-8). Αυτή η προσαρμογή είναι εργασία του βελτιστοποιητή (optimizer), ο οποίος εφαρμόζει αυτό που ονομάζεται αλγόριθμος Backpropagation: ο θεμελιώδης αλγόριθμος στη βαθιά μάθηση. Αργότερα εξηγούμε με περισσότερες λεπτομέρειες πώς λειτουργεί ο αλγόριθμος αυτός. Υπάρχουν και άλλοι αλγόριθμοι, αλλά αυτός είναι ο πιο διαδεδομένος.



Σχήμα 2-8. Η απώλεια χρησιμοποιείται ως ανάδραση για την βελτίωση των βαρών
(Πηγή: Deep Learning with Python, 2018)

Αρχικά, εκχωρούνται τυχαίες τιμές στα βάρη (weights) του δικτύου, έτσι το δίκτυο απλώς πραγματοποιεί μια σειρά από τυχαίους μετασχηματισμούς. Φυσικά, η απόδοσή του απέχει πολύ από το ιδανικό και η βαθμολογία απώλειας (loss score) είναι συνεπώς πολύ υψηλή. Αλλά με κάθε επανάληψη που επεξεργάζεται το δίκτυο, τα βάρη ρυθμίζονται λίγο προς την σωστή κατεύθυνση και η βαθμολογία απώλειας μειώνεται. Αυτός είναι ο βρόχος προπόνησης (training loop), ο οποίος, επαναλαμβανόμενος αρκετές φορές (συνήθως δεκάδες επαναλήψεις σε χιλιάδες παραδείγματα), δεσμεύουν τιμές βάρους που ελαχιστοποιούν τη συνάρτηση απώλειας. Ένα δίκτυο με ελάχιστη απώλεια (loss) είναι ένα δίκτυο για το οποίο οι έξοδοι είναι όσο πιο κοντά μπορούν να είναι στους στόχους: ένα εκπαιδευμένο δίκτυο.

2.6 Τι έχει επιτύχει μέχρι τώρα η βαθιά μάθηση

Αν και η βαθιά μάθηση είναι ένα αρκετά παλιό υποπεδίο της μηχανικής μάθησης, αναδείχθηκε μόνο στις αρχές της δεκαετίας του 2010. Μέσα σε λίγα χρόνια από τότε, δεν πέτυχε τίποτα λιγότερο από μια επανάσταση στον τομέα, με αξιοσημείωτα αποτελέσματα σε προβλήματα αντίληψης, όπως προβλήματα όρασης και ακοής που περιλαμβάνουν δεξιότητες που φαίνονται φυσικές και απλές για τον άνθρωπο, αλλά ήταν από καιρό απρόσιτο για τις μηχανές. Συγκεκριμένα, η βαθιά μάθηση έχει επιτύχει τις ακόλουθες ανακαλύψεις, όλες σε ιστορικά δύσκολους τομείς της μηχανικής μάθησης:

- Ταξινόμηση εικόνων σχεδόν σε ανθρώπινο επίπεδο
- Αναγνώριση ομιλίας σχεδόν σε ανθρώπινο επίπεδο
- Αναγνώριση χαρακτήρων σχεδόν σε ανθρώπινο επίπεδο
- Βελτιωμένη αυτόματη μετάφραση
- Βελτιωμένη μετατροπή κειμένου σε ομιλία
- Ψηφιακούς βοηθούς όπως το Google Now, το Amazon Alexa και το Apple Siri
- Βελτιωμένη διαφήμιση
- Βελτιωμένα αποτελέσματα αναζήτησης στον Ιστό
- Ικανότητα απάντησης σε ερωτήσεις φυσικής γλώσσας

Εξακολουθούμε να διερευνούμε την πλήρη έκταση του τι μπορεί να κάνει η βαθιά μάθηση.

3. Πριν από τη βαθιά μάθηση: Μια σύντομη ιστορία της μηχανικής μάθησης

Η βαθιά μάθηση έχει φτάσει σε ένα επίπεδο της δημόσιας προσοχής και των επενδύσεων της βιομηχανίας που δεν έχουμε ξαναδεί στην ιστορία της τεχνητής νοημοσύνης, αλλά δεν είναι η πρώτη επιτυχημένη μορφή μηχανικής μάθησης. Είναι ασφαλές να πούμε ότι οι περισσότεροι από τους αλγόριθμους μηχανικής μάθησης που χρησιμοποιούνται στη βιομηχανία σήμερα δεν είναι αλγόριθμοι βαθιάς μάθησης. Η βαθιά μάθηση δεν είναι πάντα το κατάλληλο εργαλείο για τον κάθε σκοπό, μερικές φορές δεν υπάρχουν αρκετά δεδομένα για να είναι εφαρμόσιμη και μερικές φορές το πρόβλημα επιλύεται καλύτερα με διαφορετικό αλγόριθμο. Μια λεπτομερής συζήτηση για τις κλασικές προσεγγίσεις μηχανικής μάθησης είναι έξω από το πεδίο αυτής της εργασίας, αλλά θα τις εξετάσουμε εν συντομία και θα περιγράψουμε το ιστορικό πλαίσιο στο οποίο αναπτύχθηκαν. Αυτό θα μας επιτρέψει να τοποθετήσουμε τη βαθιά μάθηση στο ευρύτερο πλαίσιο της μηχανικής μάθησης και να κατανοήσουμε καλύτερα από πού προέρχεται η βαθιά μάθηση και γιατί έχει σημασία.

3.1 Πιθανοτική μοντελοποίηση

Η πιθανοτική μοντελοποίηση είναι η εφαρμογή των αρχών της στατιστικής στην ανάλυση δεδομένων. Ήταν μια από τις πρώτες μορφές μηχανικής μάθησης και εξακολουθεί να χρησιμοποιείται ευρέως μέχρι σήμερα. Ένας από τους πιο γνωστούς αλγόριθμους αυτής της κατηγορίας είναι ο αλγόριθμος Naive Bayes[3]. Το Naive Bayes είναι ένας τύπος ταξινομητή μηχανικής μάθησης που βασίζεται στην εφαρμογή του θεωρήματος του Bayes όπου υποθέτουμε ότι τα χαρακτηριστικά στα δεδομένα εισόδου είναι όλα ανεξάρτητα. Αυτή η μορφή ανάλυσης δεδομένων προϋπήρχε και εφαρμόζονταν με το χέρι δεκαετίες πριν από την πρώτη της εφαρμογή σε υπολογιστή. Ένα στενά συνδεδεμένο μοντέλο είναι η λογιστική παλινδρόμηση[4] (logreg για συντομία), η οποία μερικές φορές θεωρείται ότι είναι η Α Β της σύγχρονης μηχανικής μάθησης. Το Naive Bayes και η λογιστική παλινδρόμηση εξακολουθούν να είναι χρήσιμα μέχρι σήμερα, χάρη στην απλή και ευέλικτη φύση τους. Είναι συχνά το πρώτο πράγμα που θα δοκιμάσει ένας επιστήμονας δεδομένων (data scientist) σε ένα σύνολο δεδομένων για να πάρει μια αίσθηση για την εργασία που έχει.

3 Naive Bayes: Τεχνική ταξινόμησης για τον προσδιορισμό των πιθανοτήτων των κλάσεων, όπου οι μεταβλητές είναι ανεξάρτητες δεδομένης τη κλάσης
4 Λογιστική παλινδρόμηση: Ερευνά το μη γραμμικό αποτέλεσμα μιας εξαρτημένης μεταβλητής αναφορικά με την δράση πολλών ανεξάρτητων μεταβλητών

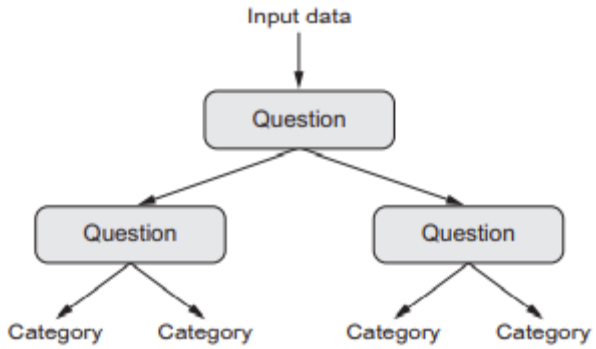
3.2 Πρώιμα νευρωνικά δίκτυα

Οι πρώτες χρήσεις των νευρωνικών δικτύων έχουν αντικατασταθεί πλήρως από τις σύγχρονες παραλλαγές. Οι βασικές ιδέες των νευρωνικών δικτύων ερευνήθηκαν σε μορφές παιχνιδιών ήδη από τη δεκαετία του 1950, η λεπτομερείς προσέγγιση χρειάστηκε δεκαετίες για να ξεκινήσει. Για πολύ καιρό, το κομμάτι που έλειπε ήταν ένας αποτελεσματικός τρόπος εκπαίδευσης μεγάλων νευρωνικών δικτύων. Αυτό άλλαξε στα μέσα της δεκαετίας του 1980, όταν πολλοί άνθρωποι χρησιμοποίησαν τον αλγόριθμο Backpropagation, έναν τρόπο εκπαίδευσης αλυσίδων παραμετρικών πράξεων χρησιμοποιώντας βελτιστοποίηση gradient-descent (κάθοδος κλήσης) και άρχισαν να τον εφαρμόζουν στα νευρωνικά δίκτυα. Η πρώτη επιτυχημένη πρακτική εφαρμογής των νευρωνικών δικτύων ήρθε το 1989 από τα Bell Labs,

όταν ο Yann LeCun συνδύασε τις προηγούμενες ιδέες των συνελικτικών νευρωνικών δικτύων και του backpropagation και τις εφάρμοσε στο πρόβλημα της ταξινόμησης χειρόγραφων ψηφίων. Το δίκτυο που προέκυψε, που ονομάστηκε LeNet, χρησιμοποιήθηκε από την Ταχυδρομική Υπηρεσία των Ηνωμένων Πολιτειών τη δεκαετία του 1990 για να αυτοματοποιήσει την ανάγνωση ταχυδρομικών κωδίκων σε φακέλους αλληλογραφίας.

3.3 Δέντρα απόφασης και τυχαία δάση (Decision trees and random forests)

Τα δέντρα αποφάσεων είναι δομές που μοιάζουν με διαγράμματα ροής που μας επιτρέπουν να ταξινομήσουμε σημεία δεδομένων εισόδου ή να προβλέψουμε τιμές εξόδου δεδομένων εισόδων (βλ. εικόνα 3-1). Είναι εύκολο να απεικονιστούν και να ερμηνευτούν. Τα δέντρα αποφάσεων που μάθαιναν από δεδομένα άρχισαν να παρουσιάζουν σημαντικό ερευνητικό ενδιαφέρον τη δεκαετία του 2000 και μέχρι το 2010.



Σχήμα 3-1. Δέντρο απόφασης, οι παράμετροι που εκπαιδεύονται είναι οι ερωτήσεις για τα δεδομένα (Πηγή: Deep Learning with Python, 2018)

Συγκεκριμένα, ο αλγόριθμος Random Forest (Τυχαία Δάση) εισήγαγε μια ισχυρή, πρακτική προσέγγιση στη μάθηση του δέντρου αποφάσεων που περιλαμβάνει τη δημιουργία ενός μεγάλου αριθμού εξειδικευμένων δέντρων αποφάσεων και στη συνέχεια τη συναρμολόγηση των εξόδων τους. Τα Random Forest μπορούν να εφαρμοστούν σε ένα ευρύ φάσμα προβλημάτων, θα μπορούσαμε να πούμε ότι είναι σχεδόν πάντα ο δεύτερος καλύτερος αλγόριθμος για κάθε ρηχή εργασία μηχανικής μάθησης (shallow machine learning).

3.4 Επιστροφή στα νευρωνικά δίκτυα

Γύρω στο 2010, αν και τα νευρωνικά δίκτυα αποκλείστηκαν σχεδόν πλήρως από την ευρύτερη επιστημονική κοινότητα, ένας αριθμός ανθρώπων που εξακολουθούσαν να εργάζονται σε νευρωνικά δίκτυα άρχισαν να κάνουν σημαντικές ανακαλύψεις: οι ομάδες του Geoffrey Hinton στο Πανεπιστήμιο του Τορόντο, του Yoshua Bengio στο Πανεπιστήμιο του Μόντρεαλ. , Yann LeCun στο Πανεπιστήμιο της Νέας Υόρκης και IDSIA στην Ελβετία. Το 2011, ο Dan Ciresan από το IDSIA άρχισε να κερδίζει ακαδημαϊκούς διαγωνισμούς ταξινόμησης εικόνων με βαθιά νευρωνικά δίκτυα εκπαιδευμένα σε GPU (Graphics Processing Unit), την πρώτη πρακτική επιτυχία της σύγχρονης βαθιάς μάθησης. Αλλά η στιγμή ορόσημο ήρθε το 2012, με την είσοδο του ομίλου Hinton στον ετήσιο μεγάλης κλίμακας διαγωνισμό ταξινόμησης εικόνων ImageNet[5]. Η πρόκληση ImageNet ήταν εμφανώς δύσκολη εκείνη την εποχή, καθώς συνίσταται στην ταξινόμηση των έγχρωμων εικόνων υψηλής ανάλυσης σε 1.000 διαφορετικές κατηγορίες μετά από εκπαίδευση σε 1,4 εκατομμύρια εικόνες. Το 2011, η ακρίβεια των πέντε κορυφαίων μοντέλων, με βάση τις κλασικές προσεγγίσεις της όρασης υπολογιστών, ήταν μόνο 74,3%. Στη συνέχεια, το 2012, μια ομάδα με επικεφαλής τον Alex Krizhevsky και με συμβουλές από τον Geoffrey Hinton μπόρεσε να επιτύχει ακρίβεια 83,6%, μια σημαντική βελτίωση. Από το 2012, τα βαθιά συνελικτικά νευρωνικά δίκτυα (convnets) έχουν γίνει ο βασικός αλγόριθμος για όλες τις εργασίες όρασης υπολογιστή. Σε μεγάλα συνέδρια υπολογιστικής όρασης το 2015 και το 2016, ήταν σχεδόν αδύνατο να βρεθούν παρουσιάσεις που δεν περιλάμβαναν convnet σε κάποια μορφή. Ταυτόχρονα, η βαθιά μάθηση έχει βρει εφαρμογές και σε πολλά άλλα είδη προβλημάτων, όπως η επεξεργασία φυσικής γλώσσας.

5 ImageNet: Ένα από τα δημοφιλέστερα σετ δεδομένων εικόνας (image dataset)

4. Τι κάνει τη βαθιά μάθηση διαφορετική

Ο κύριος λόγος που η βαθιά μάθηση απογειώθηκε τόσο γρήγορα είναι ότι προσέφερε καλύτερη απόδοση σε πολλά προβλήματα. Αλλά δεν είναι αυτός ο μόνος λόγος. Η βαθιά μάθηση κάνει επίσης πολύ πιο εύκολη την επίλυση προβλημάτων, επειδή αυτοματοποιεί πλήρως αυτό που ήταν το πιο κρίσιμο βήμα σε μια ροή εργασιών μηχανικής μάθησης: τη μηχανική χαρακτηριστικών (feature engineering). Οι άνθρωποι έπρεπε να δημιουργήσουν χειροκίνητα καλά στρώματα αναπαραστάσεων για τα δεδομένα τους. Η βαθιά μάθηση, από την άλλη πλευρά, αυτοματοποιεί πλήρως αυτό το βήμα, μαθαίνονται όλες οι δυνατότητες με ένα πέρασμα αντί να χρειάζεται να τις σχεδιάσουμε μόνοι μας. Αυτό που είναι εφευρετικό για τη βαθιά μάθηση είναι ότι επιτρέπει σε ένα μοντέλο να μαθαίνει όλα τα επίπεδα αναπαράστασης από κοινού, ταυτόχρονα, και όχι διαδοχικά (άπληστα - greedily, όπως λέγεται). Με την κοινή εκμάθηση χαρακτηριστικών, κάθε φορά που το μοντέλο προσαρμόζει ένα από τα εσωτερικά του χαρακτηριστικά, όλα τα άλλα χαρακτηριστικά που εξαρτώνται από αυτό προσαρμόζονται αυτόματα στην αλλαγή, χωρίς να απαιτείται ανθρώπινη παρέμβαση. Όλα επιβλέπονται από ένα μόνο σήμα ανάδρασης. Αυτά είναι τα δύο βασικά χαρακτηριστικά του τρόπου με τον οποίο η βαθιά μάθηση μαθαίνει από τα δεδομένα: ο σταδιακός, επίπεδο προς επίπεδο τρόπος με τον οποίο αναπτύσσονται όλο και πιο περίπλοκες αναπαραστάσεις και το γεγονός ότι αυτές οι ενδιάμεσες σταδιακές αναπαραστάσεις μαθαίνονται από κοινού, ενώ κάθε επίπεδο ενημερώνεται και ακολουθεί το ανάλογο βήμα. Μαζί, αυτές οι δύο ιδιότητες έχουν κάνει τη βαθιά μάθηση πολύ πιο επιτυχημένη από προηγούμενες προσεγγίσεις στη μηχανική μάθηση.

4.1 Η άνοδος τη βαθιάς μάθησης

Οι δύο βασικές ιδέες της βαθιάς μάθησης, τα νευρωνικά δίκτυα και ο αλγόριθμος backpropagation, ήταν ήδη καλά κατανοητές το 1989. Γιατί λοιπόν η βαθιά μάθηση απογειώθηκε μετά το 2012; Τι άλλαξε σε αυτές τις δύο δεκαετίες; Γενικά, τρεις τεχνικές δυνάμεις οδηγούν την πρόοδο στη μηχανική μάθηση:

- Υλικό (Hardware)
- Σύνολα δεδομένων και σημεία αναφοράς (Datasets and benchmarks)
- Αλγόριθμοι (Algorithms)

Επειδή το πεδίο καθοδηγείται από πειραματικά ευρήματα και όχι από θεωρία, οι αλγοριθμικοί πρόοδοι καθίστανται δυνατοί μόνο όταν είναι διαθέσιμα κατάλληλα δεδομένα και υλικό για τη δοκιμή νέων ιδεών (ή την κλιμάκωση των παλιών ιδεών). Η μηχανική μάθηση δεν είναι μαθηματικά ή φυσική, όπου μπορούν να γίνουν σημαντικές πρόοδοι με ένα στυλό και ένα κομμάτι χαρτί. Είναι μια επιστήμη μηχανικής. Τα πραγματικά σημεία περιορισμού κατά τη διάρκεια των δεκαετιών του 1990 και του 2000 ήταν τα δεδομένα και το υλικό. Αλλά να τι συνέβη εκείνη την περίοδο; Το διαδίκτυο απογειώθηκε και τα τσιπ γραφικών υψηλής απόδοσης (high-performance graphics chips) αναπτύχθηκαν για τις ανάγκες της αγοράς βιντεοπαιχνιδιών.

4.1.1 Υλικό (Hardware)

Μεταξύ 1990 και 2010, οι CPU (Central Processing Unit) έγιναν ταχύτερες. Ως αποτέλεσμα, σήμερα είναι δυνατό να εκτελούνται μικρά μοντέλα βαθιάς μάθησης μέσω φορητού

υπολογιστή (laptop), ενώ αυτό θα ήταν ακατόρθωτο πριν από 30 χρόνια. Καθ' όλη τη διάρκεια της δεκαετίας του 2000, εταιρείες όπως η NVIDIA και η AMD επένδυσαν δισεκατομμύρια δολάρια στην ανάπτυξη γρήγορων, μαζικά παράλληλων τσιπ (μονάδες επεξεργασίας γραφικών, GPUs) για να τροφοδοτήσουν τα γραφικά ολοένα και πιο φωτορεαλιστικών βιντεοπαιχνιδιών, για να αποδίδουν πολύπλοκα τρισδιάστατες σκηνές στην οθόνη μας σε πραγματικό χρόνο. Επιπλέον, ο κλάδος της βαθιάς μάθησης αρχίζει να υπερβαίνει τις GPU και επενδύει σε όλο και πιο εξειδικευμένα, αποτελεσματικά τσιπ αποκλειστικά για βαθιά μάθηση. Το 2016, η Google αποκάλυψε το έργο της, μια μονάδας επεξεργασίας τανυστών (TPU Tensor Processing Unit), ένα νέο σχέδιο τσιπ που αναπτύχθηκε από την αρχή για να τρέχει βαθιά νευρωνικά δίκτυα, το οποίο είναι 10 φορές πιο γρήγορο και πολύ πιο ενεργειακά αποδοτικό από τα κορυφαία GPU της εποχής μας.

4.1.2 Δεδομένα (Data)

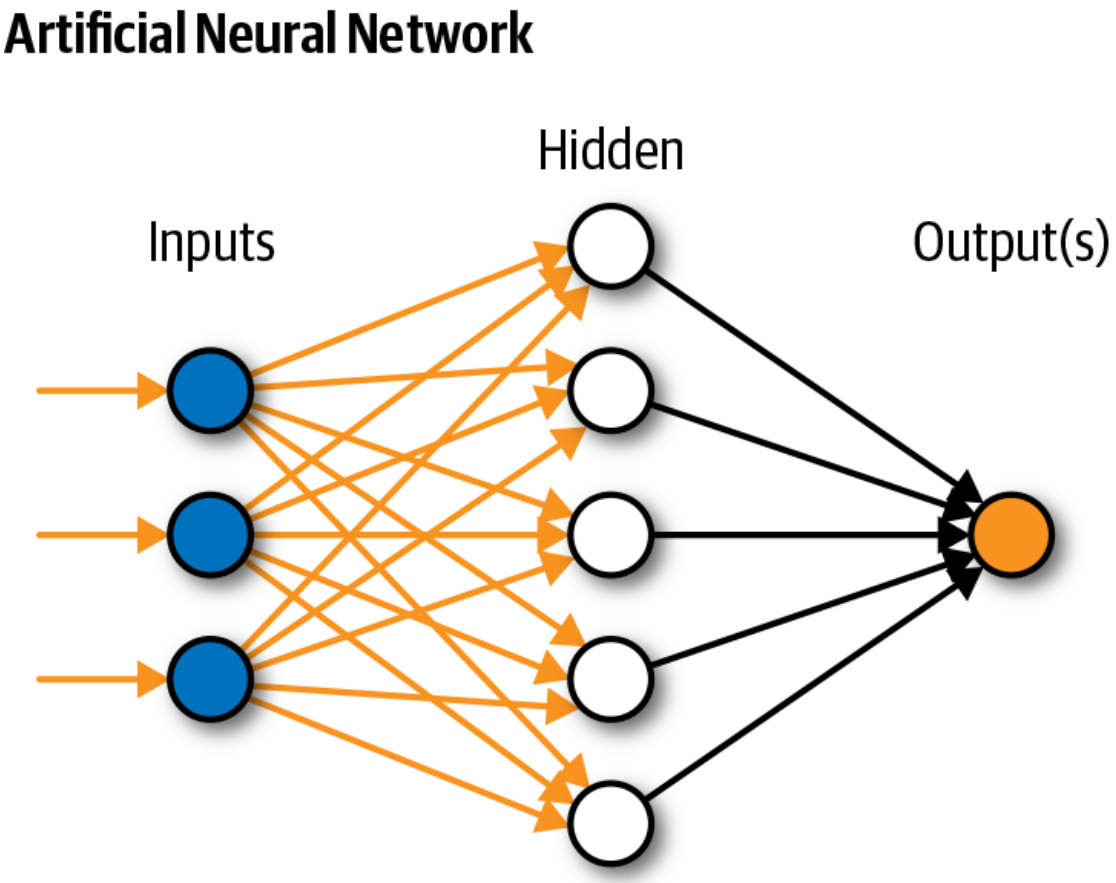
Η τεχνητή νοημοσύνη μερικές φορές προαναγγέλλεται ως η νέα βιομηχανική επανάσταση. Εάν η βαθιά μάθηση είναι η ατμομηχανή αυτής της επανάστασης, τότε τα δεδομένα είναι ο άνθρακας της. Η βασική αλλαγή ήταν η άνοδος του διαδικτύου, καθιστώντας εφικτή τη συλλογή και τη διανομή πολύ μεγάλων συνόλων δεδομένων για μηχανική μάθηση. Σήμερα, μεγάλες εταιρείες εργάζονται με σύνολα δεδομένων εικόνας, σύνολα δεδομένων βίντεο και σύνολα δεδομένων φυσικής γλώσσας που δεν θα μπορούσαν να συλλεχθούν χωρίς το διαδίκτυο. Εάν υπάρχει ένα σύνολο δεδομένων που υπήρξε καταλύτης για την άνοδο της βαθιάς μάθησης, αυτό είναι το σύνολο δεδομένων ImageNet, που αποτελείται από 1,4 εκατομμύρια εικόνες που έχουν επισημανθεί με το χέρι με 1.000 κατηγορίες εικόνων (1 κατηγορία ανά εικόνα). Αλλά αυτό που κάνει το ImageNet ξεχωριστό δεν είναι μόνο το μεγάλο του μέγεθος, αλλά και ο ετήσιος διαγωνισμός που συνδέεται με αυτό. Οι δημόσιοι διαγωνισμοί είναι ένας εξαιρετικός τρόπος για να παρακινηθούν ερευνητές και μηχανικοί να προωθήσουν τα πειράματα τους.

4.1.3 Αλγόριθμοι (Algorithms)

Εκτός από το υλικό και τα δεδομένα, μέχρι τα τέλη της δεκαετίας του 2000, μας έλειπε ένας αξιόπιστος τρόπος να εκπαιδεύσουμε πολύ βαθιά νευρωνικά δίκτυα. Ως αποτέλεσμα, τα νευρωνικά δίκτυα ήταν ακόμα αρκετά ρηχά, χρησιμοποιώντας μόνο ένα ή δύο στρώματα αναπαραστάσεων, καθώς στα βαθιά στρώματα χανώντουσαν σημαντικές πληροφορίες. Αυτό άλλαξε γύρω στο 2009-2010 με την εμφάνιση πολλών απλών αλλά σημαντικών αλγοριθμικών βελτιώσεων που επέτρεψαν καλύτερη διάδοση κλίσης. Καλύτερες συναρτήσεις ενεργοποίησης για νευρωνικά στρώματα (activation functions), καλύτερα σχήματα αρχικοποίησης βάρους (weight-initialization schemes), βελτιστοποιημένα συστήματα βάρους (weight-initialization schemes), όπως το RMSProp και το Adam. Μόνο όταν αυτές οι βελτιώσεις άρχισαν, τότε κατάφεραν τα μοντέλα να εκπαιδεύονται με 10 ή περισσότερα επίπεδα. Τέλος, το 2014, το 2015 και το 2016, ανακαλύφθηκαν ακόμη πιο προηγμένοι τρόποι για να βοηθηθεί η διάδοση της κλίσης, όπως η κανονικοποίηση παρτίδας (batch normalization), οι υπολειπόμενες συνδέσεις (residual connections) και οι διαχωρίσιμες συνελίξεις κατά βάθος (depth wise separable convolutions). Σήμερα μπορούμε να εκπαιδεύσουμε μοντέλα που έχουν βάθος χιλιάδων στρωμάτων.

5. Βαθιά νευρωνικά δίκτυα

Η πλειοψηφία των συστημάτων βαθιάς μάθησης είναι τεχνητά νευρωνικά δίκτυα με πολλαπλά στοιβαγμένα κρυφά επίπεδα. Για το λόγο αυτό, η βαθιά μάθηση έχει γίνει πλέον σχεδόν συνώνυμη με τα βαθιά νευρωνικά δίκτυα. Ωστόσο, είναι σημαντικό να επισημανθεί ότι κάθε σύστημα που χρησιμοποιεί πολλά επίπεδα για να μάθει αναπαραστάσεις υψηλού επιπέδου των δεδομένων εισόδου είναι επίσης μια μορφή βαθιάς μάθησης (π.χ. δίκτυα βαθιάς πεποίθησης και μηχανές βαθιάς Boltzmann). Ένα βαθύ νευρωνικό δίκτυο αποτελείται από μια σειρά από στοιβαγμένα επίπεδα. Κάθε στρώμα περιέχει μονάδες που συνδέονται με τις μονάδες του προηγούμενου στρώματος μέσω ενός συνόλου βαρών. Με τη στοίβαξη επιπέδων, οι μονάδες σε κάθε επόμενο επίπεδο μπορούν να αντιπροσωπεύουν όλο και πιο αναλυτικές πτυχές της αρχικής εισόδου. Το τελικό επίπεδο εξόδου είναι το αποκορύφωμα αυτής της διαδικασίας, όπου το δίκτυο εξάγει ένα σύνολο αριθμών που μπορούν να μετατραπούν σε πιθανότητες, για να αντιπροσωπεύσουν την πιθανότητα της αρχικής εισόδου.



Σχήμα 5-1. Ένα νευρωνικό δίκτυο με ένα κρυφό επίπεδο
(Πηγή ιστοσελίδα: https://www.whyofai.com/blog/ai-explained)

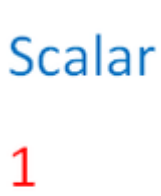
Τα βαθιά νευρωνικά δίκτυα μπορούν να έχουν οποιοδήποτε αριθμό μεσαίων ή κρυφών επιπέδων. Για παράδειγμα, το ResNet, σχεδιασμένο για αναγνώριση εικόνας, περιέχει 152 επίπεδα. Μπορούμε να χρησιμοποιήσουμε βαθιά νευρωνικά δίκτυα για να επηρεάσουμε χαρακτηριστικά υψηλού επιπέδου μιας εικόνας, όπως το χρώμα των μαλλιών ή την έκφραση ενός προσώπου, τροποποιώντας χειροκίνητα τις τιμές αυτών των κρυφών επιπέδων.

5.1 Αναπαραστάσεις δεδομένων για νευρωνικά δίκτυα

Γενικά, όλα τα τρέχοντα συστήματα μηχανικής μάθησης χρησιμοποιούν τανυστές (tensors) ως βασική δομή δεδομένων. Οι τανυστές είναι θεμελιώδεις για το πεδίο, τόσο θεμελιώδεις που το TensorFlow της Google πήρε το όνομά τους από εκεί. Τι είναι λοιπόν ο τανυστής; Στον πυρήνα του, ένας τανυστής είναι ένα δοχείο για δεδομένα, σχεδόν πάντα αριθμητικά δεδομένα. Οι πίνακες (matrices) για παράδειγμα, είναι 2D τανυστές. Οι τανυστές είναι μια γενίκευση πινάκων σε έναν αυθαίρετο αριθμό διαστάσεων (στους τανυστές, μια διάσταση ονομάζεται συνήθως άξονας, axis). Ας δούμε πώς φτάσαμε στους τανυστές.

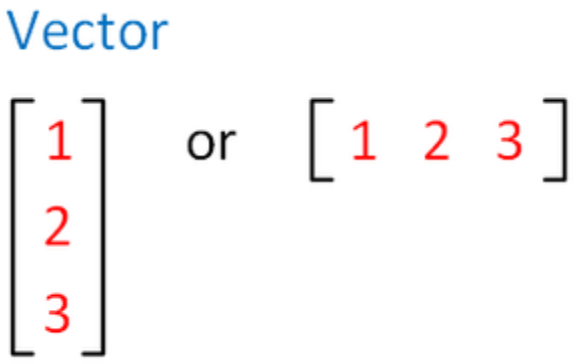
5.1.1 Κλιμακωτές, Διανύσματα, Πίνακες και Τανυστές (Scalars, Vectors, Matrices and Tensors)

- Scalars: Είναι απλώς ένας αριθμός



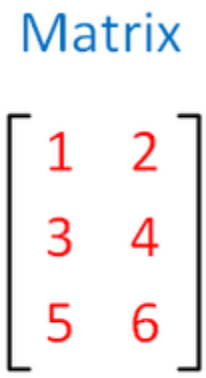
Σχήμα 5-2. Scalar

- Διανύσματα: Ένα διάνυσμα είναι ένας μονοδιάστατος πίνακας αριθμών. Οι αριθμοί είναι ταξινομημένοι σε σειρά. Μπορούμε να σκεφτούμε τα διανύσματα ως σημεία αναγνώρισης στο χώρο, με κάθε στοιχείο να δίνει τη συντεταγμένη κατά μήκος ενός διαφορετικού άξονα.



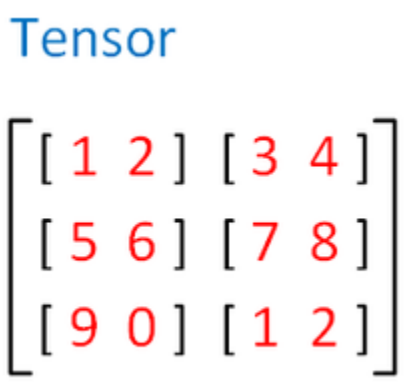
Σχήμα 5-3. Vector

- Πίνακες: Ένας πίνακας είναι ένα δισδιάστατο διάνυσμα αριθμών, επομένως κάθε στοιχείο προσδιορίζεται με δύο δείκτες αντί για έναν.



Σχήμα 5-4. Matrix

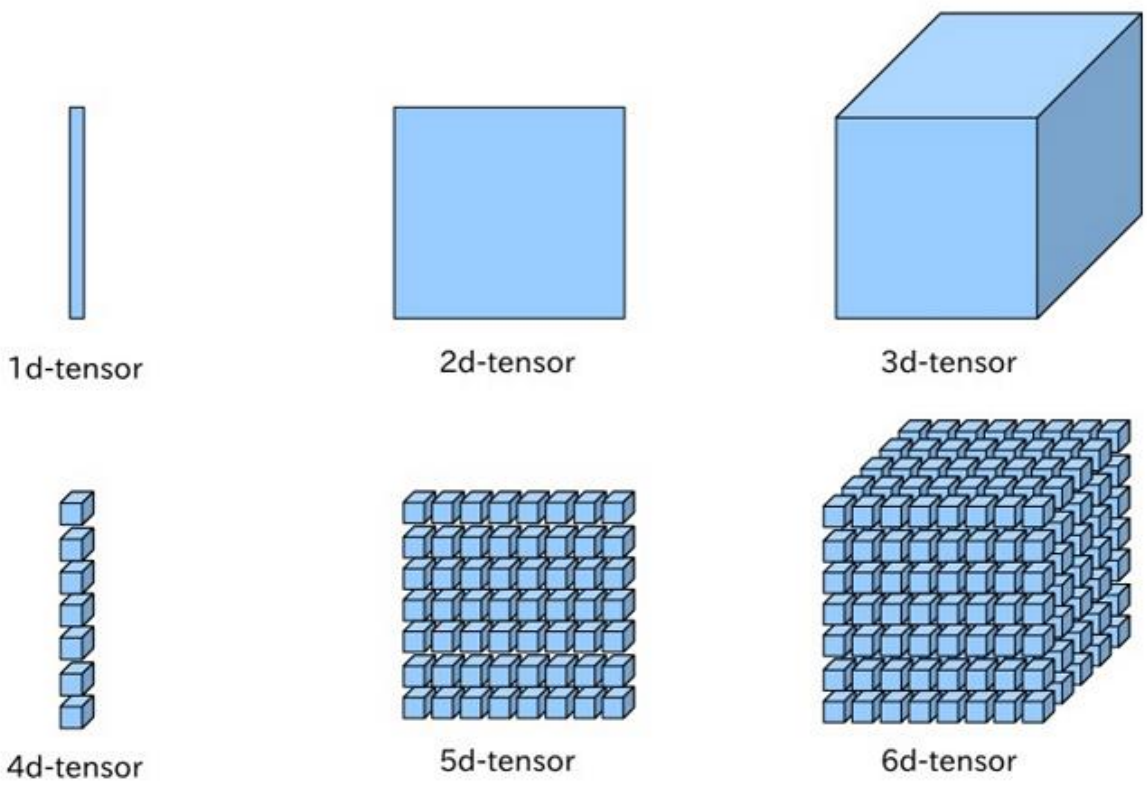
- Τανυστές: Σε ορισμένες περιπτώσεις θα χρειαστούμε έναν πίνακα με περισσότερους από δύο άξονες. Σε γενικές γραμμές, μια σειρά από αριθμούς που είναι διατεταγμένοι σε ένα κανονικό πλέγμα με μεταβλητό αριθμό αξόνων είναι γνωστός ως τανυστής.



Σχήμα 5-5. Tensor

Ένας τανυστής ορίζεται από τρία βασικά χαρακτηριστικά:

1. Αριθμός αξόνων (rank): Για παράδειγμα, ένας τρισδιάστατος τανυστής έχει τρεις άξονες και ένας πίνακας έχει δύο άξονες.
2. Σχήμα (shape): Πρόκειται για μια πλειάδα ακεραίων αριθμών που περιγράφει πόσες διαστάσεις έχει ο τανυστής κατά μήκος κάθε άξονα. Για παράδειγμα, το προηγούμενο παράδειγμα πίνακα έχει σχήμα (3, 2) και το παράδειγμα τρισδιάστατου τανυστή έχει σχήμα (6, 1, 2). Ένα διάνυσμα έχει ένα σχήμα με ένα μόνο στοιχείο, όπως το (3,), ενώ ένας βαθμωτός έχει ένα κενό σχήμα, ().
3. Τύπος δεδομένων (data type). Αυτός είναι ο τύπος των δεδομένων που περιέχεται στον τανυστή. Για παράδειγμα, ο τύπος ενός τανυστή θα μπορούσε να είναι float32, uint8, float64 και ούτω καθεξής. Σε σπάνιες περιπτώσεις, μπορεί να δούμε έναν τανυστή χαρακτήρων (char tensor).



Σχήμα 5-6. Διάφορα παραδείγματα διαστάσεων τανυστών
(Πηγή ιστοσελίδα: https://research.macrosynergy.com/analyzing-global-fixed-income-markets-with-tensors/)

5.1.2 Παραδείγματα Τανυστών με δεδομένα

Ας κάνουμε τους τανυστές δεδομένων πιο συγκεκριμένους με μερικά παραδείγματα:

- Διανυσματικά δεδομένα: τανυστές σχήματος 2D (δείγματα, χαρακτηριστικά) (samples, features)
- Δεδομένα χρονοσειράς ή δεδομένα ακολουθίας: τανυστές σχήματος 3D (δείγματα, χρονικά βήματα, χαρακτηριστικά) (samples, timesteps, features)
- Εικόνες: τανυστές σχήματος 4D (δείγματα, ύψος, πλάτος, κανάλια) (samples, height, width, channels)
- Βίντεο: τανυστές σχήματος 5D (δείγματα/παρτίδα, καρέ, ύψος, πλάτος, κανάλια) (samples, frames, height, width, channels)

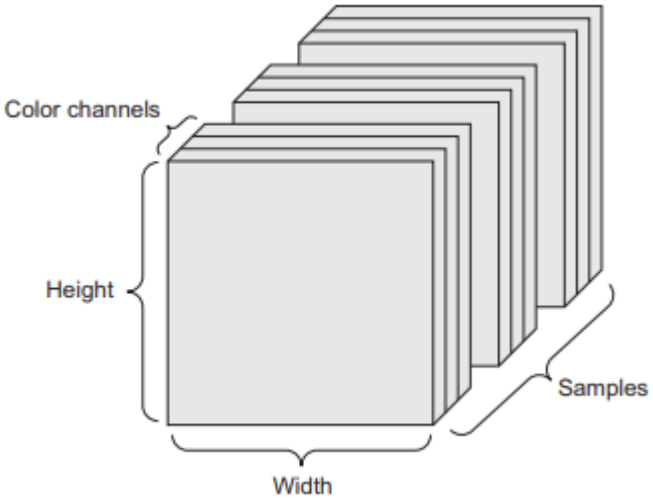
5.1.2.1 Διανυσματικά δεδομένα

Αυτή είναι η πιο συνηθισμένη περίπτωση. Σε ένα τέτοιο σύνολο δεδομένων, κάθε μεμονωμένο σημείο δεδομένων μπορεί να κωδικοποιηθεί ως διάνυσμα, και έτσι μια παρτίδα δεδομένων θα κωδικοποιηθεί ως δισδιάστατος τανυστής (δηλαδή, ένας πίνακας διανυσμάτων), όπου ο πρώτος άξονας είναι ο άξονας των δειγμάτων και ο δεύτερος ο άξονας είναι ο άξονας των χαρακτηριστικών. Ας ρίξουμε μια ματιά σε δύο παραδείγματα:

- Ένα σύνολο δεδομένων ατόμων, όπου λαμβάνουμε υπόψη την ηλικία, τον ταχυδρομικό κώδικα και το εισόδημα κάθε ατόμου. Κάθε άτομο μπορεί να χαρακτηριστεί ως ένα διάνυσμα 3 τιμών, και έτσι ένα ολόκληρο σύνολο δεδομένων 50.000 ατόμων μπορεί να αποθηκευτεί σε έναν δισδιάστατο τανυστή σχήματος (50.000, 3).
- Ένα σύνολο δεδομένων εγγράφων κειμένου, όπου αντιπροσωπεύουμε κάθε έγγραφο με τον αριθμό των φορών που εμφανίζεται κάθε λέξη σε αυτό (από ένα λεξικό 20.000 κοινών λέξεων). Κάθε έγγραφο μπορεί να κωδικοποιηθεί ως διάνυσμα 20.000 τιμών (μία μέτρηση ανά λέξη στο λεξικό) και έτσι ένα ολόκληρο σύνολο δεδομένων 500 εγγράφων μπορεί να αποθηκευτεί σε τανυστήρα σχήματος (500, 20.000).

5.1.2.2 Δεδομένα εικόνας

Οι εικόνες έχουν συνήθως τρεις διαστάσεις: ύψος, πλάτος και βάθος χρώματος. Αν και οι εικόνες σε κλίμακα του γκρι έχουν μόνο ένα κανάλι χρώματος και επομένως θα μπορούσαν να αποθηκευτούν σε τανυστές 2D, κατά σύμβαση οι τανυστές εικόνας είναι πάντα 3D, με ένα μονοδιάστατο κανάλι χρώματος για εικόνες σε κλίμακα του γκρι. Μια παρτίδα 128 εικόνων σε κλίμακα του γκρι μεγέθους 256 × 256 θα μπορούσε έτσι να αποθηκευτεί σε έναν τανυστή σχήματος (128, 256, 256, 1) και μια παρτίδα 128 έγχρωμων εικόνων θα μπορούσε να αποθηκευτεί σε έναν τανυστήρα σχήματος (128, 256, 256, 3) (βλ. εικόνα 5-7).



Σχήμα 5-7. Παράδειγμα δομή τανυστή εικόνας
(Πηγή: Deep Learning with Python, 2018)

5.2 Τα γρανάζια των νευρωνικών δικτύων: Λειτουργίες τανυστή

Κάθε πρόγραμμα υπολογιστή μπορεί τελικά να απλοποιηθεί σε ένα μικρό σύνολο δυαδικών πράξεων, σε δυαδικές εισόδους (AND, OR, NOR, και ούτω καθεξής), επίσης, όλοι οι μετασχηματισμοί που μαθαίνονται από τα βαθιά νευρωνικά δίκτυα μπορούν να απλοποιηθούν σε μια χούφτα πράξεων τανυστή που εφαρμόζονται σε τανυστές των αριθμητικών δεδομένων. Για παράδειγμα, είναι εφικτό να προσθέσουμε, να πολλαπλασιάσουμε τανυστές και ούτω καθεξής. Ένα επίπεδο ενός δικτύου μπορεί να ερμηνευθεί ως συνάρτηση, η οποία παίρνει ως είσοδο έναν τανυστή 2D και επιστρέφει έναν

άλλο 2D τανυστή, μια νέα αναπαράσταση για τον τανυστή εισόδου. Συγκεκριμένα, η λειτουργία είναι η εξής:

έξοδος = relu(dot(W, είσοδος) + b)

Σε αυτήν την έκφραση, τα W και b είναι τανυστές που είναι χαρακτηριστικά του επιπέδου. Ονομάζονται βάρη ή εκπαιδεύσιμες παράμετροι του επιπέδου. Αυτά τα βάρη περιέχουν τις πληροφορίες που μαθαίνει το δίκτυο από την έκθεση του σε δεδομένα εκπαίδευσης. Αρχικά, αυτά τα βάρη γεμίζονται με μικρές τυχαίες τιμές (ένα βήμα που ονομάζεται τυχαία προετοιμασία). Φυσικά, δεν υπάρχει λόγος να περιμένουμε ότι η έξοδος (relu[6](dot[7](W, input) + b)), όταν τα W και b είναι τυχαία, θα αποδώσουν χρήσιμες αναπαραστάσεις. Οι παραστάσεις που προκύπτουν δεν έχουν νόημα, αλλά είναι ένα σημείο εκκίνησης. Αυτό που ακολουθεί είναι η σταδιακή προσαρμογή αυτών των βαρών, με βάση ένα σήμα ανάδρασης. Αυτή η σταδιακή προσαρμογή, ονομάζεται εκπαίδευση, όπως έχουμε αναφέρει προηγουμένως. Αυτό συμβαίνει μέσα σε αυτό που ονομάζεται βρόχος εκπαίδευσης, ο οποίος λειτουργεί ως εξής:

1. Σχεδιάζουμε μια παρτίδα δειγμάτων εκπαίδευσης για x και τους αντίστοιχους στόχους y.
2. Εκτελούμε το δίκτυο στο x για να λάβουμε προβλέψεις y_pred.
3. Υπολογίζουμε την απώλεια του δικτύου στην παρτίδα, ένα μέτρο της αναντιστοιχίας μεταξύ y_pred και y.
4. Ενημερώνουμε όλα τα βάρη του δικτύου με τρόπο που να μειωθεί ελαφρώς η απώλεια σε αυτήν την παρτίδα.

Τελικά θα καταλήξουμε σε ένα δίκτυο που έχει πολύ χαμηλή απώλεια στα δεδομένα εκπαίδευσης: χαμηλή αναντιστοιχία μεταξύ των προβλέψεων y_pred και των αναμενόμενων στόχων y. Το δίκτυο έχει μάθει να χαρτογραφεί τις εισόδους του για να διορθώνει στόχους. Το δύσκολο μέρος είναι το βήμα 4: η ενημέρωση των βαρών του δικτύου. Μια καλή προσέγγιση είναι να εκμεταλλευτούμε το γεγονός ότι όλες οι λειτουργίες που χρησιμοποιούνται στο δίκτυο είναι διαφοροποιήσιμες και να υπολογίσουμε τη διαβάθμιση της απώλειας σε σχέση με τους συντελεστές του δικτύου. Στη συνέχεια, μπορούμε να μετακινήσουμε τους συντελεστές προς την αντίθετη κατεύθυνση από την κλίση, μειώνοντας έτσι την απώλεια.

6 Relu : Η πιο συνηθισμένη συνάρτηση ενεργοποίησης στην βαθιά μάθηση
7 Dot : Το εσωτερικό γινόμενο

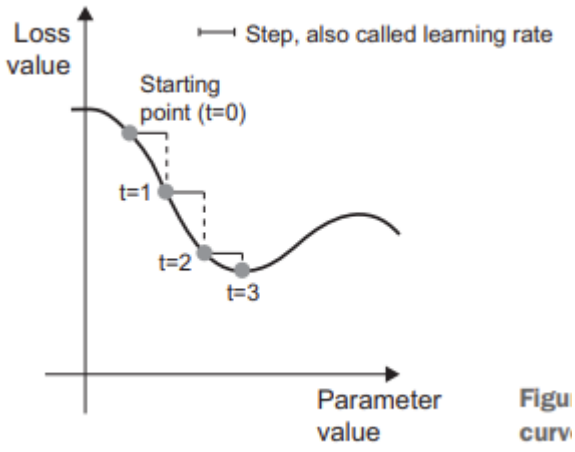
5.3 Στοχαστική κάθοδος κλίσης (Stochastic gradient descent)

Με χρήση μιας διαφοροποιήσιμης συνάρτησης, είναι θεωρητικά δυνατό να βρεθεί το ελάχιστο της. Αναλυτικότερα, είναι γνωστό ότι το ελάχιστο μιας συνάρτησης είναι ένα σημείο όπου η παράγωγος είναι 0, οπότε το μόνο που έχουμε να κάνουμε είναι να βρούμε όλα τα σημεία όπου η παράγωγος πηγαίνει στο 0 και να ελέγχουμε ποια από αυτά τα σημεία η συνάρτηση έχει τη χαμηλότερη τιμή. Εφαρμόζοντας το σε ένα νευρωνικό δίκτυο, αυτό σημαίνει ότι βρίσκουμε αναλυτικά τον συνδυασμό τιμών βάρους που αποδίδει τη μικρότερη δυνατή συνάρτηση απώλειας (loss function). Αυτό μπορεί να γίνει λύνοντας την κλίση της εξίσωσης (f)(W) = 0 για το W. Αυτή είναι μια πολυωνυμική εξίσωση N μεταβλητών, όπου N είναι ο αριθμός των συντελεστών στο δίκτυο. Αν και θα ήταν δυνατό να λυθεί μια τέτοια εξίσωση για N = 2 ή N = 3, αυτό είναι δύσκολο για τα νευρωνικά δίκτυα, όπου ο αριθμός

των παραμέτρων δεν είναι ποτέ μικρότερος από μερικές χιλιάδες και συχνά μπορεί να είναι αρκετές δεκάδες εκατομμύρια. Αντίθετα, μπορεί να χρησιμοποιηθεί ο αλγόριθμος τεσσάρων βημάτων που περιγράψαμε προηγουμένως, τροποποιώντας τις παραμέτρους σιγά σιγά με βάση την τρέχουσα τιμή απώλειας σε μια τυχαία παρτίδα δεδομένων. Επειδή έχουμε να κάνουμε με μια διαφοροποιήσιμη συνάρτηση, μπορούμε να υπολογίσουμε την κλίση της, η οποία μας δίνει έναν αποτελεσματικό τρόπο να εφαρμόσουμε το βήμα 4. Εάν ενημερώσουμε τα βάρη προς την αντίθετη κατεύθυνση από τη κλήση, η απώλεια θα είναι λίγο μικρότερη κάθε φορά:

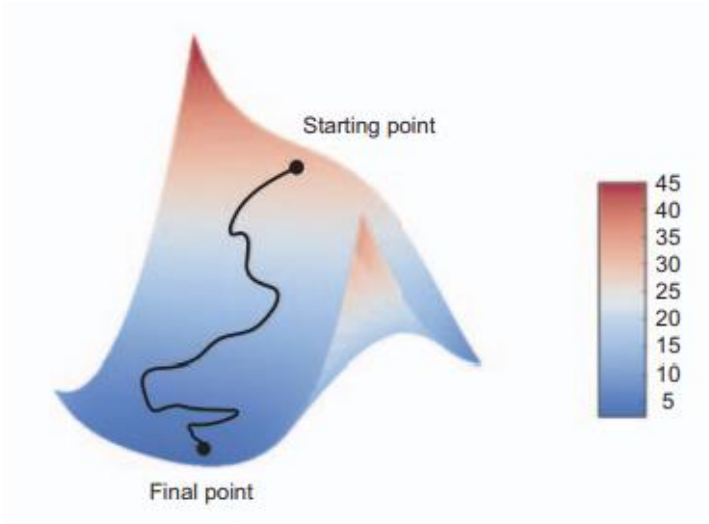
1. Σχεδιάζουμε μια παρτίδα δειγμάτων εκπαίδευσης για x και τους αντίστοιχους στόχους y.
2. Εκτελούμε το δίκτυο στο x για να λάβουμε προβλέψεις y_pred.
3. Υπολογίζουμε την απώλεια του δικτύου στην παρτίδα, ένα μέτρο της αναντιστοιχίας μεταξύ y_pred και y.
4. Υπολογίζουμε τη κλήση της απώλειας σε σχέση με τις παραμέτρους του δικτύου (ένα πέρασμα προς τα πίσω).
5. Μετακινούμε τις παραμέτρους λίγο προς την αντίθετη κατεύθυνση από τη κλήση, για παράδειγμα W -= βήμα* κλήση, μειώνοντας έτσι λίγο την απώλεια στην παρτίδα.

Τα βήματα που μόλις ακολουθήσαμε περιγράφουν την λειτουργία του mini-batch stochastic gradient descent (mini-batch SGD). Ο όρος στοχαστική αναφέρεται στο γεγονός ότι κάθε παρτίδα δεδομένων αντλείται τυχαία.



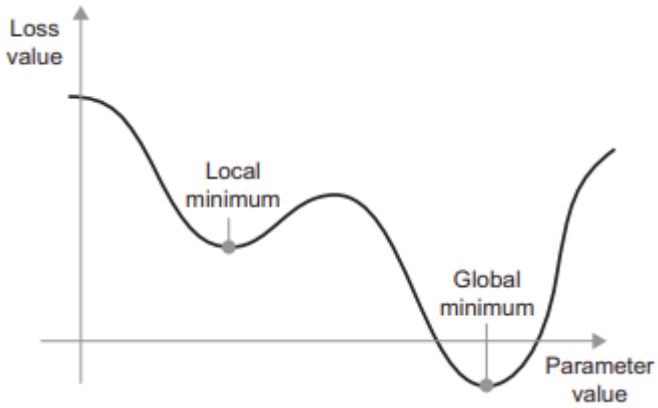
Σχήμα 5-8. Gradient Descent σε 1 διάσταση με μόνο μια παράμετρο και δείγμα εκπαίδευσης
(Πηγή: Deep Learning with Python, 2018)

Όπως μπορούμε να δούμε, είναι σημαντικό να επιλέξουμε μια λογική τιμή για τον παράγοντα βήμα. Εάν είναι πολύ μικρό, η κάθοδος στην καμπύλη θα διαρκέσει πολλές επαναλήψεις και υπάρχει πιθανότητα να κολλήσουμε σε ένα τοπικό ελάχιστο. Εάν το βήμα είναι πολύ μεγάλο, οι ενημερώσεις μπορεί να καταλήξουν να μας οδηγήσουν σε εντελώς τυχαίες τοποθεσίες στην καμπύλη. Παρόλο που το σχήμα 5-8 απεικονίζει την κάθοδο κλίσης σε ένα χώρο παραμέτρων 1D, στην πράξη θα χρησιμοποιούσαμε την κλίση σε χώρους με πολλές διαστάσεις. Κάθε συντελεστής βάρους σε ένα νευρωνικό δίκτυο είναι μια ελεύθερη διάσταση στο χώρο και μπορεί να υπάρχουν δεκάδες χιλιάδες ή και εκατομμύρια από αυτές. Διαισθητικά μπορούμε να δούμε την κάθοδο της κλίσης κατά μήκος μιας επιφάνειας απώλειας 2D, όπως φαίνεται στο σχήμα 5-9.



Σχήμα 5-9. Gradient Descent σε 2 διαστάσεις, μαζί με μια κλίμακα ύψους
(Πηγή: Deep Learning with Python, 2018)

Αλλά δεν μπορούμε να φανταστούμε πώς φαίνεται η πραγματική διαδικασία εκπαίδευσης ενός νευρωνικού δικτύου, δεν μπορεί να αναπαρασταθεί ένας χώρος 1.000.000 διαστάσεων με τρόπο που να έχει νόημα στον άνθρωπο. Ως εκ τούτου, είναι καλό να έχουμε κατά νου ότι οι διαισθήσεις που αναπτύσσουμε μέσω αυτών των χαμηλών διαστάσεων αναπαραστάσεων μπορεί να μην είναι πάντα ακριβείς στην πράξη. Αυτό ήταν ιστορικά μια πηγή ζητημάτων στον κόσμο της έρευνας της βαθιάς μάθησης. Επιπλέον, υπάρχουν πολλές παραλλαγές του SGD που διαφέρουν λαμβάνοντας υπόψη προηγούμενες ενημερώσεις βάρους κατά τον υπολογισμό της επόμενης ενημέρωσης βάρους, αντί να εξετάζουμε απλώς την τρέχουσα τιμή των κλίσεων. Υπάρχει, για παράδειγμα, το SGD με ορμή (momentum), καθώς και ο Adagrad, το RMSProp και πολλά άλλα. Τέτοιες παραλλαγές είναι γνωστές ως μέθοδοι βελτιστοποίησης ή βελτιστοποιητές (optimizers). Συγκεκριμένα, η έννοια της ορμής, η οποία χρησιμοποιείται σε πολλές από αυτές τις παραλλαγές, αξίζει την προσοχή μας. Η Ορμή (Momentum) αντιμετωπίζει δύο ζητήματα με το SGD: την ταχύτητα σύγκλισης και τα τοπικά ελάχιστα.



Σχήμα 5-10. Καμπύλη μιας παραμέτρου ως συνάρτηση μιας απώλειας του δικτύου
(Πηγή: Deep Learning with Python, 2018)

Όπως μπορούμε να δούμε στο σχήμα 5-10, γύρω από μια συγκεκριμένη τιμή παραμέτρου, υπάρχει ένα τοπικό ελάχιστο, γύρω από αυτό το σημείο, η μετακίνηση προς τα αριστερά θα είχε ως αποτέλεσμα την αύξηση της απώλειας, αλλά και τη μετακίνηση προς τα δεξιά. Εάν η υπό εξέταση παράμετρος βελτιστοποιούνταν μέσω SGD με μικρό ρυθμό εκμάθησης, τότε η

διαδικασία βελτιστοποίησης θα κολλούσε στο τοπικό ελάχιστο αντί να φτάσει στο συνολικό ελάχιστο. Μπορούμε να αποφύγουμε τέτοια ζητήματα χρησιμοποιώντας την ορμή, η οποία αντλεί έμπνευση από τη φυσική. Μια χρήσιμη νοητική εικόνα εδώ είναι να σκεφτούμε τη διαδικασία βελτιστοποίησης ως μια μικρή μπάλα που κυλά στην καμπύλη απώλειας. Εάν έχει αρκετή ορμή, η μπάλα δεν θα κολλήσει στην χαράδρα και θα καταλήξει στο ολικό ελάχιστο. Η ορμή υλοποιείται μετακινώντας την μπάλα σε κάθε βήμα με βάση όχι μόνο την τρέχουσα τιμή κλίσης (τρέχουσα επιτάχυνση) αλλά και την τρέχουσα ταχύτητα (που προκύπτει από προηγούμενη επιτάχυνση). Στην πράξη, αυτό σημαίνει ενημέρωση της παραμέτρου w με βάση όχι μόνο την τρέχουσα τιμή κλίσης αλλά και την προηγούμενη ενημέρωση παραμέτρου.

5.4 Ο αλγόριθμος Backpropagation

Στον προηγούμενο αλγόριθμο, υποθέσαμε τυχαία ότι επειδή μια συνάρτηση είναι διαφορίσιμη, μπορούμε να υπολογίσουμε ρητά την παράγωγη της. Στην πράξη, μια συνάρτηση νευρωνικού δικτύου αποτελείται από πολλές λειτουργίες τανυστών αλυσιδωτές μεταξύ τους, καθεμία από τις οποίες έχει μια απλή, γνωστή παράγωγος.

Ο λογισμός μας λέει ότι μια αλυσίδα συναρτήσεων μπορεί να εξαχθεί χρησιμοποιώντας την ακόλουθη ταυτότητα, που ονομάζεται κανόνας της αλυσίδας: f(g(x)) = f'(g(x)) * g'(x). Η εφαρμογή του κανόνα της αλυσίδας στον υπολογισμό των τιμών κλίσης ενός νευρωνικού δικτύου οδηγεί σε έναν αλγόριθμο που ονομάζεται Backpropagation (που μερικές φορές ονομάζεται επίσης διαφοροποίηση αντίστροφης λειτουργίας). Ο Backpropagation ξεκινά με την τελική τιμή απώλειας και λειτουργεί προς τα πίσω από τα πολύπλοκα στρώματα στα κάτω στρώματα, εφαρμόζοντας τον κανόνα της αλυσίδας για να υπολογιστεί η συμβολή που είχε κάθε παράμετρος στην τιμή απώλειας.

5.5 Συναρτήσεις απώλειας (loss functions) και βελτιστοποιητές (optimizers)

Αφού οριστεί η αρχιτεκτονική του δικτύου, θα πρέπει να επιλεχθούν ακόμα δύο πράγματα:

- Συνάρτηση απώλειας: Η ποσότητα που θα ελαχιστοποιηθεί κατά την εκπαίδευση. Αντιπροσωπεύει ένα μέτρο επιτυχίας για την εκάστοτε εργασία.
- Βελτιστοποιητής: Καθορίζει τον τρόπο ενημέρωσης του δικτύου με βάση τη συνάρτηση απώλειας. Εφαρμόζει μια συγκεκριμένη παραλλαγή της στοχαστικής κάθοδο κλίσης (SGD).

Ένα νευρωνικό δίκτυο που έχει πολλαπλές εξόδους μπορεί να έχει πολλαπλές συναρτήσεις απώλειας (μία ανά έξοδο). Αλλά η διαδικασία gradient-descent πρέπει να βασίζεται σε μια ενιαία κλιμακωτή τιμή απώλειας, επομένως, για δίκτυα πολλαπλών απωλειών, όλες οι απώλειες συνδυάζονται (μέσω του μέσου όρου) σε μια ενιαία βαθμωτή ποσότητα. Η επιλογή της σωστής αντικειμενικής συνάρτησης για το σωστό πρόβλημα είναι εξαιρετικά σημαντική, το δίκτυό θα προσπαθήσει να κάνει όποια συντόμευση μπορεί, για να ελαχιστοποιήσει την απώλεια του, οπότε εάν ο στόχος δεν συσχετίζεται πλήρως με την επιτυχία για την εργασία που θέλουμε, το δίκτυο θα καταλήξει να κάνει πράγματα που μπορεί να μην θέλαμε.

5.6 Αξιολόγηση μοντέλων μηχανικής μάθησης

Στη μηχανική μάθηση, ο στόχος είναι να επιτευχθούν μοντέλα που γενικεύουν (generalize), που αποδίδουν καλά σε δεδομένα που δεν έχουν δει ποτέ και η υπερ-προσαρμογή (overfitting) τους είναι το κεντρικό εμπόδιο. Το δίκτυο μπορεί να ελέγξει μόνο αυτό που μπορεί να παρατηρήσει, επομένως είναι κρίσιμο να μπορούμε να υπολογίσουμε αξιόπιστα τη δύναμη γενίκευσης του μοντέλου μας. Η υπερ-προσαρμογή (overfitting) συμβαίνει σε κάθε πρόβλημα μηχανικής μάθησης. Η εκμάθηση του τρόπου αντιμετώπισης της υπερβολικής προσαρμογής είναι απαραίτητη για την εξοικείωση με τη μηχανική εκμάθηση. Η θεμελιώδης έκβαση στη μηχανική μάθηση είναι το μέτρο μεταξύ βελτιστοποίησης και γενίκευσης. Η βελτιστοποίηση αποδίδεται στη διαδικασία προσαρμογής ενός μοντέλου για να έχει την καλύτερη δυνατή απόδοση στα δεδομένα εκπαίδευσης, ενώ η γενίκευση ορίζεται στο πόσο καλά αποδίδει το εκπαιδευμένο μοντέλο σε δεδομένα που δεν έχει ξαναδεί. Στην αρχή της εκπαίδευσης, η βελτιστοποίηση και η γενίκευση συσχετίζονται: όσο μικρότερη είναι η απώλεια δεδομένων εκπαίδευσης, τόσο μικρότερη είναι η απώλεια στα δεδομένα δοκιμής. Ενώ συμβαίνει αυτό, το μοντέλο θεωρείται ότι δεν είναι ακόμα κατάλληλο, υπάρχει ακόμη πρόοδος που πρέπει να γίνει, το δίκτυο δεν έχει ακόμη μοντελοποιήσει όλα τα σχετικά μοτίβα στα δεδομένα εκπαίδευσης. Αλλά μετά από έναν ορισμένο αριθμό επαναλήψεων στα δεδομένα εκπαίδευσης, η γενίκευση σταματά να βελτιώνεται και οι μετρήσεις επικύρωσης σταματούν και στη συνέχεια αρχίζουν να υποβαθμίζονται, το μοντέλο αρχίζει να υπερ-προσαρμόζεται (overfitting). Δηλαδή, αρχίζει να μαθαίνει μοτίβα που είναι ειδικά για τα δεδομένα εκπαίδευσης, αλλά που είναι παραπλανητικά ή άσχετα όταν πρόκειται για νέα δεδομένα. Για να αποτρέψουμε ένα μοντέλο να μάθει παραπλανητικά ή άσχετα μοτίβα που υποδεικνύονται στα δεδομένα εκπαίδευσης, η καλύτερη λύση είναι να λάβουμε περισσότερα δεδομένα εκπαίδευσης. Ένα μοντέλο που εκπαιδεύεται σε περισσότερα δεδομένα θα γενικεύεται πολύ καλύτερα. Η λειτουργία της καταπολέμησης της υπερ-προσαρμογής ονομάζεται regularization.

Ας δούμε εν συντομία μερικές από τις πιο κοινές τεχνικές regularization.

- Μείωση του μεγέθους του δικτύου

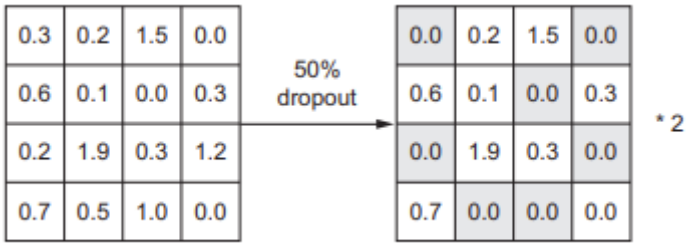
Ο ευκολότερος τρόπος για να αποφευχθεί η υπερ-προσαρμογή είναι να ελαχιστοποιηθεί το μέγεθος του μοντέλου, ο αριθμός δηλαδή των παραμέτρων που μπορούν να μαθευτούν στο μοντέλο (ο οποίος αποδίδεται από τον αριθμό των επιπέδων και τον αριθμό των μονάδων ανά επίπεδο). Η γενική ροή εργασίας για την εύρεση του κατάλληλου μεγέθους μοντέλου είναι να ξεκινάμε με σχετικά λίγα επίπεδα και παραμέτρους και να αυξάνουμε το μέγεθος των επιπέδων ή να προσθέτουμε νέα επίπεδα μέχρι να δούμε μειωμένες αποδόσεις όσον αφορά την απώλεια επικύρωσης.

- Προσθήκη κόστους βάρους (Adding weight regularization)

Τα πιο απλά μοντέλα είναι λιγότερο πιθανό να υπερ-προσαρμοστούν από τα σύνθετα. Ένα απλό μοντέλο θεωρείται, ένα μοντέλο όπου η κατανομή των τιμών των παραμέτρων έχει λιγότερη εντροπία (ή ένα μοντέλο με λιγότερες παραμέτρους). Έτσι, ένας συνηθισμένος τρόπος για να αποφευχθεί η υπερ-προσαρμογή είναι να οριστούν περιορισμοί στην πολυπλοκότητα ενός δικτύου αναγκάζοντας τα βάρη του να δεχτούν μόνο μικρές τιμές, γεγονός που καθιστά την κατανομή των τιμών βάρους πιο κανονική. Αυτό ονομάζεται weight regularization και ολοκληρώνεται προσθέτοντας ένα κόστος που σχετίζεται με την ύπαρξη μεγάλων βαρών, στη συνάρτηση απώλειας του δικτύου.

- Προσθήκη Dropout

Το Dropout είναι μια από τις πιο αποτελεσματικές και πιο συχνά χρησιμοποιούμενες τεχνικές regularization για νευρωνικά δίκτυα, που προτάθηκε από τον Geoff Hinton και τους μαθητές του στο Πανεπιστήμιο του Τορόντο. Το Dropout, που εφαρμόζεται σε ένα επίπεδο, συνίσταται στην τυχαία εγκατάλειψη (ρύθμιση στο μηδέν) ορισμένων χαρακτηριστικών εξόδου του στρώματος κατά τη διάρκεια της εκπαίδευσης. Ας υποθέσουμε ότι ένα δεδομένο επίπεδο διανύσματος [0,2, 0,5, 1,3, 0,8, 1,1] θα επέστρεφε κανονικά για ένα δεδομένο δείγμα εισόδου κατά τη διάρκεια της εκπαίδευσης. Μετά την εφαρμογή του Dropout, αυτό το διάνυσμα θα έχει μερικές μηδενικές τιμές κατανεμημένες τυχαία, για παράδειγμα, [0, 0.5, 1.3, 0, 1.1]. Το ποσοστό εγκατάλειψης είναι το κλάσμα των χαρακτηριστικών που μηδενίζονται, συνήθως ορίζεται μεταξύ 0,2 και 0,5. Κατά το χρόνο δοκιμής, καμία μονάδα δεν εγκαταλείπεται, αντίθετα, οι τιμές εξόδου του επιπέδου ελαχιστοποιούνται κατά έναν παράγοντα ίσο με το ποσοστό εγκατάλειψης, για να εξισορροπηθεί το γεγονός ότι περισσότερες μονάδες είναι ενεργές από ότι κατά τον χρόνο εκπαίδευσης.



Σχήμα 5-11. Η εφαρμογή του Dropout και η αναπροσαρμογή του κατά την εκπαίδευση
(Πηγή: Deep Learning with Python, 2018)

Αυτή η τεχνική μπορεί να φαίνεται περίεργη και αυθαίρετη. Όμως, η βασική ιδέα είναι ότι η εισαγωγή θορύβου στις τιμές εξόδου ενός στρώματος καταφέρνει να διαλύσει μοτίβα τυχαίας κατάστασης που δεν είναι σημαντικά, τα οποία το δίκτυο θα αρχίσει να απομνημονεύει εάν δεν υπήρχε θόρυβος.

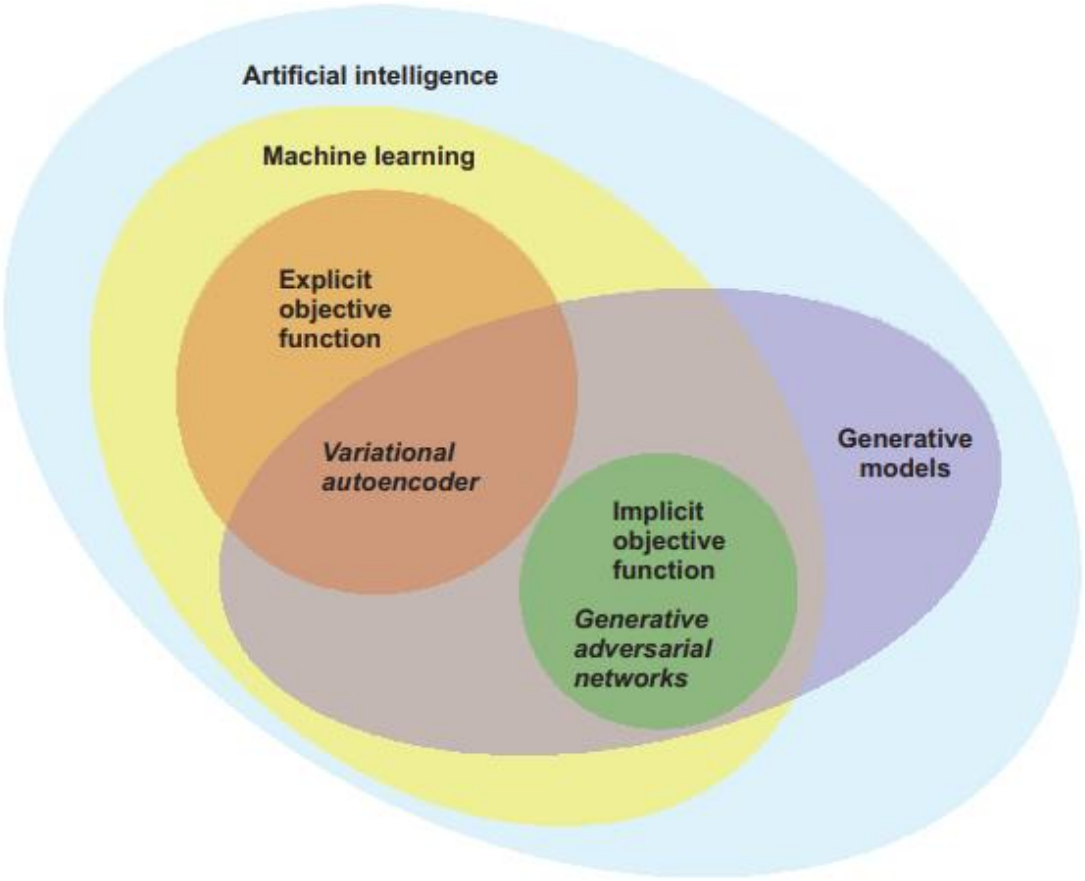
5.7 Σύνοψη Βαθιών Νευρωνικών Δικτύων

Τα Βαθιά Νευρωνικά Δίκτυα είναι ένα υποπεδίο της τεχνητής νοημοσύνης, τα οποία ξαναβγήκαν στην επιφάνεια στις αρχές του 21ου αιώνα ύστερα από κάποιες πρωτοποριακές ανακαλύψεις στον τομέα των hardware, συνόλων δεδομένων και αλγορίθμων, οι οποίες μέχρι τότε περιόριζαν την χρήση τους. Η επανάσταση αυτή έφερε στο φώς έναν πρωτοποριακό τρόπο εκμάθησης, ο οποίος με πολλά εκατοντάδες στοιβαγμένα επίπεδα που αλληλοσυνδέονται καταφέρνουν να εξάγουν σημαντικές πληροφορίες από τα μοτίβα των δεδομένων τους. Ένα πρόβλημα μηχανικής μάθησης για να λυθεί χρειάζεται να ακολουθηθούν κάποια συγκεκριμένα βήματα. Αρχικά, πρέπει να ορίσουμε το πρόβλημα, δηλαδή ποια θα είναι τα δεδομένα εισόδου και τι προσπαθούμε να προβλέψουμε. Κάνοντας υποθέσεις ότι τα διαθέσιμα δεδομένα είναι επαρκώς ενημερωτικά για να μάθουν την συσχέτιση μεταξύ εισόδων και εξόδων, μας βοηθάει ώστε να αντιληφθούμε καλύτερα το πρόβλημα. Στην συνέχεια, απαιτείται η προετοιμασία των δεδομένων εισόδου σε κατάλληλη μορφή για να εισαχθούν στο μοντέλο μας. Τέτοιες προετοιμασίες, είναι η μορφοποίηση των δεδομένων ως τανυστές, η κλιμάκωση τους σε μικρές τιμές και άλλες μεθοδολογίες αναλόγως την μορφή των δεδομένων μας. Υποθέτοντας ότι το μοντέλο λειτουργεί σωστά, χρειάζεται να κάνουμε κάποιες βασικές επιλογές ώστε να το ολοκληρώσουμε. Πρέπει να επιλέξουμε την ενεργοποίηση του τελευταίου επιπέδου, την συνάρτηση απώλειας, τον

βελτιστοποιήτη και τον μετρητή αξιολόγησης οι οποίοι είναι ανάλογοι με βάση το πρόβλημα που επιδιώκουμε να λύσουμε. Εδώ απαιτείται ιδιαίτερη προσοχή, ώστε να παρατηρήσουμε εάν το μοντέλο μας γενικεύει αρκετά ή υπερ-προσαρμόζεται. Εάν υπερ-προσαρμόζεται, υπάρχουν διάφορες τεχνικές ώστε να συντονίσουμε το δίκτυο. Μερικές από αυτές τις τεχνικές είναι η μέθοδος Dropout, η προσθαφαίρεση επιπέδων, η δοκιμή διαφορετικών αριθμών μονάδων, η αλλαγή στον ρυθμό εκμάθησης του βελτιστοποιητή και πολλά άλλα. Τέλος, αφού αναπτύξουμε μια ικανοποιητική διαμόρφωση μοντέλου, μπορούμε να εκπαιδεύσουμε το τελικό μοντέλο παραγωγής σε όλα τα διαθέσιμα δεδομένα, να εξάγουμε και να αξιολογήσουμε τα τελικά αποτελέσματα. Εάν είμαστε ικανοποιημένοι, μπορούμε να χρησιμοποιήσουμε το ίδιο τελικό μοντέλο ή με μερικές παραλλαγές σε παρόμοια προβλήματα μηχανικής μάθησης.

Κλείνοντας, ξεκινήσαμε με μια απλή έννοια των βαθιών νευρωνικών δικτύων με τύπους στατιστικών μαθηματικών και αλγορίθμων, αλλά καταλήξαμε σε πολύπλοκα μοντέλα που χρησιμοποιούν πολλές παραμέτρους με εκατοντάδες στρώματα για να εκπαιδευτούν. Παράλληλα, ανακαλύψαμε ότι υπάρχουν πολλές παράμετροι που πρέπει να συντονίσουμε και να προσέξουμε, προκειμένου το μοντέλο μας να είναι πετυχημένο. Αλλά μόλις φτάσουμε σε ένα αξιόλογο αποτέλεσμα, τα μοντέλα αυτά μπορούν να κάνουν θαύματα!

6. Variational Autoencoders



Σχήμα 6-1. Τοποθετώντας τα VAEs και τα GANs στο πεδίο της τεχνητής νοημοσύνης
(Πηγή: GANs in Action Deep Learning with Generative Adversarial Networks, 2019)

6.1 Δειγματοληψία από λανθάνοντες χώρους (latent space)

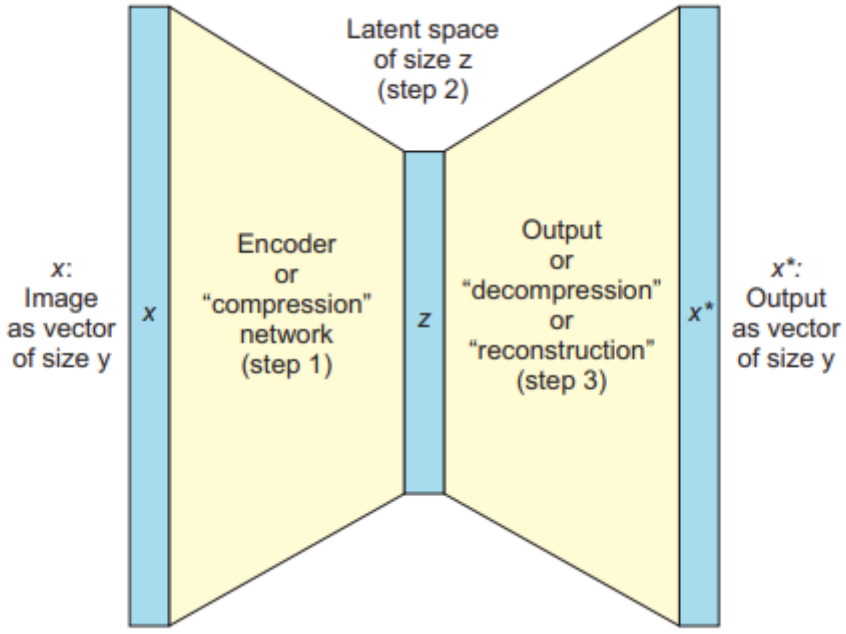
Τα δεδομένα εισόδου είναι συχνά σε υψηλές διαστάσεις. Αυτό θέτει προκλήσεις όχι μόνο για την υπολογιστική αποδοτικότητα αλλά κυρίως επειδή δυσκολεύει τη μοντελοποίηση της αναπαράστασης. Ο λανθάνοντας χώρος, λύνει αυτό το πρόβλημα, καθώς αναπαριστά τα συμπιεσμένα δεδομένα και τα χαρακτηριστικά τους, σε μια πιο απλή και χρήσιμη διάσταση την οποία μπορούμε να την ερμηνεύσουμε ευκολότερα και αποδοτικότερα. Αναλυτικότερα, η βασική ιδέα της δημιουργίας δεδομένων είναι η ανάπτυξη ενός χαμηλών διαστάσεων λανθάνοντος χώρου αναπαραστάσεων (ένας διανυσματικός χώρος) όπου οποιοδήποτε σημείο μπορεί να αντιστοιχιστεί σε ένα ρεαλιστικό δεδομένο. Το δομοστοιχείο που μπορεί να πραγματοποιήσει αυτήν την αντιστοίχιση, λαμβάνοντας ως είσοδο ένα λανθάνον σημείο και βγάζοντας ένα δεδομένο, ονομάζεται αποκωδικοποιητής, Decoder. Μόλις αναπτυχθεί ένας τέτοιος λανθάνοντας χώρος, μπορούμε να δοκιμάσουμε σημεία από αυτόν, είτε σκόπιμα είτε τυχαία, και, αντιστοιχίζοντας τα στο χώρο του δεδομένου, να δημιουργήσουμε δεδομένα που δεν έχουμε ξαναδεί ποτέ πριν. Τα VAE είναι εξαιρετικά για την εκμάθηση λανθάνοντων χώρων που είναι καλά δομημένοι, όπου συγκεκριμένες κατευθύνσεις κωδικοποιούν έναν ουσιαστικό άξονα διακύμανσης στα δεδομένα.

6.2 Από τι αποτελείται ο Αυτοκωδικοποιητής (Αυτόματος κωδικοποιητής - Autoencoder);

Οι αυτοκωδικοποιητές είναι μια αρκετά παλιά ιδέα, τουλάχιστον όταν εξετάζουμε την εποχή της μηχανικής μάθησης ως πεδίο. Ακολουθώντας την τάση για βαθιά μάθηση και εφαρμόζοντας την στους αυτοκωδικοποιητές παρατηρούμε ότι αποτελούνται από δύο νευρωνικά δίκτυα:

- Ένα δίκτυο κωδικοποιητών που συμπιέζει δεδομένα εισόδου υψηλών διαστάσεων σε ένα διάνυσμα αναπαράστασης χαμηλότερων διαστάσεων (η δουλειά του κωδικοποιητή είναι να πάρει τα δεδομένα εισόδου και να την αντιστοιχίσει σε ένα σημείο του λανθάνοντος χώρου)
- Ένα δίκτυο αποκωδικοποιητή που αποσυμπιέζει ένα δεδομένο διάνυσμα αναπαράστασης πίσω στον αρχικό τομέα

Το δίκτυο είναι εκπαιδευμένο να βρίσκει βάρη για τον κωδικοποιητή και τον αποκωδικοποιητή που ελαχιστοποιούν την απώλεια μεταξύ της αρχικής εισόδου και της ανακατασκευής της εισόδου αφού περάσει από τον κωδικοποιητή και τον αποκωδικοποιητή. Το διάνυσμα αναπαράστασης είναι μια συμπίεση της αρχικής εικόνας σε ένα χαμηλότερης διάστασης, λανθάνον χώρο. Η ιδέα είναι ότι επιλέγοντας οποιοδήποτε σημείο στον λανθάνοντα χώρο, θα πρέπει να είμαστε σε θέση να δημιουργήσουμε νέα δεδομένα περνώντας αυτό το σημείο μέσα από τον αποκωδικοποιητή, καθώς ο αποκωδικοποιητής έχει μάθει πώς να μετατρέπει σημεία στον λανθάνοντα χώρο σε δεδομένα. Στην πράξη, οι αυτόματοι κωδικοποιητές έχουν συνήθως περισσότερες από δύο διαστάσεις, προκειμένου να έχουν μεγαλύτερη ελευθερία να καταγράφουν μεγαλύτερες λεπτομέρειες στα δεδομένα.

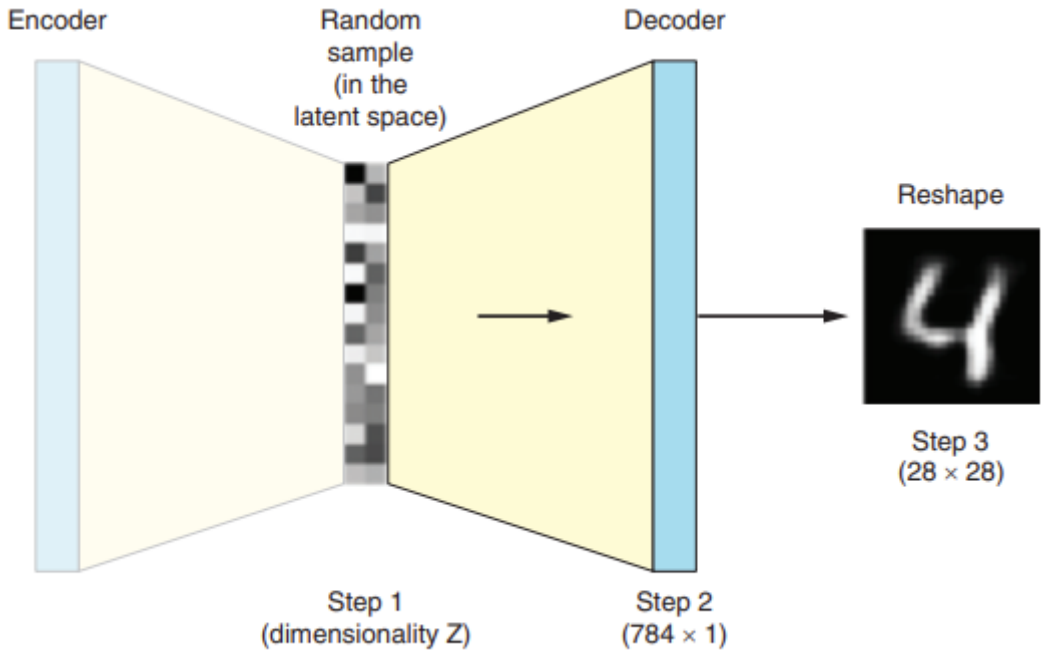


Σχήμα 6-2. Παράδειγμα αυτοκωδικοποιητή
(Πηγή: GANs in Action Deep Learning with Generative Adversarial Networks, 2019)

Παράδειγμα εκπαίδευσης αυτοκωδικοποιητή:

1. Λαμβάνουμε δεδομένα χ από τη βάση δεδομένων μας, τα οποία τροφοδοτούμε στο κωδικοποιητή. Αυτός στην συνέχεια συμπιέζει τα δεδομένα εισόδου, ώστε να τα αντιστοιχίσει σε ένα σημείο του λανθάνοντος χώρου.
2. Ο αποκωδικοποιητής αποσυμπιέζει και ανακατασκευάζει τα δεδομένα του λανθάνοντα χώρου πίσω στον αρχικό τομέα ως χ*.
3. Μετράμε την απώλεια ανακατασκευής (reconstruction loss), τη διαφορά μεταξύ x και x*. Αυτό γίνεται με χρήση μιας απόστασης (για παράδειγμα, με το μέσου σφάλματος) μεταξύ των pixel των x και x*. Αυτό μας δίνει μια αντικειμενική συνάρτηση (|| x –x* ||) για βελτιστοποίηση μέσω μιας έκδοσης του gradient descent.

Προσπαθούμε λοιπόν να βρούμε τις παραμέτρους του κωδικοποιητή και του αποκωδικοποιητή που θα ελαχιστοποιήσουν την απώλεια ανακατασκευής που ενημερώνουμε χρησιμοποιώντας gradient descent.



Σχήμα 6-3. Ακόμη ένα παράδειγμα αυτοκωδικοποιητή αναπαραστώντας καλύτερα τον λανθάνοντα χώρο (Πηγή: GANs in Action Deep Learning with Generative Adversarial Networks, 2019)

6.3 Χρήση αυτόματων κωδικοποιητών

Παρά την απλότητά τους, υπάρχουν πολλοί λόγοι για να ενδιαφερόμαστε για τους αυτόματους κωδικοποιητές:

- Πρώτα απ 'όλα, παίρνουμε συμπίεση. Αυτό συμβαίνει επειδή το ενδιάμεσο βήμα (1) από το σχήμα 6-3 γίνεται μια μειωμένη εικόνα ή αντικείμενο στη διάσταση του λανθάνοντος χώρου. Προφανώς δεν γίνεται χωρίς απώλειες πληροφορίας, αλλά είμαστε ελεύθεροι να χρησιμοποιήσουμε αυτήν την παρενέργεια, εάν επιθυμούμε.
- Ακόμα χρησιμοποιώντας τον λανθάνοντα χώρο, μπορούμε να σκεφτούμε πολλές πρακτικές εφαρμογές, όπως έναν ταξινομητή μιας κατηγορίας, όπου μπορούμε να δούμε τα στοιχεία σε έναν μειωμένο, πιο γρήγορα αναζητήσιμο λανθάνοντα χώρο για

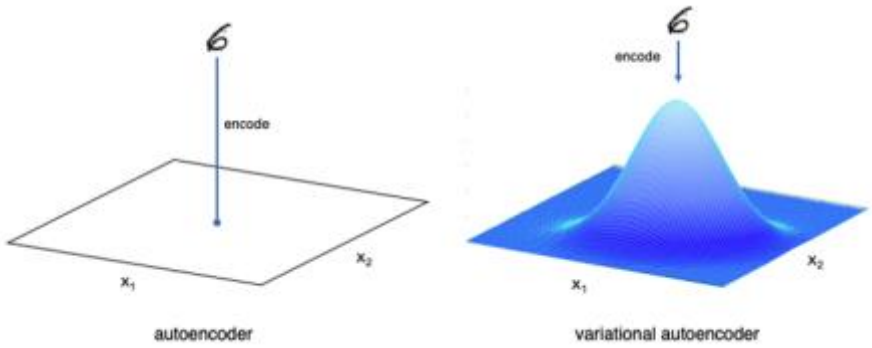
να ελέγξουμε για ομοιότητα με την κατηγορία στόχο. Αυτό μπορεί να λειτουργήσει σε ρυθμίσεις αναζήτησης (ανάκτηση πληροφοριών) ή ανίχνευσης ανωμαλιών (συγκρίνοντας την εγγύτητα στον λανθάνοντα χώρο).

- Μια άλλη περίπτωση χρήσης είναι η διαγραφή δεδομένων ή ο χρωματισμός ασπρόμαυρων εικόνων. Για παράδειγμα, εάν έχουμε μια παλιά θορυβώδες φωτογραφία ή ένα βίντεο, του Β' Παγκοσμίου Πολέμου, μπορούμε να τις κάνουμε λιγότερο θορυβώδεις και να προσθέσουμε ξανά χρώμα. Εξ ου και η ομοιότητα με τα GAN, τα οποία επίσης τείνουν να υπερέχουν σε αυτούς τους τύπους των εφαρμογών.
- Ορισμένες αρχιτεκτονικές GAN, όπως το BEGAN (αναλυτικότερα στο Κεφ. 8.1), χρησιμοποιούν αυτοκωδικοποιητές ως μέρος της αρχιτεκτονικής τους για να τους βοηθήσουν να σταθεροποιήσουν την εκπαίδευσή τους, κάτι που είναι εξαιρετικά σημαντικό.
- Η εκπαίδευση των αυτόματων κωδικοποιητών δεν απαιτεί δεδομένα με ετικέτα.
- Μπορούμε να χρησιμοποιήσουμε αυτοκωδικοποιητές για τη δημιουργία νέων εικόνων. Οι αυτοκωδικοποιητές έχουν εφαρμοστεί σε όλα τα είδη εικόνας, αλλά συνήθως όσο υψηλότερη είναι η ανάλυση της εικόνας, τόσο χειρότερη είναι η απόδοση, καθώς η έξοδος τείνει να φαίνεται θολή.

Έτσι, όλα αυτά τα πράγματα μπορούν να γίνουν μόνο και μόνο επειδή βρήκαμε μια νέα αναπαράσταση των δεδομένων που ήδη είχαμε. Αυτή η αναπαράσταση είναι χρήσιμη επειδή αναδεικνύει τις βασικές πληροφορίες, οι οποίες είναι εγγενώς συμπιεσμένες, αλλά είναι επίσης πιο εύκολο να χειριστούμε ή να δημιουργήσουμε νέα δεδομένα με βάση τη λανθάνουσα αναπαράσταση.

6.4 Variational Autoencoder (VAE)

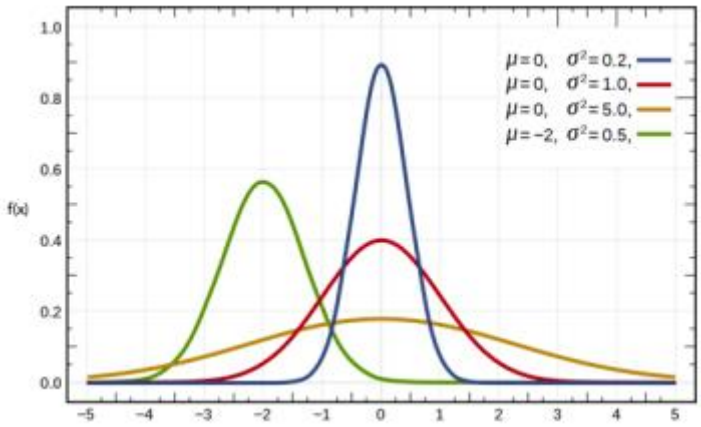
Οι Variational Autoencoders, ανακαλύφθηκαν από τους Kingma και Welling τον Δεκέμβριο του 2013 και τους Rezende, Mohamed και Wierstra τον Ιανουάριο του 2014. Ποια είναι όμως η διαφορά μεταξύ ενός Variational Autoencoder και ενός "κανονικού"; Όλα έχουν να κάνουν με τον λανθάνοντα χώρο. Στην περίπτωση ενός VAE, επιλέγουμε να αναπαραστήσουμε τον λανθάνοντα χώρο ως κατανομή με έναν μαθημένο μέσο όρο και μια τυπική απόκλιση και όχι απλώς ως ένα σύνολο αριθμών. Συνήθως, επιλέγουμε έναν πολυμεταβλητή Gaussian.



Σχήμα 6-4. Η διαφορά μεταξύ ενός αυτόματου κωδικοποιητή και ενός VAE, στο λανθάνοντα χώρο
(Πηγή: Generative Deep Learning, Teaching Machines to Paint, Write, Compose and Play, 2019)

Εκτενέστερα, σε έναν αυτόματο κωδικοποιητή, κάθε εικόνα αντιστοιχίζεται απευθείας σε ένα σημείο του λανθάνοντος χώρου. Σε έναν VAE, κάθε εικόνα αντιστοιχίζεται σε μια πολυμεταβλητή κανονική κατανομή (multivariate normal distribution) γύρω από ένα σημείο στον λανθάνοντα χώρο, όπως φαίνεται στο σχήμα 6-4. Η κανονική κατανομή, γνωστή ως και Gaussian κατανομή, είναι μια κατανομή πιθανότητας που χαρακτηρίζεται από ένα χαρακτηριστικό σχήμα καμπύλης καμπάνας. Σε μία διάσταση, ορίζεται από δύο μεταβλητές: το μέσο όρο (μ) και τη διακύμανση (σ^2). Η τυπική απόκλιση (σ) είναι η τετραγωνική ρίζα της διακύμανσης. Η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής σε μία διάσταση είναι:

$$f\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Σχήμα 6-5. Κανονική κατανομή σε 1 διάσταση
(Πηγή: Generative Deep Learning, Teaching Machines to Paint, Write, Compose and Play, 2019)

Το σχήμα 6-5. δείχνει πολλές κανονικές κατανομές σε μία διάσταση, για διαφορετικές τιμές του μέσου όρου και της διακύμανσης. Η κόκκινη καμπύλη είναι η τυπική κανονική, δηλαδή η κατανομή με μέσο όρο ίσο με 0 και διακύμανση ίση με 1.
Μπορούμε να πάρουμε δείγμα ένα σημείο z από μια κανονική κατανομή με μέσο μ και τυπική απόκλιση σ χρησιμοποιώντας την ακόλουθη εξίσωση: z = μ + σε όπου ϵ λαμβάνεται δείγμα από μια τυπική κανονική κατανομή.

Η έννοια της κανονικής κατανομής εκτείνεται σε περισσότερες από μία διαστάσεις, η συνάρτηση πυκνότητας πιθανότητας για μια γενική πολυμεταβλητή κανονική κατανομή σε k διαστάσεις είναι η εξής:

$$f(x_1, \ldots, x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right)}{\sqrt{(2\pi)^k|\Sigma|}}$$

Στην πράξη, οι κλασικοί αυτοκωδικοποιητές δεν οδηγούν σε ιδιαίτερα χρήσιμους λανθάνοντες χώρους. Τα VAE, ωστόσο, προσθέτουν λίγη στατιστική δύναμη που τους αναγκάζει να μάθουν συνεχείς, εξαιρετικά δομημένους λανθάνοντες χώρους. Το οποίο αποδεικνύει να είναι ένα ισχυρό εργαλείο για τη δημιουργία νέων δεδομένων. Ένα VAE, αντί να συμπιέσει τα δεδομένα εισόδου του σε έναν σταθερό διάνυσμα στον λανθάνοντα χώρο, μετατρέπει τα δεδομένα σε διάνυσμα μιας στατιστικής κατανομής: με μέσο όρο και μια διακύμανση. Ουσιαστικά, αυτό σημαίνει ότι υποθέτουμε ότι τα δεδομένα εισόδου έχουν δημιουργηθεί από μια στατιστική διαδικασία και ότι θα πρέπει να ληφθεί υπόψη η τυχαιότητα αυτής της διαδικασίας κατά την κωδικοποίηση και την αποκωδικοποίηση. Στη συνέχεια, το VAE χρησιμοποιεί τον μέσο όρο και τη διακύμανση για τυχαία δειγματοληψία ενός στοιχείου της κατανομής και αποκωδικοποιεί αυτό το στοιχείο πίσω στην αρχική είσοδο. Η στοχαστικότητα αυτής της διαδικασίας βελτιώνει και αναγκάζει τον λανθάνοντα χώρο να κωδικοποιεί ουσιαστικές αναπαραστάσεις. Με άλλα λόγια, οι VAE προσπαθούν να βρουν τις σωστές παραμέτρους που ορίζουν μια κατανομή.

Σχήμα 6-6. Ένα VAE με διάνυσμα κατανομής χαμόγελου
(Πηγή: Deep Learning with Python, 2018)

6.5 Περίληψη των VAE

Σε αυτό το κεφάλαιο είδαμε πώς οι VAE είναι ένα ισχυρό εργαλείο στην εργαλειοθήκη της γεννητικής μοντελοποίησης. Ξεκινήσαμε διερευνώντας πώς μπορούν να χρησιμοποιηθούν απλοί αυτοκωδικοποιητές για τη χαρτογράφηση δεδομένων υψηλών διαστάσεων σε έναν λανθάνοντα χώρο χαμηλής διάστασης, έτσι ώστε να μπορούν να εξαχθούν χαρακτηριστικά υψηλού επιπέδου από τα μεμονωμένα δεδομένα. Ωστόσο, υπάρχουν ορισμένα μειονεκτήματα στη χρήση απλών αυτοκωδικοποιητών ως παραγωγικού μοντέλου, η δειγματοληψία από τον εκμαθημένο λανθάνοντα χώρο ήταν περιορισμένη. Οι VAE λύνουν αυτά τα προβλήματα, εισάγοντας την στοχαστικότητα στο μοντέλο και περιορίζοντας τον τρόπο κατανομής των σημείων στον λανθάνοντα χώρο. Με μερικές μικρές προσαρμογές στην κατανομή του λανθάνοντα χώρου, μπορούμε να μετατρέψουμε τον αυτοκωδικοποιητή μας σε έναν VAE, δίνοντάς του έτσι τη δύναμη να είναι ένα εύστοχο μοντέλο παραγωγής. Στην πράξη, αυτή τείνει να μην είναι η κύρια χρήση τους γιατί έχουν περιορισμούς και επίσης υπάρχουν άλλες μέθοδοι, που μας παρακινούν να προχωρήσουμε στα GANs.

7. Εισαγωγή στα GANs

Η ιδέα του αν οι μηχανές μπορούν να σκεφτούν είναι παλαιότερη από τον ίδιο τον υπολογιστή. Το 1950, ο διάσημος μαθηματικός, και επιστήμονας υπολογιστών Άλαν Τούρινγκ έγραψε ένα βιβλίο που θα άφηνε το όνομά του για τις επόμενες γενιές, «Υπολογιστικά Μηχανήματα και Νοημοσύνη. ” Στο βιβλίο, ο Τούρινγκ πρότεινε ένα τεστ που το ονόμασε παιχνίδι μίμησης, πιο γνωστό σήμερα ως Τούρινγκ τεστ. Σε αυτό το υποθετικό σενάριο, ένας άγνωστος παρατηρητής συνομιλεί με δύο οντότητες πίσω από μια κλειστή πόρτα: Ένας άνθρωπος. Το άλλο, ένας υπολογιστής. Ο Turing εξηγεί ότι εάν ο παρατηρητής δεν μπορεί να ξεχωρίσει ποιο είναι το άτομο και ποιο το μηχάνημα, ο υπολογιστής περνάει το τεστ και πρέπει να θεωρηθεί έξυπνο. Όποιος έχει προσπαθήσει να συμμετάσχει σε διάλογο με ένα αυτοματοποιημένο chatbot ή έναν ευφυή βοηθό που λειτουργεί με φωνή, γνωρίζει ότι οι υπολογιστές έχουν πολύ δρόμο να διανύσουν για να περάσουν αυτό το απλό τεστ. Ωστόσο, σε άλλες εργασίες, οι υπολογιστές όχι μόνο έχουν φτάσει την ανθρώπινη απόδοση, αλλά και την έχουν ξεπεράσει, ακόμη και σε τομείς που μέχρι πρόσφατα θεωρούνταν απρόσιτοι ακόμη και για τους πιο έξυπνους αλγόριθμους, όπως η υπεράνθρωπη ακριβής αναγνώριση προσώπου για παράδειγμα. Οι αλγόριθμοι μηχανικής μάθησης είναι εξαιρετικοί στην αναγνώριση μοτίβων σε υπάρχοντα δεδομένα και στη χρήση αυτής της γνώσης για εργασίες όπως η ταξινόμηση (classification: η ανάθεση της σωστής κατηγορίας σε ένα παράδειγμα) και η παλινδρόμηση (regression: η εκτίμηση μιας αριθμητικής τιμής με βάση μια ποικιλία εισόδων). Όταν τους ζητηθεί να δημιουργήσουν νέα δεδομένα, ωστόσο, οι υπολογιστές πάντα δυσκολευόντουσαν. Όλα αυτά άλλαξαν το 2014 όταν ο Ian Goodfellow, τότε διδάκτορας στο Πανεπιστήμιο του Μόντρεαλ, εφηύρε τα Γεννητικά Ανταγωνιστικά Δίκτυα - Generative Adversarial Networks (GANs). Αυτή η τεχνική επέτρεψε στους υπολογιστές να παράγουν ρεαλιστικά δεδομένα χρησιμοποιώντας όχι ένα, αλλά δύο ξεχωριστά νευρωνικά δίκτυα. Τα GAN δεν ήταν το πρώτο πρόγραμμα υπολογιστή που χρησιμοποιήθηκε για τη δημιουργία δεδομένων, αλλά τα αποτελέσματα και η ευελιξία τους τα ξεχώρισαν από όλα τα υπόλοιπα. Τα GAN έχουν επιτύχει αξιοσημείωτα αποτελέσματα που θεωρούνταν από καιρό σχεδόν αδύνατα για τεχνητά συστήματα, όλα αυτά χωρίς την ανάγκη για τεράστιες ποσότητες επισημασμένων δεδομένων (labeled dataset). Μόλις το 2014, όταν εφευρέθηκαν τα GAN, το καλύτερο που μπορούσαν να παράγουν οι μηχανές ήταν μια θολή όψη, και ακόμη και αυτό επικροτήθηκε ως πρωτοποριακή επιτυχία. Μέχρι το 2017, μόλις τρία χρόνια αργότερα, η πρόοδος στα GAN επέτρεψε στους υπολογιστές να συνθέτουν ψεύτικα πρόσωπα των οποίων η ποιότητα τους συναγωνίζεται τις φωτογραφίες υψηλής ανάλυσης.

7.1 Τι είναι τα Γεννητικά Ανταγωνιστικά Δίκτυα

Τα Generative Adversarial Networks (GANs) είναι μια κατηγορία τεχνικών μηχανικής εκμάθησης που αποτελούνται από δύο μοντέλα που εκπαιδεύονται ταυτόχρονα: το ένα (ο Δημιουργός - Generator) που είναι εκπαιδευμένο να δημιουργεί πλαστά δεδομένα και το άλλο (ο Διευκρινιστής - Discriminator) εκπαιδευμένο να διακρίνει τα πλαστά δεδομένα από πραγματικά παραδείγματα. Η λέξη Γεννητικά (Generative) υποδηλώνει τον γενικό σκοπό του μοντέλου: τη δημιουργία νέων δεδομένων. Τα δεδομένα που θα μάθει να παράγει ένα GAN εξαρτώνται από την επιλογή του σετ εκπαίδευσης. Για παράδειγμα, αν θέλουμε ένα GAN να συνθέτει εικόνες που μοιάζουν με του Leonardo da Vinci, θα χρησιμοποιούσαμε ένα εκπαιδευτικό σύνολο δεδομένων του έργου τέχνης του da Vinci. Ο όρος Ανταγωνιστικά (Adversarial) δείχνει την ανταγωνιστική δυναμική που μοιάζει με παιχνίδι μεταξύ των δύο μοντέλων που αποτελούν το πλαίσιο των GAN: του Generator και του Discriminator. Ο

στόχος του Δημιουργού είναι να δημιουργήσει παραδείγματα που δεν διακρίνονται από τα πραγματικά δεδομένα στο σετ εκπαίδευσης. Στο παράδειγμά μας, αυτό σημαίνει την παραγωγή έργων ζωγραφικής που μοιάζουν ακριβώς με του da Vinci. Ο στόχος του Διευκρινιστή είναι να διακρίνει τα πλαστά παραδείγματα που παράγονται από τον Δημιουργό, από τα πραγματικά παραδείγματα που προέρχονται από το σύνολο δεδομένων εκπαίδευσης. Στο παράδειγμά μας, ο Διευκρινιστής παίζει το ρόλο ενός ειδικού τέχνης που αξιολογεί την αυθεντικότητα των πινάκων που πιστεύεται ότι είναι του da Vinci. Τα δύο δίκτυα προσπαθούν συνεχώς να ξεγελάσουν το ένα το άλλο, όσο καλύτερος είναι ο Δημιουργός στη δημιουργία πειστικών δεδομένων, τόσο καλύτερος πρέπει να είναι ο Διευκρινιστής στη διάκριση των πραγματικών παραδειγμάτων από τα πλαστά. Τέλος, η λέξη Δίκτυα (Networks) υποδηλώνει την κατηγορία των μοντέλων μηχανικής μάθησης που χρησιμοποιούνται συχνότερα για την αναπαράσταση του Generator και του Discriminator, τα νευρωνικά δίκτυα. Ανάλογα με την πολυπλοκότητα της υλοποίησης του GAN, αυτά μπορεί να κυμαίνονται από απλά νευρωνικά δίκτυα έως και συνελικτικά νευρωνικά δίκτυα ή ακόμα και πιο περίπλοκες παραλλαγές.

7.2 Πως λειτουργούν τα GANs

Με πιο τεχνικούς όρους, ο στόχος του Δημιουργού είναι να παράγει παραδείγματα που αποτυπώνουν τα χαρακτηριστικά του συνόλου δεδομένων εκπαίδευσης, τόσο πολύ ώστε τα δείγματα που δημιουργεί να φαίνονται δυσδιάκριτα από τα δεδομένα εκπαίδευσης. Ο Δημιουργός μαθαίνει μέσω της ανατροφοδότησης που λαμβάνει από τις ταξινομήσεις του Διευκρινιστή. Ο στόχος του Διευκρινιστή είναι να προσδιορίσει εάν ένα συγκεκριμένο παράδειγμα είναι πραγματικό (που προέρχεται από το σύνολο δεδομένων εκπαίδευσης) ή πλαστό (που δημιουργήθηκε από τον Δημιουργό). Αντίστοιχα, κάθε φορά που ο Διευκρινιστής ξεγελιέται και ταξινομεί μια ψεύτικη εικόνα ως πραγματική, ο Δημιουργός ξέρει ότι έκανε κάτι καλά. Αντίστροφα, κάθε φορά που ο Διευκρινιστής απορρίπτει σωστά μια εικόνα που έχει παραχθεί από τον Δημιουργό ως ψεύτικη, ο Δημιουργός λαμβάνει την ανατροφοδότηση ότι πρέπει να βελτιώσει. Ο Διευκρινιστής συνεχίζει επίσης να βελτιώνεται. Όπως κάθε ταξινομητής, μαθαίνει πόσο μακριά απέχουν οι προβλέψεις του από τις αληθινές ετικέτες (πραγματικές ή πλαστές). Έτσι, καθώς ο Δημιουργός βελτιώνεται στην παραγωγή δεδομένων με ρεαλιστική εμφάνιση, ο Διευκρινιστής γίνεται καλύτερος στο να ξεχωρίζει τα πλαστά δεδομένα από τα πραγματικά οπότε τα δύο δίκτυα συνεχίζουν να βελτιώνονται ταυτόχρονα.

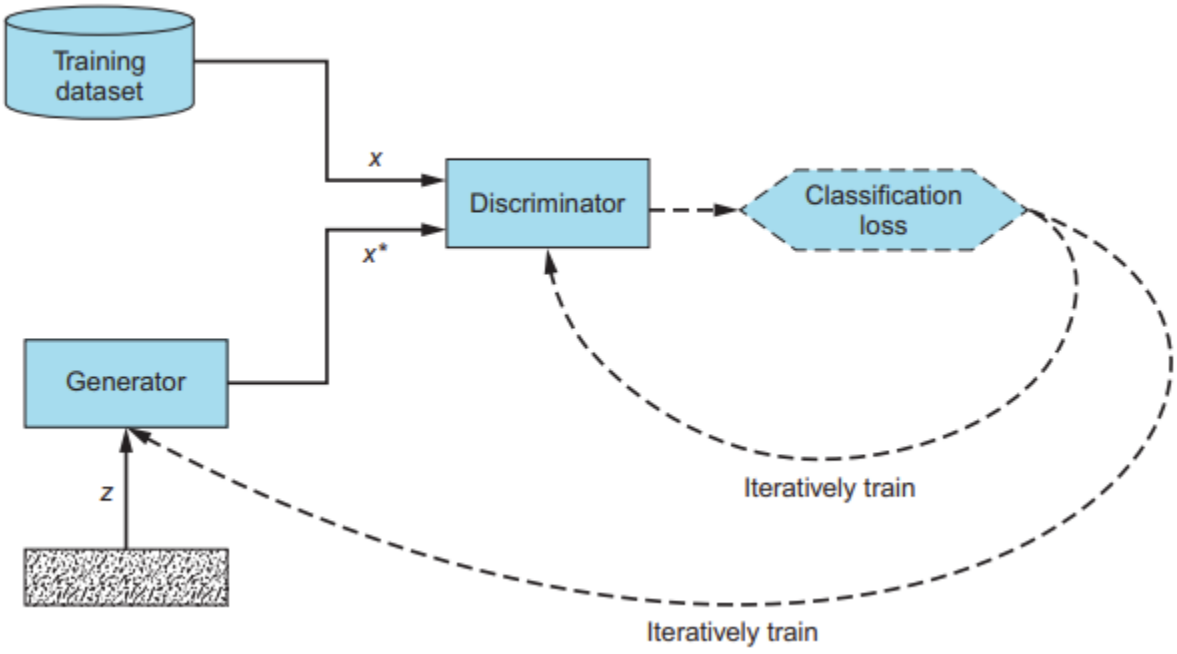
	Generator	Discriminator
Είσοδος	Διάνυσμα τυχαίων αριθμών	Ο Discriminator λαμβάνει είσοδο από 2 πηγές: • Αληθινά δεδομένα από το σετ εκπαίδευσης • Ψευδή δεδομένα από τον Generator
Έξοδος	Ψευδή παραδείγματα που προσπαθούν να είναι όσο το δυνατόν πιο πειστικά	Προβλεπόμενη πιθανότητα ότι τα δεδομένα είναι αληθινά
Στόχος	Δημιουργία πλαστών δεδομένων τα οποία είναι δυσδιάκριτα από τα αληθινά δεδομένα εκπαίδευσης	Διάκριση μεταξύ πλαστών παραδειγμάτων που προέρχονται από τον Generator και των πραγματικών παραδειγμάτων που προέρχονται από το σύνολο δεδομένων εκπαίδευσης

Εικόνα 7-1. Τα δύο υποδίκτυα GAN, σε συνάρτηση με τις εισόδους, εξόδους και στόχους του καθενός

(Πηγή: GANs in Action Deep Learning with Generative Adversarial Networks, 2019)

Ας δούμε τις λεπτομέρειες του διαγράμματος:

1. Δεδομένα εκπαίδευσης (training data): Το σύνολο δεδομένων των πραγματικών παραδειγμάτων που θέλουμε ο Δημιουργός να μάθει να μιμείται με σχεδόν τέλεια ποιότητα. Αυτό το σύνολο δεδομένων χρησιμεύει ως είσοδος (x) στο δίκτυο του Διευκρινιστή.
2. Διάνυσμα τυχαίου θορύβου: Η πρωτογενής είσοδος (z) στο δίκτυο του Δημιουργού. Αυτή η είσοδος είναι ένα διάνυσμα τυχαίων αριθμών που χρησιμοποιεί ο Δημιουργός ως σημείο εκκίνησης για τη σύνθεση πλαστών παραδειγμάτων.
3. Δίκτυο Δημιουργού: Ο Δημιουργός παίρνει ένα διάνυσμα τυχαίων αριθμών (z) ως είσοδο και εξάγει πλαστά παραδείγματα (x*). Στόχος του είναι να κάνει τα ψεύτικα παραδείγματα που παράγει να μην διακρίνονται από τα πραγματικά παραδείγματα από σύνολο δεδομένων εκπαίδευσης.
4. Δίκτυο Διευκρινιστή: Ο Διευκρινιστής λαμβάνει ως είσοδο είτε ένα πραγματικό παράδειγμα (x) που προέρχεται από το σετ εκπαίδευσης είτε ένα ψεύτικο παράδειγμα (x*) που παράγεται από τον Δημιουργό. Για κάθε παράδειγμα, ο Διευκρινιστής καθορίζει και εξάγει την πιθανότητα εάν το παράδειγμα είναι πραγματικό.
5. Επαναληπτική εκπαίδευση/συντονισμός: Για καθεμία από τις προβλέψεις του Διευκρινιστή, προσδιορίζουμε πόσο καλή είναι και χρησιμοποιούμε τα αποτελέσματα για να συντονίσουμε επαναληπτικά τα δίκτυα Διευκρινιστή και Δημιουργό μέσω backpropagation:
   – Τα βάρη (weights) και τα biases του Διευκρινιστή ενημερώνονται για να μεγιστοποιηθεί η ακρίβεια ταξινόμησης του (μεγιστοποιώντας την πιθανότητα σωστής πρόβλεψης: x ως πραγματική και x* ως ψεύτικη).
   – Τα βάρη (weights) και τα biases του Δημιουργού ενημερώνονται για να μεγιστοποιηθεί η πιθανότητα ο Διευκρινιστής να ταξινομήσει εσφαλμένα το x* ως πραγματικό.



Σχήμα 7-2. Εικονογράφηση της λειτουργίας GAN
(Πηγή: GANs in Action Deep Learning with Generative Adversarial Networks, 2019)

Τυπικά, ο Generator και ο Discriminator αντιπροσωπεύονται από διαφοροποιήσιμες συναρτήσεις, όπως τα νευρωνικά δίκτυα, το καθένα με τη δική του συνάρτηση κόστους (loss function). Τα δύο δίκτυα εκπαιδεύονται με backpropagation χρησιμοποιώντας την απώλεια του Διευκρινιστή. Ο Διευκρινιστής προσπαθεί να ελαχιστοποιήσει την απώλεια τόσο για τα πραγματικά όσο και για τα ψεύτικα παραδείγματα, ενώ ο Δημιουργός προσπαθεί να μεγιστοποιήσει την απώλεια του Διευκρινιστή για τα πλαστά παραδείγματα που παράγει. Είναι σημαντικό ότι το σύνολο δεδομένων εκπαίδευσης καθορίζει το είδος των παραδειγμάτων που θα μάθει να μιμείται ο Δημιουργός.

7.3 Αλγόριθμος εκπαίδευσης GAN:

Για κάθε επανάληψη εκπαίδευσης κάνε (for each training iteration do)

1. Εκπαίδευσε τον Διευκρινιστή:
a. Πάρε ένα τυχαίο πραγματικό παράδειγμα x από το σύνολο δεδομένων εκπαίδευσης.
b. Πάρε ένα νέο διάνυσμα τυχαίου θορύβου z και, χρησιμοποιώντας το δίκτυο Δημιουργό, σύνθεσε ένα ψεύτικο παράδειγμα x*.
c. Χρησιμοποίησε το δίκτυο Διευκρινιστή για να ταξινομήσεις τα x και x*.
d. Υπολόγισε τα σφάλματα ταξινόμησης και backpropagate το συνολικό σφάλμα για να ενημερώσεις τις εκπαιδεύσιμες παραμέτρους του Διευκρινιστή, επιδιώκοντας να ελαχιστοποιήσεις τα σφάλματα ταξινόμησης.

2. Εκπαίδευσε τον Δημιουργό:
a. Απόκτησε ένα νέο διάνυσμα τυχαίου θορύβου z και, χρησιμοποιώντας το δίκτυο Δημιουργού, σύνθεσε ένα ψεύτικο παράδειγμα x*.
b. Χρησιμοποίησε το δίκτυο Διευκρινιστή για να ταξινομήσεις το x*.
c. Υπολόγισε το σφάλμα ταξινόμησης και backpropagate το σφάλμα για να ενημερώσεις τις εκπαιδεύσιμες παραμέτρους του Δημιουργού, επιδιώκοντας να μεγιστοποιήσεις το σφάλμα του Διευκρινιστή.

Τέλος για (End for)

7.4 Συναρτήσεις κόστους (Cost functions)

Έστω $J^{(G)}$ η συνάρτηση κόστους του Generator και $J^{(D)}$ η συνάρτηση κόστους του Discriminator. Τα GAN διαφέρουν από τα συμβατικά νευρωνικά δίκτυα σε δύο βασικές απόψεις. Πρώτον, η συνάρτηση κόστους, J, ενός παραδοσιακού νευρωνικού δικτύου ορίζεται αποκλειστικά με βάση τις δικές του εκπαιδεύσιμες παραμέτρους, θ. Μαθηματικά, αυτό εκφράζεται ως J(θ). Αντίθετα, τα GAN αποτελούνται από δύο δίκτυα των οποίων οι συναρτήσεις κόστους εξαρτώνται και από τις δύο παραμέτρους των δικτύων. Δηλαδή, η συνάρτηση κόστους του Generator είναι $J^{(G)}((G), (D))$, και η συνάρτηση κόστους του Discriminator είναι $J^{(D)}((G), (D))$. Η δεύτερη διαφορά είναι ότι ένα παραδοσιακό νευρωνικό δίκτυο μπορεί να συντονίσει όλες τις παραμέτρους του θ, κατά τη διάρκεια της εκπαιδευτικής διαδικασίας. Σε ένα GAN, κάθε δίκτυο μπορεί να συντονίσει μόνο τα δικά του βάρη (w) και biases (b). Ο Generator μπορεί να συντονιστεί μόνο το $\theta^{(G)}$ και ο Discriminator μπορεί να συντονιστεί μόνο το $\theta^{(D)}$ κατά τη διάρκεια της εκπαίδευσης. Αντίστοιχα, κάθε δίκτυο έχει τον έλεγχο μόνο ενός μέρους αυτού που καθορίζει την απώλειά του.

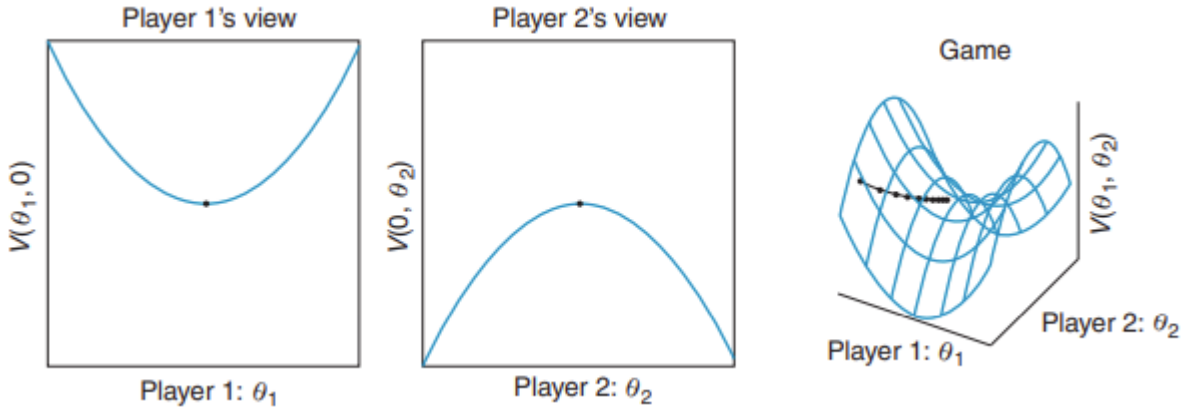
7.5 Διαδικασία εκπαίδευσης - Φτάνοντας σε ισορροπία

Πώς γνωρίζουμε πότε ένα GAN είναι πλήρως εκπαιδευμένο, ώστε να μπορούμε να προσδιορίσουμε τον κατάλληλο αριθμό επαναλήψεων εκπαίδευσης; Με ένα κανονικό νευρωνικό δίκτυο, έχουμε συνήθως έναν σαφή στόχο να επιτύχουμε και να μετρήσουμε. Για παράδειγμα, όταν εκπαιδεύουμε έναν ταξινομητή, μετράμε το σφάλμα ταξινόμησης στο σετ εκπαίδευσης (training set) και επικύρωσης (validation set) και σταματάμε τη διαδικασία όταν το σφάλμα επικύρωσης αρχίσει να χειροτερεύει, να υπερ-προσαρμόζεται (overfit). Σε ένα GAN, τα δύο δίκτυα έχουν ανταγωνιστικούς στόχους, όταν το ένα δίκτυο γίνεται καλύτερο, το άλλο χειροτερεύει. Πώς καθορίζουμε πότε θα σταματήσουμε; Όσοι είναι εξοικειωμένοι με τη θεωρία παιγνίων μπορούν να αναγνωρίσουν αυτή τη ρύθμιση ως ένα παιχνίδι μηδενικού αθροίσματος, μια κατάσταση στην οποία τα κέρδη ενός παίκτη ισούνται με τις απώλειες του άλλου παίκτη. Όταν ένας παίκτης βελτιώνεται κατά ένα ορισμένο ποσό, ο άλλος παίκτης χειροτερεύει κατά το ίδιο ποσό. Όλα τα παίγνια μηδενικού αθροίσματος έχουν μια ισορροπία Nash[8], ένα σημείο στο οποίο κανένας παίκτης δεν μπορεί να βελτιώσει την κατάστασή του ή να αποδώσει αλλάζοντας τις ενέργειές του. Το GAN φτάνει στην ισορροπία Nash όταν πληρούνται οι ακόλουθες προϋποθέσεις: Ο Δημιουργός παράγει ψεύτικα παραδείγματα που δεν διακρίνονται από τα πραγματικά δεδομένα του συνόλου δεδομένων εκπαίδευσης. Ο Διευκρινιστής μπορεί στην καλύτερη περίπτωση να μαντέψει τυχαία εάν ένα συγκεκριμένο παράδειγμα είναι πραγματικό ή ψεύτικο (δηλαδή, να κάνει μια εικασία 50/50 εάν ένα παράδειγμα είναι πραγματικό). Για παράδειγμα, όταν καθένα από τα ψεύτικα παραδείγματα (x*) είναι πραγματικά δυσδιάκριτα από τα πραγματικά παραδείγματα (x) που προέρχονται από το σύνολο δεδομένων εκπαίδευσης, δεν μπορεί να χρησιμοποιήσει τίποτα ο Διευκρινιστής για να τα ξεχωρίσει το ένα από το άλλο. Επειδή τα μισά από τα παραδείγματα που λαμβάνει είναι αληθινά και τα μισά είναι πλαστά, το καλύτερο που μπορεί να κάνει ο Διευκρινιστής είναι να γυρίσει ένα νόμισμα και να ταξινομήσει κάθε παράδειγμα ως πραγματικό ή πλαστό με πιθανότητα 50%. Ο Δημιουργός είναι επίσης σε ένα σημείο όπου δεν έχει τίποτα να κερδίσει από περαιτέρω συντονισμό. Επειδή τα παραδείγματα που παράγει δεν διακρίνονται ήδη από τα πραγματικά, ακόμη και μια μικροσκοπική αλλαγή στη διαδικασία που χρησιμοποιεί για να μετατρέψει το διάνυσμα τυχαίου θορύβου (z) σε ψεύτικο παράδειγμα (x*) μπορεί να δώσει στο Διευκρινιστή μια ένδειξη για το πώς να διακρίνει το ψεύτικο παράδειγμα από τα πραγματικά δεδομένα, κάνοντας τον Δημιουργό χειρότερο. Με την επίτευξη ισορροπίας, το GAN λέγεται ότι έχει συγκλίνει. Εδώ είναι το πιο περίπλοκο σημείο. Στην πράξη, είναι σχεδόν αδύνατο να βρεθεί η ισορροπία Nash, γιατί τα GAN λόγω της τεράστιας πολυπλοκότητας τους, δυσκολεύονται στην επίτευξη σύγκλισης. Πράγματι, η σύγκλιση GAN παραμένει ένα από τα πιο σημαντικά ανοιχτά ερωτήματα στην έρευνα GAN.

8 Ισορροπία Nash: Το Nash equilibrium είναι ένα θεώρημα λήψης αποφάσεων, πήρε το όνομά του από τον Αμερικανό οικονομολόγο και μαθηματικό John Forbes Nash Jr.

$$E_x[log(D(x))] + E_z[log(1 - D(G(z)))]$$

Σχήμα 7-3. Συνάρτηση απώλειας GAN
(Πηγή ιστοσελίδα: https://neptune.ai/blog/gan-loss-functions)



Σχήμα 7-4. Γραφική παράσταση του Generator και του Discriminator σε δισδιάστατη μορφή
(Πηγή: GANs in Action Deep Learning with Generative Adversarial Networks, 2019)

7.6 Πίνακας σύγχυσης - Confusion matrix

Οι ταξινομήσεις του Discriminator μπορούν να εκφραστούν με όρους ενός πίνακα σύγχυσης, μια αναπαράσταση πίνακα όλων των πιθανών αποτελεσμάτων στη δυαδική ταξινόμηση. Στην περίπτωση του Διευκρινιστή, αυτά είναι τα εξής:

- Αληθινό θετικό: Πραγματικό παράδειγμα σωστά ταξινομημένο ως πραγματικό. D(x)=1
- Ψευδές αρνητικό: Πραγματικό παράδειγμα εσφαλμένα ταξινομημένο ως ψεύτικο. D(x)=0
- Αληθινό αρνητικό: Το ψεύτικο παράδειγμα σωστά ταξινομημένο ως ψεύτικο. D(x*)=0
- Ψευδές θετικό: Ψεύτικο παράδειγμα εσφαλμένα ταξινομημένο ως πραγματικό. D(x*)=1

Input	Discriminator output	
	Close to 1 (real)	*Close to 0 (fake)*
Αληθινό(x)	Αληθινό θετικό	Ψευδές αρνητικό
Πλαστό(x*)	Ψευδές θετικό	Αληθινό αρνητικό

Σχήμα 7-5. Πίνακας Σύγχυσης ενός GAN
(Πηγή: GANs in Action Deep Learning with Generative Adversarial Networks, 2019)

Χρησιμοποιώντας την ορολογία του πίνακα σύγχυσης, ο Διευκρινιστής προσπαθεί να μεγιστοποιήσει τις αληθινές θετικές και αληθινές αρνητικές ταξινομήσεις ή, ισοδύναμα, να ελαχιστοποιήσει τις ψευδώς θετικές και ψευδώς αρνητικές ταξινομήσεις. Αντίθετα, ο στόχος του Δημιουργού είναι να μεγιστοποιήσει τις ψευδώς θετικές ταξινομήσεις του Διευκρινιστη, αυτές είναι οι περιπτώσεις στις οποίες ο Δημιουργός ξεγελά επιτυχώς τον Διευκρινιστή ώστε να πιστέψει ότι ένα ψεύτικο παράδειγμα είναι πραγματικό. Ο Δημιουργός δεν ενδιαφέρεται για το πόσο καλά ταξινομεί ο Διευκρινιστής τα πραγματικά παραδείγματα, νοιάζεται μόνο για τις ταξινομήσεις των πλαστών δειγμάτων δεδομένων από τον Διευκρινιστή.

7.7 Εκπαίδευση και κοινές προκλήσεις των GAN

Δεδομένου ότι τα GAN εφευρέθηκαν για πρώτη φορά το 2014, η εκπαίδευση και η αξιολόγησή τους έχει παραμείνει το πιο δύσκολο κομμάτι τους. Ερευνητές έχουν δοκιμάσει και έχουν προτείνει πολλές διαφορετικές λύσεις για να διευκολύνουν αυτό το έργο. Οι περισσότερες από αυτές μπορούν να εξηγηθούν μαθηματικά, αλλά μερικές από αυτές τις λύσεις είναι αρκετά εμπειρικές.

Λίστα με τα κυριότερα προβλήματα:

- Σύμπτυξη λειτουργίας (Mode collapse): Το mode collapse συμβαίνει όταν ο Generator μπορεί να παράγει μόνο έναν τύπο εξόδου/δείγμα ή ένα μικρό σύνολο εξόδων. Αυτό μπορεί να συμβεί λόγω προβλημάτων στην εκπαίδευση, όπως όταν ο Generator βρίσκει έναν τύπο δεδομένων που είναι εύκολα σε θέση να ξεγελάσει τον Discriminator και έτσι συνεχίζει να δημιουργεί αυτόν τον έναν τύπο. Επειδή δεν υπάρχει κίνητρο για τον Generator να αλλάξει, ολόκληρο το μοντέλο θα βελτιστοποιηθεί υπερβολικά στη συγκεκριμένη έξοδο.

- Αργή σύγκλιση (Slow convergence): Αυτό είναι ένα μεγάλο πρόβλημα με τα GAN, όπου γενικά η ταχύτητα σύγκλισης και ο διαθέσιμος υπολογισμός είναι οι κύριοι περιορισμοί.

- Υπερ-γενίκευση (Overgeneralization): Ανισορροπία μεταξύ Generator και Discriminator που προκαλεί υπερ-προσαρμογή. Για παράδειγμα, μπορεί να δούμε μια αγελάδα με πολλά σώματα αλλά μόνο ένα κεφάλι ή το αντίστροφο. Αυτό συμβαίνει όταν το GAN υπερ-γενικεύει και μαθαίνει πράγματα που δεν πρέπει να υπάρχουν με βάση τα πραγματικά δεδομένα.

- Αποτυχία σύγκλισης (Convergence failure): Προκαλείται όταν ο Discriminator απορρίπτει επιτυχώς δείγματα του Generator με υψηλή εμπιστοσύνη. Σε αυτή την περίπτωση, δεν έχει βρεθεί η ισορροπία μεταξύ των δύο δικτύων, με αποτέλεσμα ο Generator να μην βρίσκει τρόπο να ξεγελάσει τον Discriminator και το δίκτυο να μην εκπαιδεύεται καθόλου.

Όσον αφορά την εκπαίδευση των GAN, πολλές τεχνικές μπορούν να μας βοηθήσουν να βελτιώσουμε τη διαδικασία εκπαίδευσης. Όμως χρειάζεται ιδιαίτερη προσοχή διότι τα GAN είναι αρκετά ευάλωτα και ευαίσθητα σε τέτοιες αλλαγές. Κάποιες από τις μεθοδολογίες είναι οι παρακάτω:

1. Προσθήκη βάθους δικτύου
2. Συνάρτηση απώλειας
3. Τεχνικές Εκπαίδευσης

1. Προσθήκη βάθους δικτύου

Όπως συμβαίνει με πολλούς αλγόριθμους μηχανικής μάθησης, ο ευκολότερος τρόπος για να κάνουμε τη μάθηση πιο σταθερή είναι να μειώσουμε την πολυπλοκότητα. Εάν μπορούμε να ξεκινήσουμε με έναν απλό αλγόριθμο και σιγά σιγά να προσθέτουμε, θα έχουμε μεγαλύτερη σταθερότητα κατά τη διάρκεια της εκπαίδευσης, ταχύτερη σύγκλιση και ενδεχομένως άλλα οφέλη.

Συνάρτηση Απώλειας

Ένας τρόπος για να σκεφτούμε την ανταγωνιστική φύση των GAN για δύο παίκτες είναι να φανταστούμε ότι παίζουμε ένα επιτραπέζιο παιχνίδι που μπορεί να τελειώσει ανά πάσα στιγμή. Ως παίκτης, πρέπει να είμαστε σε θέση όχι μόνο να γνωρίζουμε τον στόχο του παιχνιδιού και επομένως τι προσπαθούν να επιτύχουν και οι δύο παίκτες, αλλά και να καταλάβουμε πόσο κοντά είμαστε στη νίκη. Άρα έχουμε τους κανόνες και την μέτρηση απόστασης (νίκης). Αλλά όπως σε κάθε διαφορετικό παιχνίδι αλλάζει η μέτρηση της νίκης, έτσι ομοίως αλλάζει και σε κάθε GAN. Η μέτρηση της νίκης με άλλα λόγια είναι η συνάρτηση απώλειας.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right]$$

Σχήμα 7-6. Συνάρτηση απώλειας του Discriminator
(Πηγή ιστοσελίδα: https://jonathan-hui.medium.com/gan-wasserstein-gan-wgan-gp-6a1a2aa1b490)

Το D σημαίνει συνάρτηση του Discriminator (αντιστοίχιση δεδομένου σε πιθανότητα) και G σημαίνει συνάρτηση Generator (αντιστοίχιση λανθάνοντος διανύσματος σε ένα δεδομένο). Έτσι ο Discriminator προσπαθεί να ελαχιστοποιήσει την πιθανότητα να μπερδέψει ένα πραγματικό δείγμα με ένα πλαστό (πρώτο μέλος συνάρτησης) ή ένα πλαστό δείγμα για ένα πραγματικό (το δεύτερο μέλος συνάρτησης).

$$J^G = -J^D \quad \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)$$

,

Σχήμα 7-7. Συνάρτηση απώλειας του Generator
(Πηγή: GANs in Action Deep Learning with Generative Adversarial Networks, 2019),
(Πηγή ιστοσελίδα: https://jonathan-hui.medium.com/gan-wasserstein-gan-wgan-gp-6a1a2aa1b490)

Επειδή έχουμε μόνο δύο "παίκτες" και ανταγωνίζονται ο ένας τον άλλον, είναι λογικό ότι η απώλεια του Generator θα ήταν αρνητική έναντι του Discriminator. Συνδυάζοντας τα όλα μαζί, έχουμε δύο συναρτήσεις απώλειας και η μία είναι η αρνητική τιμή της άλλης. Η αντιπαλότητα φαίνεται ξεκάθαρα.

Το Wasserstein GAN

Το Wasserstein GAN, ή WGAN για συντομία, εισήχθη το 2017. Είναι μια επέκταση του GAN που αναζητά έναν εναλλακτικό τρόπο εκπαίδευσης του Δημιουργού για την καλύτερη προσέγγιση της κατανομής των δεδομένων που παρατηρούνται σε ένα δεδομένο σύνολο δεδομένων εκπαίδευσης. Αντί να χρησιμοποιεί έναν Διευκρινιστή για να ταξινομήσει ή να προβλέψει την πιθανότητα των δημιουργούμενων εικόνων ως πραγματικών ή πλαστών, το WGAN αλλάζει ή αντικαθιστά το μοντέλο Διευκρινιστή με έναν Κριτικό (Critic), που βαθμολογεί την πραγματική ή την ψεύτικη εικόνα μιας δεδομένης εικόνας. Αυτή η αλλαγή υποκινείται από ένα μαθηματικό επιχείρημα ότι η εκπαίδευση του Δημιουργού πρέπει να επιδιώκει την ελαχιστοποίηση της απόστασης μεταξύ της κατανομής των δεδομένων που παρατηρείται στο σύνολο δεδομένων εκπαίδευσης και της κατανομής που παρατηρείται στα

παραγόμενα παραδείγματα. Αποδεικνύεται ότι ένα νευρωνικό δίκτυο Critic μπορεί να εκπαιδευτεί ώστε να προσεγγίζει την απόσταση Wasserstein και, με τη σειρά του, να χρησιμοποιηθεί για την αποτελεσματική εκπαίδευση ενός μοντέλου Δημιουργού. Είναι σημαντικό ότι η απόσταση Wasserstein έχει τις ιδιότητες ότι είναι συνεχής και διαφοροποιήσιμη και συνεχίζει να παρέχει μια γραμμική κλίση, ακόμη και αφού ο Κριτικός είναι καλά εκπαιδευμένος. Το πλεονέκτημα του WGAN είναι ότι η διαδικασία εκπαίδευσης είναι πιο σταθερή και λιγότερο ευαίσθητη στην αρχιτεκτονική του μοντέλου και στην επιλογή των υπερ-παραμέτρων. Ίσως το πιο σημαντικό, η απώλεια του Διευκρινιστή φαίνεται να σχετίζεται με την ποιότητα των εικόνων που δημιουργούνται από τον Δημιουργό. Συγκεκριμένα, όσο μικρότερη είναι η απώλεια του Critic κατά την αξιολόγηση των παραγόμενων εικόνων, τόσο υψηλότερη είναι η αναμενόμενη ποιότητα των δημιουργούμενων εικόνων. Αυτό είναι σημαντικό καθώς σε αντίθεση με άλλα GAN που αναζητούν σταθερότητα όσον αφορά την εύρεση μιας ισορροπίας μεταξύ δύο μοντέλων, το WGAN επιδιώκει τη σύγκλιση, μειώνοντας τις απώλειες του Δημιουργού. Το WGAN χρησιμοποιεί το earth mover's distance[9] ως συνάρτηση απώλειας που σαφώς συσχετίζεται με την οπτική ποιότητα των δειγμάτων που παράγονται.

9 earth mover's distance (EMD): Είναι ένα μέτρο της απόστασης μεταξύ δύο κατανομών πιθανότητας σε μια περιοχή D

WGAN $\nabla_w \frac{1}{m}\sum_{i=1}^{m}\left[f\left(x^{(i)}\right) - f\left(G\left(z^{(i)}\right)\right)\right]$ $\nabla_\theta \frac{1}{m}\sum_{i=1}^{m} f\left(G\left(z^{(i)}\right)\right)$

Σχήμα 7-8. Συνάρτηση απώλειας WGAN και του Generator(Πηγή ιστοσελίδα: https://jonathan-hui.medium.com/gan-wasserstein-gan-wgan-gp-6a1a2aa1b490

Συνολικά, το WGAN χρησιμοποιείται ευρέως και έχει γίνει πρότυπο σε μεγάλο μέρος της έρευνας και πρακτικής των GAN.

3. Τεχνικές Εκπαίδευσης

Αυτές είναι απλές τεχνικές, που συχνά πρέπει απλώς να τις δοκιμάσουμε για να δούμε εάν λειτουργούν στο μοντέλο μας.

- Gradient Penalty

  Η ιδέα του Gradient Penalty είναι να επιβάλει έναν περιορισμό ή μια ποινή στις εξόδους του Διευκρινιστή, με αποτέλεσμα η εκπαίδευση να είναι πιο σταθερή.

- Εκπαίδευση του Διευκρινιστή περισσότερο

  Το να εκπαιδεύσουμε περισσότερο τον Διευκρινιστή είναι μια προσέγγιση με μεγάλη επιτυχία. Συνήθως υπάρχουν περισσότερες ενημερώσεις για το Discriminator ανά κύκλο εκπαίδευσης. Μια κοινή αναλογία είναι πέντε ενημερώσεις βάρους του Discriminator ανά μία του Generator. Επίσης, η προεκπαίδευση του Discriminator πριν καν ο Generator έχει την ευκαιρία να παράγει οτιδήποτε είναι επίσης μια καλή προσέγγιση.

- Αποφυγεί των αραιών κλίσεων

  Είναι διαισθητικά λογικό ότι οι αραιές κλίσεις (όπως αυτές που παράγονται από το ReLU ή το MaxPool) θα κάνουν την εκπαίδευση πιο δύσκολη, διότι εστιάζουν στην αποκοπή/αφαίρεση δεδομένων. Υπάρχουν απλές λύσεις εδώ, όπως η χρήση του Leaky ReLU και η αποφυγή του Pooling. Συνολικά, προσπαθούμε να ελαχιστοποιήσουμε την απώλεια πληροφοριών και να κάνουμε τη ροή των πληροφοριών όσο πιο λογική μπορεί να είναι.

7.8 Περίληψη των GAN

Συνοψίζοντας, τα GANs αποτελούνται από δύο νευρωνικά δίκτυα, το ένα ονομάζεται Generator και το άλλο Discriminator. Αυτά τα δύο δίκτυα εκπαιδεύονται ταυτόχρονα, αλλά με μια ανταγωνιστική δυναμική μεταξύ τους, με σκοπό το κάθε δίκτυο ξεχωριστά να ξεγελάσει και να υπερέχει από το άλλο. Ο Generator προσπαθεί να δημιουργήσει πλαστά δεδομένα μέσα από τον λανθάνοντα χώρο, τα οποία είναι δυσδιάκριτα από τα δεδομένα εκπαίδευσης. Αντίθετα, ο Discriminator, προσπαθεί να διακρίνει τα πλαστά από τα πραγματικά δεδομένα εκπαίδευσης. Η έξοδος του Discriminator, είναι μια πιθανότητα που υποδεικνύει εάν τα δεδομένα φαίνονται πλαστά ή όχι. Με οδηγό αυτή την πιθανότητα, ο Generator προσπαθεί να βελτιώσει τα πλαστά δεδομένα που παράγει, έως ότου αυτή η πιθανότητα φτάσει στο ελάχιστο της, τότε λέμε ότι το δίκτυο μας έχει φτάσει σε ισορροπία. Όμως, η ισορροπία αυτή είναι πολύ δύσκολο να βρεθεί καθώς το συνολικό δίκτυο αποτελείται από πολλές παραμέτρους και εκπαιδεύσιμα βάρη. Η αξιολόγηση και τα κριτήρια παύσης της εκπαίδευσης είναι ένα δύσκολο θέμα για τα μοντέλα παραγωγής. Υπάρχουν διάφορες τεχνικές εκπαίδευσης, κάποιες μαθηματικά εξηγήσιμες και κάποιες περισσότερο εμπειρικές όπου βοηθούν στην ομαλότερη και αποδοτικότερη εκπαίδευση. Οι σημαντικότερες από αυτές είναι η χρήση του WGAN και της συνάρτησης απώλειας του, η προσθήκη βάθους δικτύου και η προεκπαίδευση του Discriminator.

8. Η εξέλιξη των GANs

8.1 Ταξινομία των GANs

Όπως αναφέρθηκε προηγουμένως, τα GAN είναι εμφανώς δύσκολο να εκπαιδευτούν και να αξιολογηθούν σωστά. Κατά τη διάρκεια των χρόνων από τότε που πρωτο-εφευρέθηκαν για πρώτη φορά τα GAN, πολλοί ερευνητές έχουν βρει πολλούς έξυπνους τρόπους για να μετριάσουν αυτά τα προβλήματα. Έχουν φτάσει τόσο μακριά που έχουν εφεύρει νέες και καλύτερες παραλλαγές των GAN ανάλογα με το πρόβλημα που έχουν προγραμματιστεί να λύσουν. Θα ρίξουμε μια πιο προσεκτική ματιά σε μερικές από τις πιο σημαντικές παραλλαγές.

Deep Convolutional GAN (DCGAN)

Τα DCGAN είναι πολύ σημαντικό ορόσημο στην ιστορία των GAN. Διότι ήταν από τις πρώτες πετυχημένες προσπάθειες παραλλαγής των GAN, χρησιμοποιώντας την τεχνολογία των νευρωνικών συνελίξεων. Πριν προχωρήσουμε παραπέρα, μια σύντομη εισήγηση στα συνελικτικά νευρωνικά δίκτυα.

8.1.1 Συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks)

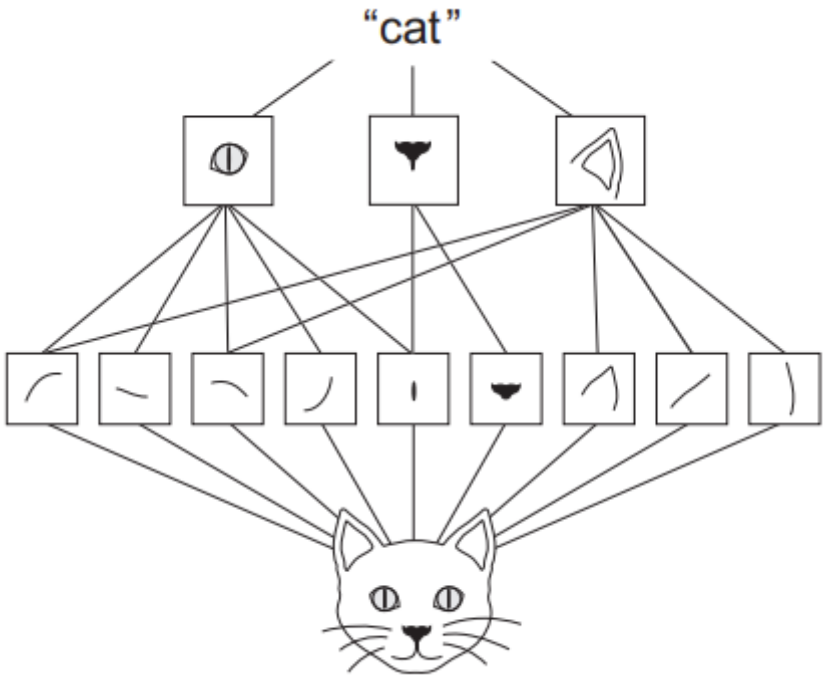
8.1.1.1 Συνελικτικά φίλτρα

Σε αντίθεση με ένα κανονικό νευρωνικό δίκτυο του οποίου οι νευρώνες είναι διατεταγμένοι σε επίπεδα, πλήρως συνδεδεμένα στρώματα, τα στρώματα σε ένα CNN είναι διατεταγμένα σε τρεις διαστάσεις (πλάτος × ύψος × βάθος). Οι συνελίξεις πραγματοποιούνται ολισθαίνοντας ένα ή περισσότερα φίλτρα πάνω από το επίπεδο εισόδου. Κάθε φίλτρο έχει ένα σχετικά μικρό πεδίο (πλάτος × ύψος), αλλά εκτείνεται πάντα σε όλο το βάθος της εισόδου. Σε κάθε βήμα καθώς ολισθαίνει κατά μήκος της εισόδου, κάθε φίλτρο εξάγει μια ενιαία τιμή ενεργοποίησης, το εσωτερικό γινόμενο μεταξύ των τιμών εισόδου και των καταχωρήσεων φίλτρου. Αυτή η διαδικασία έχει ως αποτέλεσμα έναν δισδιάστατο χάρτη ενεργοποίησης για κάθε φίλτρο. Οι χάρτες ενεργοποίησης που παράγονται από κάθε φίλτρο στη συνέχεια στοιβάζονται ο ένας πάνω στον άλλο για να παραχθεί ένα τρισδιάστατο στρώμα εξόδου, το βάθος εξόδου είναι ίσο με τον αριθμό των φίλτρων.

Μια βασική διαφορά μεταξύ αυτών των δύο διαφορετικών τύπων επιπέδων είναι ότι τα συνελικτικά στρώματα μαθαίνουν τοπικά μοτίβα. Αυτό το βασικό χαρακτηριστικό δίνει στα CNN δύο ενδιαφέρουσες ιδιότητες:

- Τα μοτίβα που μαθαίνουν είναι μεταφραστικά αμετάβλητα. Δηλαδή, αφού μάθει ένα συγκεκριμένο μοτίβο στην κάτω δεξιά γωνία μιας εικόνας, ένα CNN μπορεί να το αναγνωρίσει οπουδήποτε, για παράδειγμα, στην επάνω αριστερή γωνία. Ένα απλό βαθύ νευρωνικό δίκτυο θα έπρεπε να μάθει ξανά το μοτίβο εάν εμφανιζόταν σε μια νέα τοποθεσία. Αυτό καθιστά τα δεδομένα CNN αποτελεσματικά κατά την επεξεργασία εικόνων (επειδή ο οπτικός κόσμος είναι ουσιαστικά αμετάβλητος στη μετάφραση): χρειάζονται λιγότερα δείγματα εκπαίδευσης για να μάθουν αναπαραστάσεις που έχουν ισχύ γενίκευσης.

- Μπορούν να μάθουν χωρικές ιεραρχίες προτύπων (βλ. εικόνα 8-1). Ένα πρώτο επίπεδο συνέλιξης θα μάθει μικρά τοπικά μοτίβα, όπως άκρες, ένα δεύτερο επίπεδο συνέλιξης θα μάθει μεγαλύτερα μοτίβα που κατασκευάζονται από τα χαρακτηριστικά των πρώτων στρωμάτων και ούτω καθεξής. Αυτό επιτρέπει στα CNN να μαθαίνουν αποτελεσματικά όλο και πιο περίπλοκες και αφηρημένες οπτικές έννοιες.



Σχήμα 8-1. Παράδειγμα χωρικής ιεραρχίας με μια εικόνα γάτας
(Πηγή: Deep Learning with Python, 2018)

Οι συνελίξεις ορίζονται από δύο βασικές παραμέτρους:

- Μέγεθος των φίλτρων που εξάγονται από τις εισόδους: Αυτές είναι συνήθως 3 × 3 ή 5 × 5.
- Βάθος του χάρτη χαρακτηριστικών εξόδου: Ο αριθμός των φίλτρων που υπολογίζεται από τη συνέλιξη.

8.1.1.2 Κατανόηση των Εφέ Συνόρων (Border Effects) και του Padding

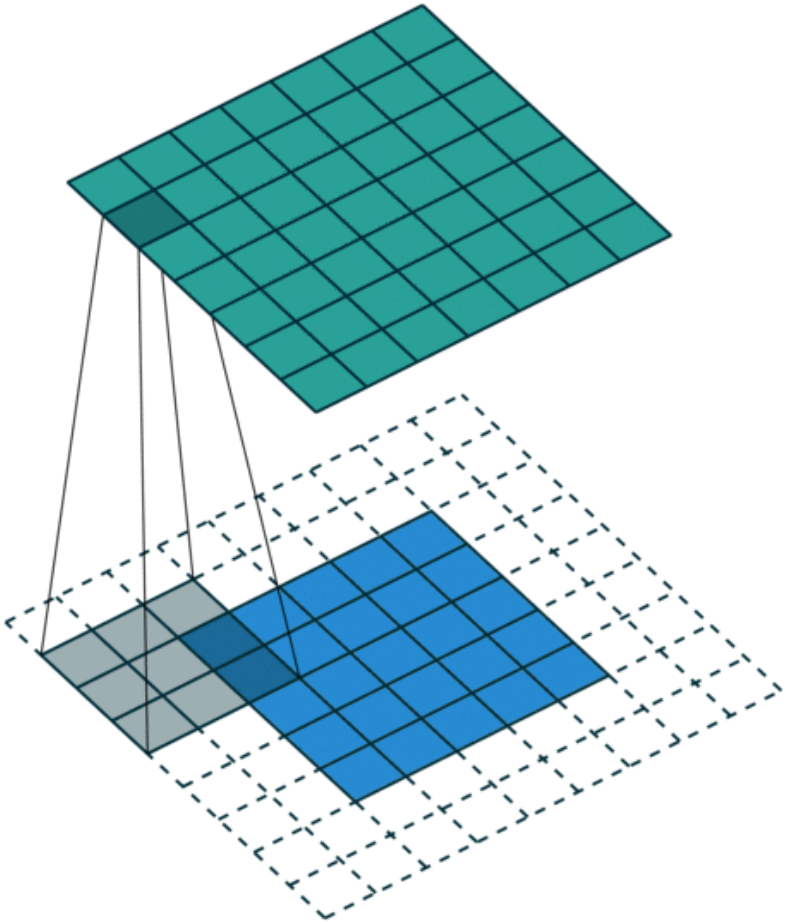
Η μείωση του μεγέθους της εισόδου αναφέρεται ως εφέ συνόρων. Προκαλείται από την αλληλεπίδραση του φίλτρου με το περίγραμμα της εικόνας.

Από προεπιλογή, ένα φίλτρο ξεκινά στα αριστερά της εικόνας με την αριστερή πλευρά του φίλτρου να βρίσκεται στα αριστερά pixel της εικόνας. Στη συνέχεια, το φίλτρο τοποθετείται κατά μήκος της εικόνας μία στήλη τη φορά έως ότου η δεξιά πλευρά του φίλτρου βρίσκεται στα δεξιά pixel της εικόνας όπου ολοκληρώνεται η διαδικασία.

Μια εναλλακτική προσέγγιση για την εφαρμογή ενός φίλτρου σε μια εικόνα είναι να διασφαλιστεί ότι κάθε pixel στην εικόνα έχει την ευκαιρία να βρίσκεται στο κέντρο του φίλτρου. Αυτή η εναλλακτική πηγάζει ένα πρόβλημα όμως, καθώς τα pixel στην άκρη της εισόδου εκτίθενται μόνο στην άκρη του φίλτρου. Ξεκινώντας το φίλτρο έξω από το πλαίσιο

της εικόνας, δίνει στα pixel στο περίγραμμα της εικόνας περισσότερη ευκαιρία για αλληλεπίδραση με το φίλτρο, μεγαλύτερη ευκαιρία για τον εντοπισμό χαρακτηριστικών από το φίλτρο και με τη σειρά του, μια έξοδος που έχει το ίδιο σχήμα με την εικόνα εισόδου.

Για παράδειγμα, στην περίπτωση εφαρμογής ενός φίλτρου 3×3 σε εικόνα εισόδου 5×5, μπορούμε να προσθέσουμε ένα περίγραμμα ενός pixel γύρω από το εξωτερικό της εικόνας. Αυτό έχει ως αποτέλεσμα τη δημιουργία τεχνητής εικόνας εισόδου 9×9. Όταν εφαρμόζεται το φίλτρο 3×3, προκύπτει ένας χάρτης χαρακτηριστικών 7×7. Οι προστιθέμενες τιμές pixel θα μπορούσαν να έχουν την τιμή μηδέν που δεν επηρεάζει τη λειτουργία του εσωτερικού γινομένου όταν εφαρμόζεται το φίλτρο, η διαδικασία αυτή ονομάζεται padding.

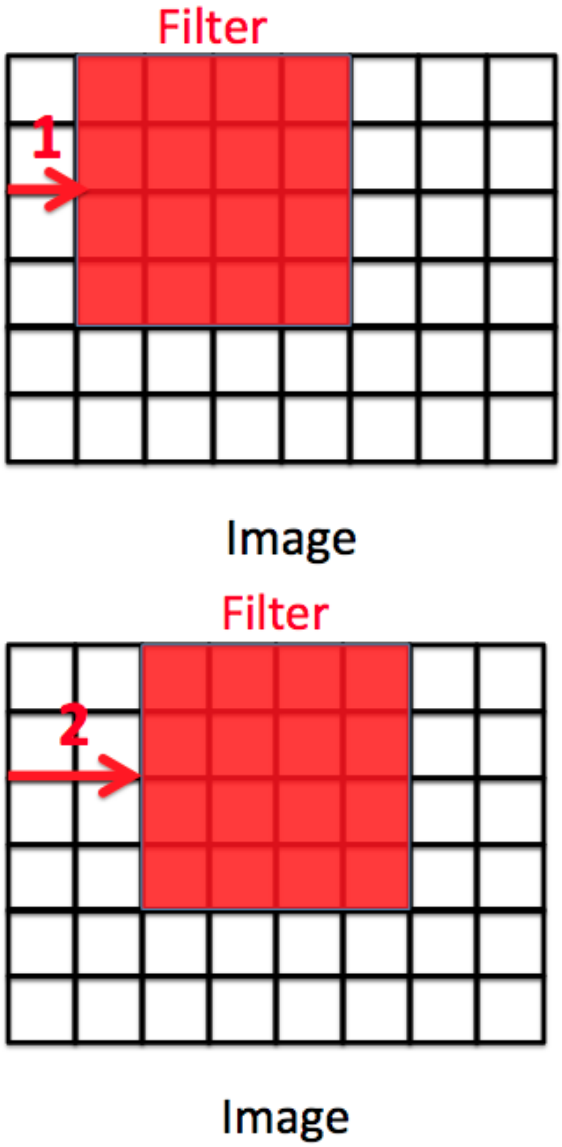


Σχήμα 8-2. Η διαδικασία padding με φίλτρο 3×3 σε εικόνα εισόδου 5×5
(Πηγή ιστοσελίδα: https://theano-pymc.readthedocs.io/en/latest/tutorial/conv_arithmetic.html)

8.1.1.3 Κατανόηση των Διασκελισμών (Strides)

Ο άλλος παράγοντας που μπορεί να επηρεάσει το μέγεθος της παραγωγής είναι η έννοια των stride, το μέγεθος της κίνησης μεταξύ των εφαρμογών του φίλτρου στην εικόνα εισόδου και είναι σχεδόν πάντα συμμετρικός σε διαστάσεις ύψους και πλάτους. Ο προεπιλεγμένος βηματισμός είναι το 1. Το stride μπορεί να αλλάξει, γεγονός που επηρεάζει τόσο τον τρόπο εφαρμογής του φίλτρου στην εικόνα όσο και, με τη σειρά του, το μέγεθος του χάρτη χαρακτηριστικών που προκύπτει. Για παράδειγμα, μπορούμε να αλλάξουμε το μέγεθος του

stride σε 2, που σημαίνει ότι το πλάτος και το ύψος του χάρτη χαρακτηριστικών μειώνονται κατά συντελεστή 2 (επιπλέον των αλλαγών που προκαλούνται από εφέ περιγράμματος).



Σχήμα 8-3. Απεικόνιση του stride με βήμα 1 (αριστερά) και βήμα 2 (δεξιά)
(Πηγή ιστοσελίδα: https://medium.com/machine-learning-algorithms/what-is-stride-in-convolutional-neural-network-e3b4ae9baedb)

8.1.2.1 Σύντομο ιστορικό του DCGAN

Το DCGAN, το οποίο εισήχθη το 2016 από τους Alec Radford, Luke Metz και Soumith Chintala, σηματοδότησε μια από τις πιο σημαντικές πρώιμες καινοτομίες στα GAN από την έναρξη της τεχνικής δύο χρόνια νωρίτερα. Η χρήση των CNNs επιδεινώνει πολλές από τις δυσκολίες που μαστίζουν την εκπαίδευση GAN, συμπεριλαμβανομένης της αστάθειας και του κορεσμού κλίσης. Με το DCGAN, ο Radford και οι συνεργάτες του εισήγαγαν τεχνικές και βελτιστοποιήσεις που επέτρεψαν στα CNNs να κλιμακωθούν στο πλήρες πλαίσιο των GAN χωρίς την ανάγκη τροποποίησης της υποκείμενης αρχιτεκτονικής. Μία από τις βασικές τεχνικές που ο Radford χρησιμοποιείσαι ήταν το batch normalization, το οποίο βοηθά στη

σταθεροποίηση της εκπαιδευτικής διαδικασίας ομαλοποιώντας τις εισόδους σε κάθε στρώμα όπου εφαρμόζεται.

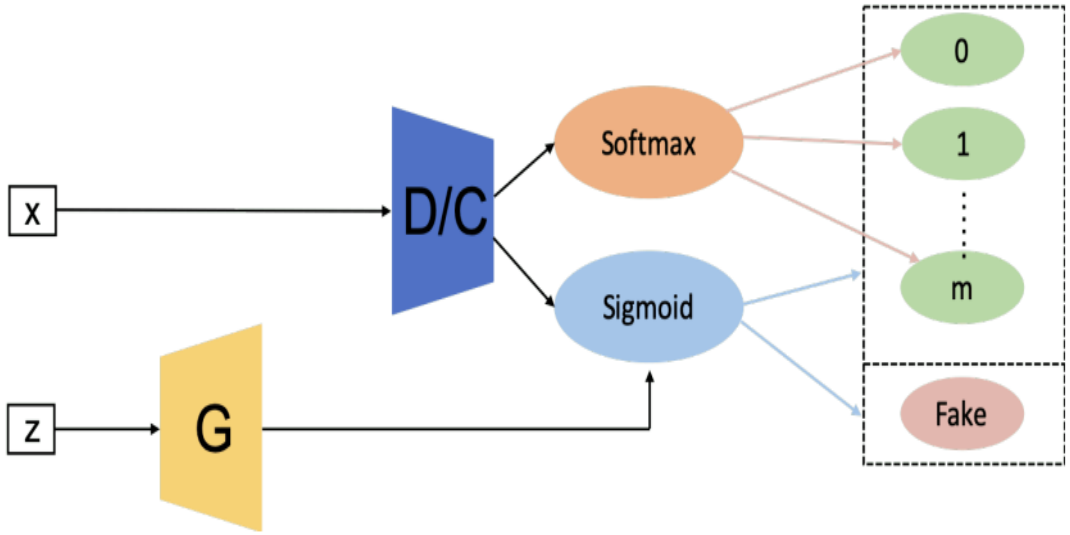
8.1.2.2 Ομαλοποίηση παρτίδας (Batch normalization)

Η ομαλοποίηση παρτίδων εισήχθη από τους επιστήμονες της Google Sergey Ioffe και Christian Szegedy το 2015. Η διορατικότητα τους ήταν τόσο απλή όσο και πρωτοποριακή. Ακριβώς όπως ομαλοποιούμε τις εισόδους δικτύου, πρότειναν να ομαλοποιηθούν οι είσοδοι σε κάθε επίπεδο, για κάθε μίνι-παρτίδα εκπαίδευσης καθώς ρέει μέσω του δικτύου. Ομαλοποίηση είναι η κλιμάκωση των δεδομένων έτσι ώστε να έχουν μηδενική μέση και μοναδιαία διακύμανση (zero mean and unit variance). Αυτό επιτυγχάνεται παίρνοντας κάθε σημείο δεδομένων x, αφαιρώντας το μέσο μ, και διαιρώντας το αποτέλεσμα με την τυπική απόκλιση.

Η ομαλοποίηση έχει πολλά πλεονεκτήματα. Ίσως το πιο σημαντικό, διευκολύνει τις συγκρίσεις μεταξύ χαρακτηριστικών με πολύ διαφορετικές κλίμακες και, κατ' επέκταση, καθιστά τη διαδικασία εκπαίδευσης λιγότερο ευαίσθητη στην κλίμακα των χαρακτηριστικών. Καθώς οι τιμές εισόδου ρέουν μέσω του δικτύου, από το ένα επίπεδο στο άλλο, κλιμακώνονται από τις εκπαιδεύσιμες παραμέτρους σε καθένα από αυτά τα επίπεδα. Και καθώς οι παράμετροι ρυθμίζονται από την backpropagation, η κατανομή των εισόδων κάθε επιπέδου είναι επιρρεπής σε αλλαγές στις επόμενες επαναλήψεις εκπαίδευσης, γεγονός που αποσταθεροποιεί τη διαδικασία εκμάθησης. Η ομαλοποίηση παρτίδας το επιλύει με την κλιμάκωση των τιμών σε κάθε μίνι παρτίδα με το μέσο όρο και τη διακύμανση αυτής της μίνι παρτίδας.

Semi-supervised GAN (SGAN)

Το SGAN προτείνεται στο πλαίσιο της ημι-εποπτευόμενης μάθησης. Η ημι-εποπτευόμενη μάθηση είναι ένα πολλά υποσχόμενο ερευνητικό πεδίο μεταξύ της εποπτευόμενης και της μη εποπτευόμενης μάθησης. Η ημι-εποπτευόμενη μάθηση έχει ετικέτες για ένα μικρό υποσύνολο δεδομένων. Σε σύγκριση με τα αρχικά GAN, ο Discriminator του SGAN είναι πολλαπλών εξόδων, έχει softmax[10] και sigmoid[8] για την ταξινόμηση των πραγματικών δεδομένων και τη διάκριση πραγματικών και πλαστών δειγμάτων αντίστοιχα. Τα αποτελέσματα δείχνουν ότι τόσο ο Discriminator όσο και ο Generator στο SGAN είναι βελτιωμένοι σε σύγκριση με το αρχικό GAN.

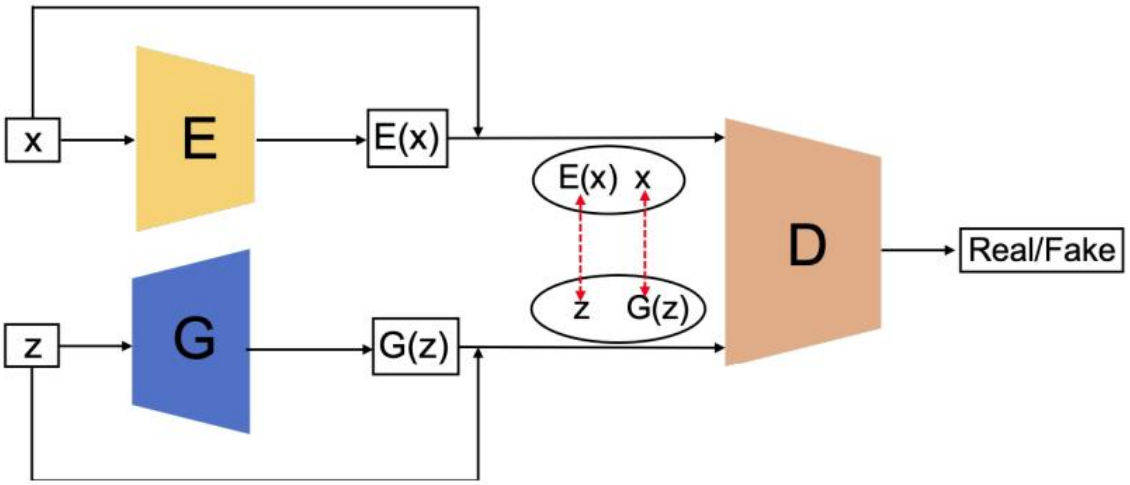


Σχήμα 8-4. Αρχιτεκτονική SGAN
(Πηγή: Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy, 2020)

10 softmax και sigmoid: Συναρτήσεις εξόδου

Bidirectional GAN (BiGAN)

Τα παραδοσιακά GAN δεν έχουν μέσα εκμάθησης της αντίστροφης χαρτογράφησης, προβάλλοντας δεδομένα πίσω στον λανθάνοντα χώρο. Το BiGAN έχει σχεδιαστεί για αυτό το σκοπό. Η συνολική αρχιτεκτονική αποτελείται από Encoder (E), Generator(G) και Discriminator (D). Το E κωδικοποιεί τα πραγματικά δείγματα δεδομένων στο E(x) ενώ το G αποκωδικοποιεί το z σε G(z). Ως αποτέλεσμα, το D στοχεύει να αξιολογήσει τη διαφορά μεταξύ κάθε ζεύγους των (E(x), x) και (G(z), z). Καθώς το E και το G δεν επικοινωνούν απευθείας, το E δεν βλέπει ποτέ το G(z) και το G ποτέ το E(x). Ουσιαστικά, ο κωδικοποιητής και ο αποκωδικοποιητής πρέπει να μάθουν να αναστρέφουν ο ένας τον άλλον για να ξεγελάσουν τον Discriminator.

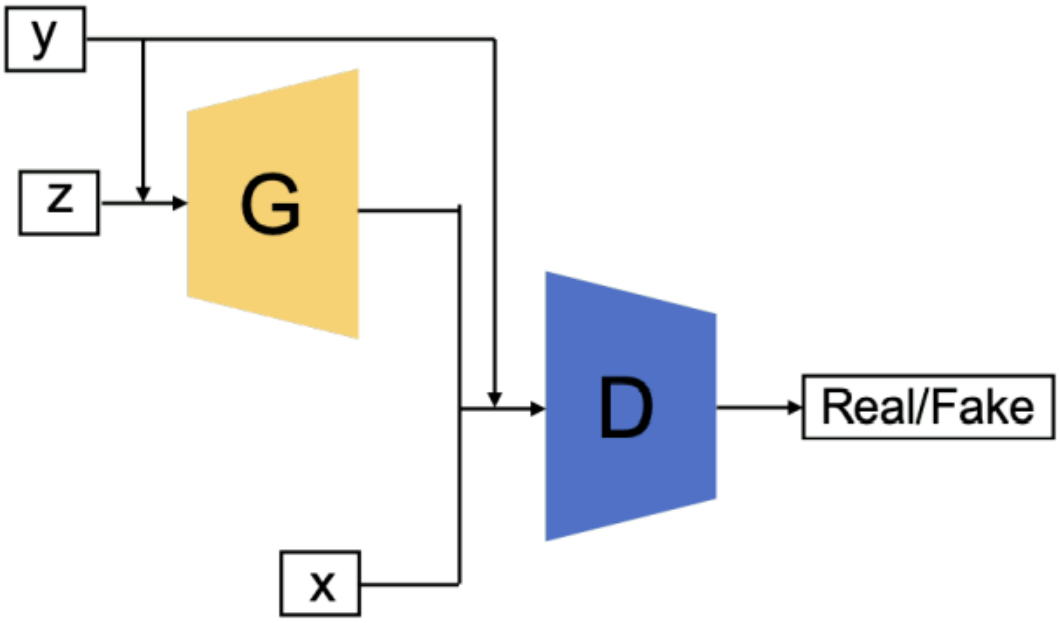


Σχήμα 8-5. Αρχιτεκτονική BiGAN
(Πηγή: Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy, 2020)

Conditional GAN (CGAN)

Το CGAN εισάγεται με τη ρύθμιση τόσο του Discriminator όσο και του Generator με τροφοδοσία ετικετών κλάσης. Το CGAN τροφοδοτεί την πρόσθετη πληροφορία y (ετικέτα κλάσης) τόσο στον Discriminator όσο και στον Generator. Θα πρέπει να σημειωθεί ότι το y κωδικοποιείται κανονικά μέσα στον Generator και τον Discriminator πριν συνδεθεί με το κωδικοποιημένο z και κωδικοποιημένο x. Για παράδειγμα, τόσο το z όσο και το y αντιστοιχίζονται σε κρυφά στρώματα με μεγέθη στρώματος 200 και 1000 αντίστοιχα πριν συνδυαστούν μεταξύ τους (η συνδυασμένη διάσταση του στρώματος είναι 200 + 1000 = 1200) στον Generator. Με αυτόν τον τρόπο, το CGAN ενισχύει τη διακριτική ικανότητα του χρήστη. Η συνάρτηση απώλειας του CGAN είναι ελαφρώς διαφορετική από την αρχική GAN, στην οποία τα x και y εξαρτώνται από το z. Επωφελούμενο από τις επιπλέον κωδικοποιημένες πληροφορίες y, το CGAN δεν είναι μόνο σε θέση να χειρίζεται μονοτροπικά αλλά και πολυτροπικά σύνολα δεδομένων.

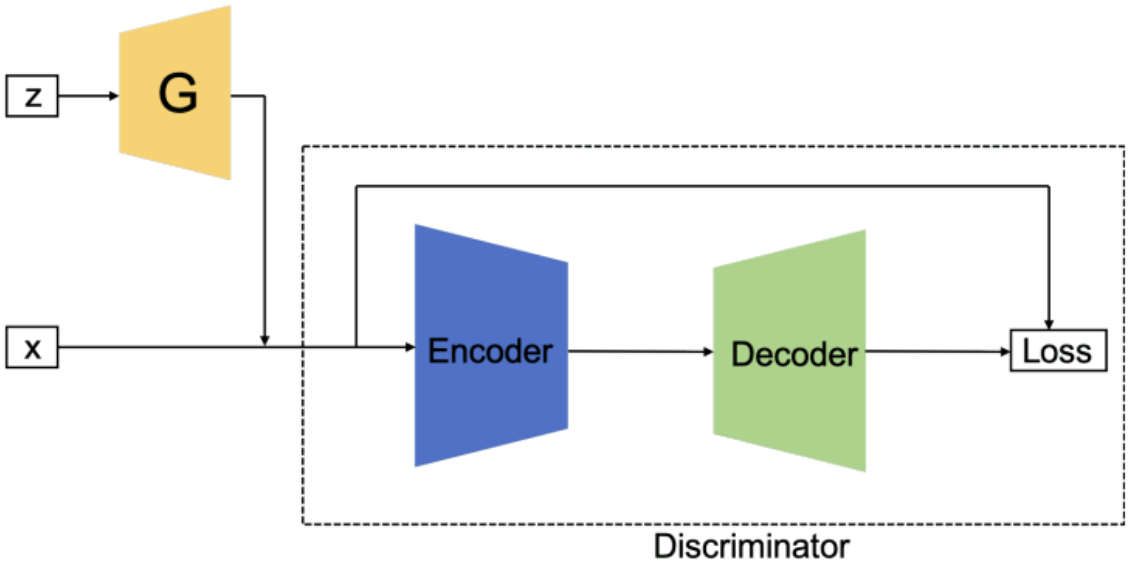
$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_r} \log[D(\mathbf{x}|\mathbf{y})] + \mathbb{E}_{\mathbf{z}\sim p_\mathbf{z}} \log\left[1 - D(G(\mathbf{z}|\mathbf{y}))\right]$$



Σχήμα 8-6. Αρχιτεκτονική CGAN
(Πηγή: Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy, 2020)

Boundary Equilibrium GAN (BEGAN)

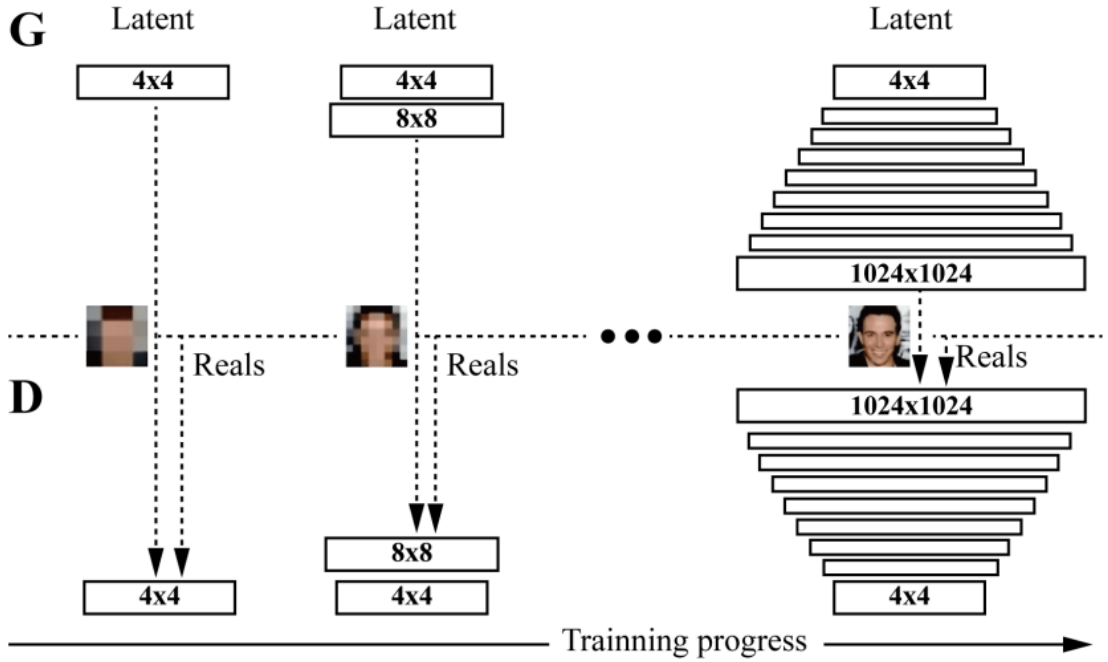
Το BEGAN χρησιμοποιεί μια αρχιτεκτονική αυτόματου κωδικοποιητή για τον Discriminator. Η απώλεια του αυτόματου κωδικοποιητή μπορεί να δημιουργηθεί αντίστοιχα για τα G και D. Κατά την εκπαίδευση του αυτόματου κωδικοποιητή (D), ο στόχος είναι να μεγιστοποιηθεί η απώλεια ανακατασκευής πραγματικών εικόνων και να μεγιστοποιηθεί η απώλεια ανακατασκευής για τις παραγόμενες εικόνες. Σε σύγκριση με την παραδοσιακή βελτιστοποίηση, το BEGAN ταιριάζει με τις κατανομές απώλειας του αυτόματου κωδικοποιητή χρησιμοποιώντας μια απώλεια που προέρχεται από την απόσταση Wasserstein αντί να αντιστοιχίζει απευθείας τις κατανομές δεδομένων. Αυτή η τροποποίηση βοηθά τον G να δημιουργήσει εύκολα ανακατασκευή δεδομένη για τον αυτόματο κωδικοποιητή στην αρχή, επειδή τα δεδομένα που δημιουργούνται είναι κοντά στο 0 και η πραγματική κατανομή δεδομένων δεν έχει μάθει ακόμη με ακρίβεια, γεγονός που εμποδίζει τον D να κερδίσει εύκολα το G στο πρώιμο στάδιο της εκπαίδευσης. Για τον κωδικοποιητή και τον αποκωδικοποιητή, εφαρμόστηκαν εκθετικές γραμμικές μονάδες (ELUs) στις εξόδους τους.



Σχήμα 8-7. Αρχιτεκτονική BEGAN
(Πηγή: Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy, 2020)

Progressive GAN (PROGAN)

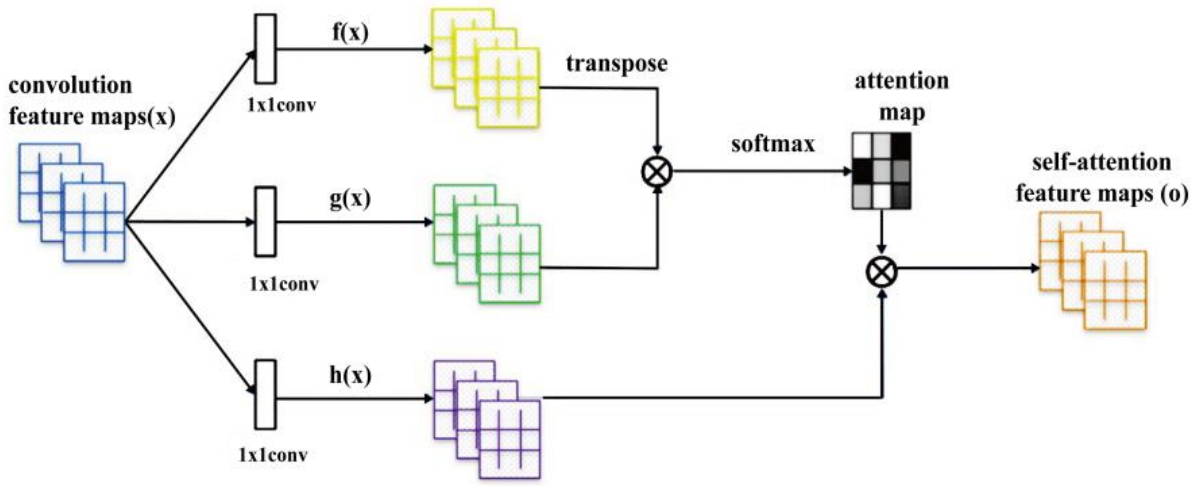
Το PROGAN περιλαμβάνει προοδευτικά βήματα προς την επέκταση της αρχιτεκτονικής του δικτύου. Αυτή η αρχιτεκτονική χρησιμοποιεί την ιδέα των προοδευτικών νευρωνικών δικτύων. Αυτή η τεχνολογία δεν υποφέρει από απομνημόνευση και είναι σε θέση να αναπτύξει προηγούμενες γνώσεις μέσω πλευρικών συνδέσεων σε χαρακτηριστικά που είχαν μάθει προηγουμένως. Κατά συνέπεια, εφαρμόζεται ευρέως για την εκμάθηση σύνθετων ακολουθιών εργασιών. Το σχήμα δείχνει τη διαδικασία εκπαίδευσης για το PROGAN. Η εκπαίδευση ξεκινά με μια εικόνα χαμηλής ανάλυσης 4 × 4 pixel. Τόσο ο G όσο και ο D αρχίζουν να αναπτύσσονται με την εξέλιξη της εκπαίδευσης. Είναι σημαντικό ότι όλες οι μεταβλητές παραμένουν εκπαιδεύσιμες σε όλη αυτή τη διαδικασία ανάπτυξης. Αυτή η προοδευτική στρατηγική εκπαίδευσης επιτρέπει ουσιαστικά πιο σταθερή μάθηση και για τα δύο δίκτυα. Αυξάνοντας σιγά σιγά την ανάλυση, τίθεται συνεχώς στα δίκτυα μια πολύ πιο απλή εργασία σε σύγκριση με τον τελικό στόχο της ανακάλυψης μιας χαρτογράφησης από λανθάνοντα διανύσματα. Όλα τα τρέχοντα υπερσύγχρονα GAN χρησιμοποιούν αυτόν τον τύπο εκπαιδευτικής στρατηγικής και έχει ως αποτέλεσμα εντυπωσιακές, αληθοφανείς εικόνες.



Σχήμα 8-8. Αρχιτεκτονική PROGAN
(Πηγή: Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy, 2020)

Self-Attention GAN (SAGAN)

Τα παραδοσιακά CNN (Convolutional Neural Networks) μπορούν να καταγράφουν μόνο τοπικές χωρικές πληροφορίες και το πεδίο λήψης μπορεί να μην καλύπτει αρκετή δομή, γεγονός που προκαλεί δυσκολία στα GAN που βασίζονται στα CNN στην εκμάθηση συνόλων δεδομένων εικόνων πολλών κλάσεων (για παράδειγμα ImageNet) και ενδέχεται να μετατοπιστούν τα βασικά στοιχεία στις παραγόμενες εικόνες. Έχουν προταθεί μηχανισμοί αυτοπροσοχής για τη διασφάλιση μεγάλων δεκτικών πεδίων και χωρίς να θυσιάζεται η υπολογιστική απόδοση για τα CNN. Η αυτοπροσοχή λειτουργεί συμπληρωματικά των συνελίξεων και βοηθά με την μοντελοποίηση εξαρτήσεων μεγάλης εμβέλειας και πολλαπλών επιπέδων σε όλη την περιφέρεια της εικόνας. Ο Generator που λειτουργεί με αυτοπροσοχή, μπορεί να σχεδιάσει εικόνες στις οποίες οι λεπτομέρειες σε κάθε τοποθεσία της εικόνας είναι προσεκτικά συντονισμένες με λεπτές λεπτομέρειες σε μακρινά σημεία της εικόνας. Επιπλέον, ο Discriminator μπορεί επίσης να επιβάλει με μεγαλύτερη ακρίβεια περίπλοκους γεωμετρικούς περιορισμούς σε όλη την έκταση της εικόνας. Επωφελούμενο από τον μηχανισμό αυτοπροσοχής, το SAGAN είναι σε θέση να μάθει παγκόσμιες, μακράς εμβέλειας εξαρτήσεις για τη δημιουργία εικόνων. Έχει επιτύχει εξαιρετική απόδοση στη δημιουργία εικόνων πολλαπλών κλάσεων με βάση τα σύνολα δεδομένων ImageNet.



Σχήμα 8-9. Αρχιτεκτονική SAGAN
(Πηγή: Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy, 2020)

BigGAN

Το BigGAN έχει επίσης επιτύχει κορυφαίες επιδόσεις στα σύνολα δεδομένων ImageNet. Ο σχεδιασμός του βασίζεται στο SAGAN και έχει αποδείξει ότι η απόδοση μπορεί να ωφεληθεί με την κλιμάκωση της εκπαίδευσης GAN, δηλαδή την αύξηση του αριθμού των καναλιών για κάθε επίπεδο και την αύξηση του μεγέθους της παρτίδας. Στο BigGAN, μια κατανομή Gauss χρησιμοποιείται κατά τη διάρκεια της εκπαίδευσης και μια περικομμένη Gaussian χρησιμοποιείται κατά τη διάρκεια της εξαγωγής συμπερασμάτων. Αυτό το τέχνασμα περικοπής παρέχει μια αντιστάθμιση μεταξύ ποιότητας ή πιστότητας εικόνας και ποικιλίας εικόνας. Ένα πιο στενό εύρος δειγματοληψίας οδηγεί σε καλύτερη ποιότητα, ενώ ένα μεγαλύτερο εύρος δειγματοληψίας έχει ως αποτέλεσμα μεγαλύτερη ποικιλία στις δειγματοληπτικές εικόνες. Συνοπτικά η αρχιτεκτονική του BigGAN: 1) Χρησιμοποιεί μοντέλο αυτοπροσοχής εμπνευσμένο από το SAGAN. 2) Οι πληροφορίες κλάσης παρέχονται στον Generator μέσω κανονικοποίησης παρτίδας κατηγορίας 3) Ο Discriminator εκπαιδεύεται διπλάσια από ότι ο Generator 4) Προτού δημιουργηθούν οι εικόνες για αξιολόγηση, υπολογίζονται τα βάρη του μοντέλου σε σχέση με τον μέσο όρο των προηγούμενων επαναλήψεων. 5) Ορισμένες μετατροπές στο δίκτυο, όπως ορθογώνια αρχικοποίηση βάρους και μεγαλύτερο μέγεθος παρτίδας.

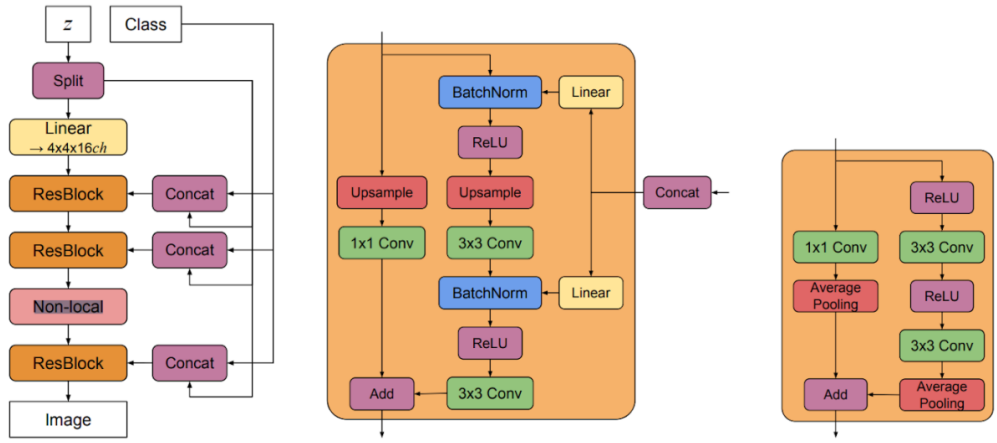
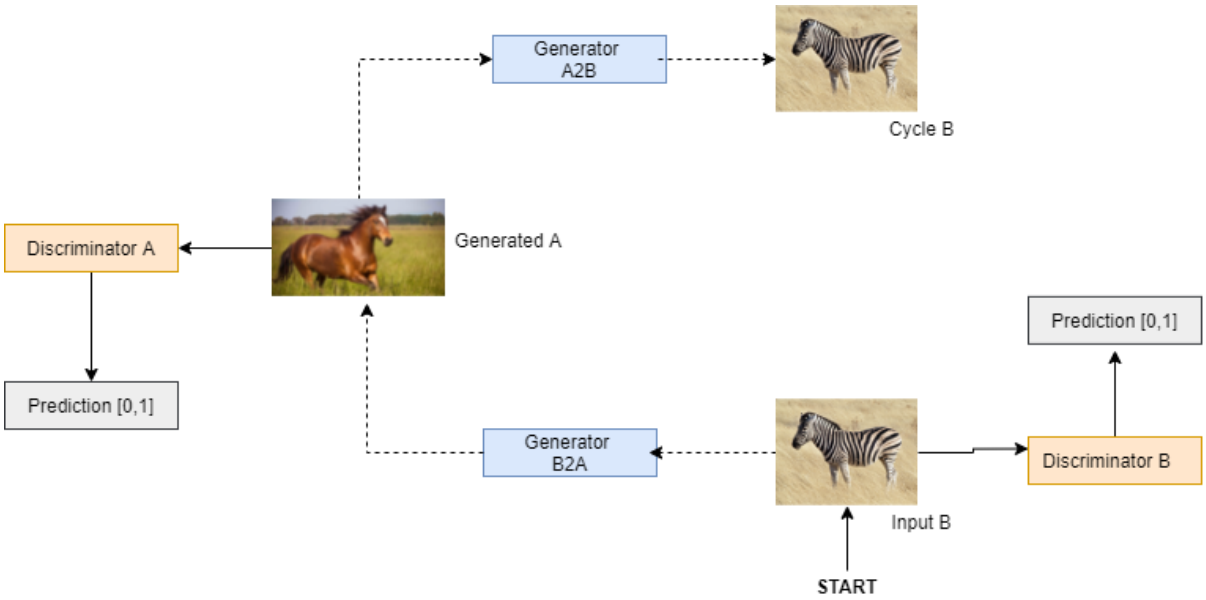


Figure 15: (a) A typical architectural layout for BigGAN's **G**; details are in the following tables. (b) A Residual Block (*ResBlock up*) in BigGAN's **G**. (c) A Residual Block (*ResBlock down*) in BigGAN's **D**.

Σχήμα 8-10. Αρχιτεκτονική BigGAN
(Πηγή ιστοσελίδα: https://paperswithcode.com/method/biggan)

CycleGAN

Το CycleGAN είναι μια επέκταση της αρχιτεκτονικής GAN που περιλαμβάνει την ταυτόχρονη εκπαίδευση δύο μοντέλων Generator και δύο μοντέλων Discriminator. Ο ένας Generator παίρνει εικόνες από τον πρώτο τομέα ως είσοδο και εξάγει εικόνες για τον δεύτερο τομέα και ο άλλος Generator παίρνει εικόνες από τον δεύτερο τομέα ως είσοδο και δημιουργεί εικόνες για τον πρώτο τομέα. Στη συνέχεια, χρησιμοποιούνται μοντέλα Discriminator για να προσδιοριστεί πόσο αληθοφανείς είναι οι δημιουργημένες εικόνες και να ενημερωθούν ανάλογα τα μοντέλα Generator. Αυτή η επέκταση από μόνη της μπορεί να είναι αρκετή για τη δημιουργία εύλογων εικόνων σε κάθε τομέα, αλλά όχι επαρκής για τη δημιουργία μεταφράσεων των εικόνων εισόδου. Το CycleGAN χρησιμοποιεί μια πρόσθετη επέκταση στην αρχιτεκτονική που ονομάζεται συνέπεια κύκλου (cycle consistency). Αυτή είναι η ιδέα ότι μια έξοδος εικόνας από τον πρώτο Generator θα μπορούσε να χρησιμοποιηθεί ως είσοδος στον δεύτερο Generator και η έξοδος του δεύτερου Generator θα πρέπει να ταιριάζει με την αρχική εικόνα. Το αντίστροφο ισχύει επίσης: ότι μια έξοδος από τον δεύτερο Generator μπορεί να τροφοδοτηθεί ως είσοδος στον πρώτο Generator και το αποτέλεσμα θα πρέπει να ταιριάζει με την είσοδο του δεύτερου Generator. Το CycleGAN ενθαρρύνει τη συνοχή του κύκλου προσθέτοντας μια πρόσθετη απώλεια για τη μέτρηση της διαφοράς μεταξύ της παραγόμενης εξόδου του δεύτερου Generator και της αρχικής εικόνας και της αντίστροφης. Αυτό λειτουργεί ως τακτοποίηση των μοντέλων Generator, καθοδηγώντας τη διαδικασία δημιουργίας εικόνας στον νέο τομέα προς τη μετάφραση εικόνας.



Σχήμα 8-11. Αρχιτεκτονική CycleGAN
(Πηγή ιστοσελίδα: https://towardsdatascience.com/cycle-gan-with-pytorch-ebe5db947a99)

8.2 Πρακτικές εφαρμογές GAN

Τα GAN έχουν προχωρήσει πολύ τα τελευταία χρόνια, και οι χρήσεις τους συνεχώς και αυξάνονται. Έχουν μελετηθεί εκτενώς σε πολλές διαφορετικές εφαρμογές, οι οποίες θα αλλάξουν ριζικά πολλές πτυχές της ζωής μας, μερικές από τις κυριότερες εφαρμογές είναι οι εξής:

GANs στην ιατρική

Οι εφαρμογές μηχανικής μάθησης στην ιατρική αντιμετωπίζουν μια σειρά από προκλήσεις που βοηθούν το πεδίο να επωφεληθεί από τα GAN. Ίσως το πιο σημαντικό, είναι η δυσκολία να προμηθεύονται σύνολα δεδομένων εκπαίδευσης αρκετά μεγάλα για εποπτευόμενους αλγόριθμους μηχανικής μάθησης λόγω των δυσκολιών που σχετίζονται με τη συλλογή ιατρικών δεδομένων. Η λήψη δειγμάτων ιατρικών καταστάσεων τείνει να είναι απαγορευτικά δαπανηρή και μη πρακτική. Σε αντίθεση με άλλα σύνολα δεδομένων, δεδομένα ιατρικών καταστάσεων είναι πιο δύσκολο να βρεθούν και συχνά απαιτούν εξειδικευμένο εξοπλισμό για τη συλλογή τους. Μια ακόμη δυσκολία σχετίζεται με το απόρρητο των ασθενών που περιορίζουν τον τρόπο συλλογής και χρήσης ιατρικών δεδομένων. Εκτός από τις δυσκολίες στην απόκτηση ιατρικών συνόλων δεδομένων, είναι επίσης δύσκολο να επισημανθούν σωστά αυτά τα δεδομένα, μια διαδικασία που απαιτεί συχνά σχολιασμούς από άτομα με ειδικές γνώσεις για μια δεδομένη πάθηση. Ως αποτέλεσμα, πολλές ιατρικές εφαρμογές δεν μπορούν να επωφεληθούν από την πρόοδο στη βαθιά μάθηση και την τεχνητή νοημοσύνη. Οι ιατρικοί ερευνητές προσπαθούν να ξεπεράσουν την πρόκληση των ανεπαρκών συνόλων δεδομένων χρησιμοποιώντας τεχνικές αύξησης δεδομένων (data augmentation). Όπως μπορούμε να φανταστούμε, η τυπική αύξηση δεδομένων έχει πολλούς περιορισμούς. Οι μικρές τροποποιήσεις δίνουν παραδείγματα που δεν αποκλίνουν πολύ από την αρχική εικόνα. Ως αποτέλεσμα, τα πρόσθετα παραδείγματα δεν προσθέτουν μεγάλη ποικιλία για να βοηθήσουν τον αλγόριθμο να μάθει να γενικεύει. Στην περίπτωση της ιατρικής διαγνωστικής, θέλουμε διαφορετικά παραδείγματα της ίδιας υποκείμενης παθολογίας. Ο εμπλουτισμός ενός συνόλου δεδομένων με συνθετικά παραδείγματα, όπως αυτά που παράγονται από τα GAN, έχει τη δυνατότητα να εμπλουτίσει περαιτέρω τα διαθέσιμα δεδομένα πέρα από τις παραδοσιακές τεχνικές αύξησης (data augmentation).

Ανακάλυψη και ανάπτυξη φαρμάκων

Η χρήση των GAN για την ανάπτυξη φαρμάκων μπορεί να ακούγεται σαν όνειρο, αλλά τα GAN έχουν ήδη χρησιμοποιηθεί για τη δημιουργία μοριακών αρχιτεκτονικών, δεδομένου ενός επιθυμητού συνόλου χημικών και βιολογικών ιδιοτήτων. Οι ερευνητές μπορούν να εκπαιδεύσουν τον Generator με την υπάρχουσα βάση δεδομένων για να βρεθούν νέες ενώσεις που μπορούν ενδεχομένως να χρησιμοποιηθούν για τη θεραπεία νέων ασθενειών. Οι ερευνητές δεν χρειάζεται να περάσουν χειροκίνητα ολόκληρη τη βάση δεδομένων για να αναζητήσουν ενώσεις που μπορούν να βοηθήσουν στην καταπολέμηση νέων ασθενειών. Ο αλγόριθμος εντοπίζει αυτόματα τέτοιες ενώσεις και βοηθά στη μείωση του χρόνου που απαιτείται για την έρευνα και την ανάπτυξη τέτοιων φαρμάκων. Οι φαρμακευτικές εταιρείες ξοδεύουν δισεκατομμύρια για την έρευνα και την ανάπτυξη νέων φαρμάκων. Τα GAN για την ανάπτυξη φαρμάκων μπορούν να μειώσουν σημαντικά αυτό το κόστος.

Ρομποτική

Η ρομποτική έχει εξίσου μεγάλη επιρροή τις τελευταίες δεκαετίες και πολλές εφαρμογές. Όμως, η εκπαίδευση ενός ανθρωποειδούς ρομπότ για παράδειγμα, περνάει από πολλά στάδια ώστε να λειτουργήσει τελικά σωστά. Το πιο δύσκολο και χρονοβόρο στάδιο είναι η εκπαίδευσή του στο εξωτερικό περιβάλλον, καθώς η φυσική του περιβάλλοντος είναι συχνά πολύ πιο περίπλοκη από τις προσομοιώσεις. Εδώ τα GANs έρχονται να λύσουν αυτό το πρόβλημα. Μπορούν να παράγουν ποικιλόμορφες προσομοιώσεις σε δύο ή και τρεις διαστάσεις, οι οποίες θα φορτώνονται στο ρομπότ το οποίο θα μαθαίνει από μόνο του στο δικό του περιβάλλον. Οπότε δεν θα χρειάζεται τον προγραμματιστή να προσομοιώνει συνεχώς καινούρια περιβάλλοντα. Ύστερα, όταν το ρομπότ θεωρηθεί ικανό να λειτουργήσει στον φυσικό κόσμο θα είναι ήδη αρκετά έτοιμο και έτσι, θα μειωθεί κατά πολύ ο χρόνος εκπαίδευσης του. Επίσης, οι έτοιμες προσομοιώσεις θα μπορούν να αναδιανέμονται στα υπόλοιπα ρομπότ τα οποία δεν θα χρειάζονται τόσο εκτενή εκπαίδευση.

Βιομηχανία

Σε όλες τις βιομηχανίες όλα τα προϊόντα περνάνε πρώτα από ένα σχεδιαστικό και διαγνωστικό στάδιο πριν παραχθούν. Τα DCGANs εμπλουτισμένα με σχεδιαστικά δεδομένα μπορούν να αυτοματοποιήσουν αυτό το στάδιο σε συνδυασμό με CAD προγράμματα και να παράγουν νέα και ίσως πρωτοποριακά σχέδια. Ταυτόχρονα, έχουν γίνει επιτυχημένες έρευνες στο διαγνωστικό κομμάτι, όπου τα γεννητικά μοντέλα μπορούν να διαγνώσουν τυχόν λάθη ή βλάβες στα μηχανικά κομμάτια του σχεδίου, ώστε να προβλεφθούν πριν την τελική παραγωγή τους.

Ενέργεια

Η ενέργεια είναι βασικό κριτήριο στις κατασκευές. Οι ανησυχίες σχετικά με το κόστος και τους περιβαλλοντικούς ρύπους, είναι ένα κριτήριο που κάθε κατασκευαστής αναλογίζεται, καθώς είναι σημαντικό να ελαχιστοποιείται όσο το δυνατόν περισσότερο. Η ακριβής πρόβλεψη της χρήσης ενέργειας του κτιρίου γίνεται ολοένα και πιο ζωτικής σημασίας. Κάποιες πειραματικές έρευνες έχουν δείξει πως η χρήση γεννητικών μοντέλων έχουν παράγει ελπιδοφόρα αποτελέσματα για την σωστή κατασκευή κτιρίων. Όσο περνάει ο χρόνος, η ενέργεια αποκτά όλο και ζωτικότερη σημασία, και συνδέεται ρητά με κάθε κτιριακή μονάδα. Οπότε η ελαχιστοποίηση της χρήσης της ενέργειας είναι ένα σημαντικό μέλημα για κάθε ιδιοκτήτη.

GAN στη μουσική

Για έναν αλγόριθμο παραγωγής, οι εικόνες μπορεί να φαίνονται εύκολο να δημιουργηθούν. Ο ήχος, ωστόσο, είναι ένα διαφορετικό είδος πρόκλησης, επειδή κάθε δείγμα εξαρτάται σε μεγάλο βαθμό από τα προηγούμενα. Είναι επίσης σημαντικό για το μοντέλο να μπορεί να δημιουργήσει μια δομή μελωδίας και έναν χαρακτηριστικό τρόπο που εξαρτάται από τη σχέση μεταξύ διαφορετικών τόνων και συγχορδιών. Συνολικά, τα αποτελέσματα είναι ελπιδοφόρα και αισθητικά ελκυστικά. Ωστόσο, η δομή φαίνεται επαναλαμβανόμενη με τρόπο που υποδηλώνει ότι η διαδικασία παραγωγής στερείται καινοτομίας. Η εκπαίδευση ενός παραγωγικού αλγορίθμου για τη δημιουργία μουσικής είναι πράγματι μια δύσκολη εργασία, ειδικά όταν έχουμε διαφορετικά όργανα με ανεξάρτητες ιδιότητες. Ωστόσο, δεν πρέπει να εγκαταλείψουμε τις τεράστιες δυνατότητες των GAN. Οι μηχανικοί τεχνητής

νοημοσύνης και οι επιστήμονες δεδομένων εργάζονται συνεχώς για τη βελτίωση αυτών των υπαρχόντων μοντέλων, παράλληλα με τη δημιουργία νέων.

Τα GAN στη λογοτεχνία

Όπως η μουσική, η δημιουργία κειμένου απαιτεί τη συνεκτίμηση της ακολουθίας λέξεων πριν από κάθε νέα προσθήκη. Ωστόσο, η εργασία εδώ είναι απλούστερη, καθώς η εισαγωγή (λέξεις) διακρίνεται εύκολα, ειδικά με τη βοήθεια προηγμένων τεχνικών NLP[11] και γλωσσικών μοντέλων. Η δημιουργία κειμένου τον τελευταίο καιρό συνδέθηκε με τον φόβο της απειλής της τεχνητής νοημοσύνης. Το ίδιο με σχεδόν κάθε εφαρμογή τεχνητής νοημοσύνης, οι άνθρωποι πάντα φοβούνται τη διάδοση της τεχνητής νοημοσύνης και την πιθανότητα καταστροφικών συνεπειών. Όσον αφορά τη δημιουργία κειμένου, τα άρθρα που δημιουργούνται θα μπορούσαν να διαδοθούν ως ψευδείς ειδήσεις χωρίς να έχουν την παραμικρή αμφιβολία για την αυθεντικότητά τους.

11 NLP (Natural Language Processing): Αναφέρεται στον κλάδο της επιστήμης των υπολογιστών που ασχολείται με το να δώσει στους υπολογιστές την ικανότητα να κατανοούν το κείμενο και τις προφορικές λέξεις με τον ίδιο τρόπο που μπορούν οι άνθρωποι.

Συμπίεση δεδομένων

Το Διαδίκτυο μας επιτρέπει να μεταφέρουμε τεράστιο όγκο δεδομένων σε οποιαδήποτε τοποθεσία, αλλά αυτό έχει ένα τίμημα, την συμπίεση δεδομένων. Τα GAN μας δίνουν τη δυνατότητα να αυξήσουμε την ανάλυση δεδομένων. Μπορούμε να μεταφέρουμε εικόνες και βίντεο χαμηλής ανάλυσης στην επιθυμητή θέση τους και, στη συνέχεια, τα GAN μπορούν να χρησιμοποιηθούν για τη βελτίωση της ποιότητας των δεδομένων, κάτι που απαιτεί λιγότερο εύρος ζώνης. Αυτό ανοίγει μια ολόκληρη σειρά από δυνατότητες.

E-Commerce

Το NLP και η τεχνητή νοημοσύνη στο ηλεκτρονικό εμπόριο έχουν προχωρήσει πολύ. Όταν κοιτάξουμε απλοϊκά το εμπόριο, παρατηρούμε ένα μοτίβο. Όλα ξεκινάνε και τελειώνουν με την εμπειρία του πελάτη. Εάν ο πελάτης αισθάνεται ότι τον εκτιμούν και τον κατανοούν, πιθανότατα θα επιστρέψει ξανά στις υπηρεσίες. Για αυτόν τον λόγο, είναι σημαντικό να έχουμε μια αυτοματοποιημένη εξατομικευμένη εμπειρία πελάτη. Για να κατανοήσουμε καλύτερα την εξατομίκευση, μπορούμε να την φανταστούμε το μάρκετινγκ ως ένας προς έναν. Με άλλα λόγια, εστιάζει στο να δώσει στους πελάτες αυτό που θέλουν και αυτό που αναζητούν. Τα NLP βελτιώνουν την αλληλεπίδραση μεταξύ ανθρώπων και μηχανών, προσφέροντας στους πελάτες οικειότητα μέσα στην ιστοσελίδα και άμεση εξυπηρέτηση.

Βελτίωση της κυβερνοασφάλειας (cyber security)

Τα περιστατικά απειλών στον κυβερνοχώρο έχουν αυξηθεί τα τελευταία χρόνια. Όλο και περισσότερα δεδομένα κοινοποιούνται πρόθυμα από τους ανθρώπους, με τη μορφή εικόνων και βίντεο, στο Διαδίκτυο και, ως εκ τούτου, γίνονται μια εύκολη πηγή για λανθασμένη χρήση. Το Adversarial Attack[10] είναι μια τέτοια μέθοδος που χρησιμοποιείται από χάκερ. Οι χάκερ χειρίζονται εικόνες προσθέτοντας κακόβουλα δεδομένα σε αυτές. Αυτό ξεγελάει το ίδιο το νευρωνικό δίκτυο και θέτει σε κίνδυνο την επιδιωκόμενη λειτουργία του αλγορίθμου. Αυτό, με τη σειρά του, μπορεί να έχει ως αποτέλεσμα την αποκάλυψη και παραβίαση

ανεπιθύμητων πληροφοριών. Τα GAN μπορούν να εκπαιδευτούν για τον εντοπισμό τέτοιων περιπτώσεων απάτης. Μπορούν να χρησιμοποιηθούν για να κάνουν τα μοντέλα βαθιάς μάθησης πιο ισχυρά. Το νευρωνικό δίκτυο μπορεί να εκπαιδευτεί ώστε να αναγνωρίζει τυχόν κακόβουλες πληροφορίες που ενδέχεται να προστεθούν σε εικόνες και άλλα δεδομένα από χάκερ. Ερευνητές και αναλυτές δημιουργούν επίτηδες ψεύτικα παραδείγματα και τα χρησιμοποιούν για να εκπαιδεύσουν το νευρωνικό δίκτυο. Το δίκτυο βελτιώνεται από μόνο του καθώς αναλύει πολλές εικόνες.

10 Adversarial Attack: Μια τεχνική που επιχειρεί να ξεγελάσει μοντέλα με παραπλανητικά δεδομένα, είναι μια αυξανόμενη απειλή στην ερευνητική κοινότητα της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Ο πιο συνηθισμένος λόγος είναι η πρόκληση δυσλειτουργίας σε ένα μοντέλο μηχανικής μάθησης. Μια τέτοια επίθεση μπορεί να συνεπάγεται την παρουσίαση ενός μοντέλου με ανακριβή ή παραπλανητικά δεδομένα καθώς εκπαιδεύεται ή την εισαγωγή δεδομένων που έχουν σχεδιαστεί με κακόβουλο τρόπο για να εξαπατήσει ένα ήδη εκπαιδευμένο μοντέλο.

Δημιουργία μοντέλων κινουμένων σχεδίων

Η βιομηχανία των βιντεοπαιχνιδιών μπορεί να επωφεληθεί πάρα πολύ από τα GAN. Τα GAN μπορούν να χρησιμοποιηθούν για την αυτόματη δημιουργία τρισδιάστατων μοντέλων που απαιτούνται σε βιντεοπαιχνίδια, ταινίες κινουμένων σχεδίων ή κινούμενα σχέδια. Το δίκτυο μπορεί να δημιουργήσει νέα τρισδιάστατα μοντέλα με βάση το υπάρχον σύνολο δεδομένων 2D εικόνων που παρέχονται. Το νευρωνικό δίκτυο μπορεί να αναλύσει τις 2D φωτογραφίες για να αναδημιουργήσει τα τρισδιάστατα μοντέλα τους σε σύντομο χρονικό διάστημα. Αυτό θα βοηθήσει σημαντικά τους σχεδιαστές να εξοικονομήσουν χρόνο και να χρησιμοποιήσουν το χρόνο τους αλλού για άλλες σημαντικές εργασίες.

Επεξεργασία φωτογραφιών

Όταν κάποιος σκέφτεται να χρησιμοποιήσει τα GAN για την επεξεργασία φωτογραφιών, πρέπει να σκεφτεί πέρα από τις συνηθισμένες βελτιώσεις. Τα GAN μπορούν να χρησιμοποιηθούν για την ανακατασκευή εικόνων προσώπων για τον εντοπισμό αλλαγών σε χαρακτηριστικά όπως το χρώμα των μαλλιών, τις εκφράσεις του προσώπου ή το φύλο κ.λπ. Αυτό μπορεί να βοηθήσει τις αρχές να εντοπίσουν εγκληματίες που μπορεί να έχουν υποβληθεί σε χειρουργικές επεμβάσεις για να τροποποιήσουν την εμφάνισή τους. Ομοίως, η γήρανση του προσώπου, με τη βοήθεια Γεννητικών Ανταγωνιστικών Δικτύων, μπορούν να χρησιμοποιηθούν για τη δημιουργία εικόνων προσώπων ανθρώπων σε διάφορες ηλικίες. Αυτό, πάλι, μπορεί να βοηθήσει στον εντοπισμό ατόμων που έχουν εξαφανιστεί ή έχουν διαφύγει εδώ και χρόνια. Επιπλέον, τα GAN μπορούν να χρησιμοποιηθούν για τη βελτίωση των εικόνων ώστε να γίνουν πιο ελκυστικές και ενημερωτικές. Ορισμένες λεπτομέρειες μπορούν να αφαιρεθούν από την εικόνα για να γίνει πιο λεπτομερής. Για παράδειγμα, υπάρχει ένα GAN το οποίο μπορεί να αφαιρέσει τη βροχή και το χιόνι από φωτογραφίες. Άλλη σημαντική εφαρμογή είναι η μετατροπή παλιών ασπρόμαυρων φωτογραφιών σε έγχρωμες.

8.3 Περιορισμοί και Κίνδυνοι

Τα GAN έχουν λύσει πολλά προβλήματα για τα μοντέλα παραγωγής και έχουν εμπνεύσει άλλες μεθόδους τεχνητής νοημοσύνης, αλλά εξακολουθούν να έχουν περιορισμούς. Η εκπαιδευτική διαδικασία πρέπει να διασφαλίζει την ισορροπία και τον συγχρονισμό δύο αντίπαλων δικτύων, διαφορετικά είναι δύσκολο να επιτευχθούν καλά αποτελέσματα εκπαίδευσης. Ωστόσο, είναι δύσκολο να ελεγχθεί ο συγχρονισμός των δύο αντίπαλων δικτύων, επομένως η διαδικασία εκπαίδευσης μπορεί να είναι ασταθής. Επιπλέον, αν και τα δείγματα που δημιουργούνται από τα GAN είναι διαφορετικά, υπάρχει το πρόβλημα του mode collapse. Αν και τα GAN έχουν ορισμένους περιορισμούς, είναι αδιαμφισβήτητο ότι η ερευνητική πρόοδος των GAN έχει αποκαλύψει τις ευρείες προοπτικές τους. Νέες τεχνικές αφιερωμένες στη μείωση των περιορισμών εμφανίζονται συνεχώς. Για παράδειγμα, το Wasserstein GAN ξεπερνά σε μεγάλο βαθμό το πρόβλημα της αστάθειας της εκπαίδευσης και ταυτόχρονα λύνει εν μέρει το πρόβλημα του mode collapse. Το πώς να αποφύγουμε εντελώς το mode collapse και να βελτιστοποιήσουμε περαιτέρω τη διαδικασία εκπαίδευσης παραμένει μια ερευνητική κατεύθυνση των GAN. Επιπλέον, η θεωρία για τη σύγκλιση του μοντέλου και την ύπαρξη σημείου ισορροπίας παραμένουν σημαντικά ερευνητικά θέματα στο εγγύς μέλλον. Ένας από τους μεγαλύτερους κινδύνους των GAN είναι η δυνατότητα δημιουργίας παραπλανητικών δεδομένων. Αυτό μπορεί να έχει σοβαρές συνέπειες σε τομείς όπως η πολιτική, η οικονομία και η ασφάλεια, όπου η παραπληροφόρηση μπορεί να έχει σημαντικό αντίκτυπο. Για παράδειγμα, τα GAN έχουν χρησιμοποιηθεί για τη δημιουργία πλαστών βίντεο που χειραγωγούν την εμφάνιση και τις ενέργειες των ανθρώπων, τα οποία μπορούν να χρησιμοποιηθούν για κακόβουλους σκοπούς, όπως προπαγάνδα ή εκβιασμό. Ένας ακόμη κίνδυνος των GAN είναι ότι μπορούν να διαιωνίσουν και να ενισχύσουν τις υπάρχουσες προκαταλήψεις. Αυτό συμβαίνει επειδή τα GAN συνήθως εκπαιδεύονται σε υπάρχοντα δεδομένα, τα οποία μπορεί να είναι μεροληπτικά. Εάν αυτά τα μεροληπτικά δεδομένα χρησιμοποιηθούν για την εκπαίδευση του Generator, θα δημιουργήσουν προκατειλημμένα ή μεροληπτικά δεδομένα, διαιωνίζοντας το πρόβλημα. Οι πιθανές συνέπειες των GAN δεν περιορίζονται σε εκείνες που είναι σκόπιμα κακόβουλες ή επιβλαβείς. Υπάρχει επίσης η πιθανότητα ακούσιων συνεπειών. Για παράδειγμα, τα GAN θα μπορούσαν να χρησιμοποιηθούν για τη δημιουργία συνθετικών ιατρικών εικόνων που χρησιμοποιούνται για την εκπαίδευση διαγνωστικών μοντέλων, αλλά εάν αυτές οι συνθετικές εικόνες περιέχουν σφάλματα, αυτό θα μπορούσε να οδηγήσει σε εσφαλμένες διαγνώσεις και να βλάψει τους ασθενείς. Λόγω των πιθανών κινδύνων των GAN, υπάρχει αυξανόμενη ανάγκη για ρύθμιση και έλεγχο της χρήσης τους. Αυτό περιλαμβάνει μέτρα όπως η ανάπτυξη προτύπων για την ποιότητα των παραγόμενων δεδομένων, η διασφάλιση ότι τα δεδομένα που δημιουργούνται δεν χρησιμοποιούνται για να βλάψουν άτομα ή κοινότητες και η διασφάλιση ότι τα GAN χρησιμοποιούνται με ηθικό και υπεύθυνο τρόπο. Μακροπρόθεσμα, ο τρόπος χρήσης των GAN για την προώθηση της ανάπτυξης και της εφαρμογής της τεχνητής νοημοσύνης, έχει και μπορεί να φέρει επανάσταση σε πολλούς τομείς, αλλά συνοδεύεται από σημαντικούς κινδύνους. Είναι σημαντικό να εξεταστούν προσεκτικά οι πιθανές συνέπειες της χρήσης GAN και να αναπτυχθούν μέτρα για τη ρύθμιση και τον έλεγχο της χρήσης τους, προκειμένου να ελαχιστοποιηθεί η πιθανότητα βλάβης. Με αυτόν τον τρόπο, μπορούμε να διασφαλίσουμε ότι τα GAN χρησιμοποιούνται προς όφελος της κοινωνίας και όχι εις βάρος της.

8.4 Αντίκτυπος και Προοπτική

Τα GAN έχουν σημαντικό αντίκτυπο στο πεδίο της μηχανικής μάθησης. Μία από τις βασικές συνεισφορές των GAN είναι η ικανότητα δημιουργίας νέων δειγμάτων δεδομένων υψηλής ποιότητας. Τα GAN έχουν χρησιμοποιηθεί για τη δημιουργία ρεαλιστικών εικόνων, βίντεο, κειμένου, ομιλίας, μουσικής και άλλων τύπων δεδομένων που μπορούν να χρησιμοποιηθούν για την αύξηση των υπαρχόντων συνόλων δεδομένων ή τη δημιουργία εντελώς νέων συνόλων δεδομένων. Πέρα από αυτά τα πεδία, τα GAN έχουν εφαρμοστεί σε μια ποικιλία άλλων εργασιών, όπως στην δημιουργία νέων φαρμάκων, στην κυβερνοασφάλεια και στην εξυπηρέτηση πελατών. Αυτό τονίζει την ευελιξία και τις δυνατότητες των GAN να εφαρμοστούν σε πολλά διαφορετικά πεδία και εφαρμογές. Παρά τον αντίκτυπό τους, τα GAN εξακολουθούν να αντιμετωπίζουν ορισμένους περιορισμούς που πρέπει να αντιμετωπιστούν. Ξεπερνώντας αυτούς τους περιορισμούς, το μέλλον των GANs φαίνεται λαμπρό και υπάρχει ένας αυξανόμενος όγκος έρευνας που στοχεύει στην προώθηση της τελευταίας τεχνολογίας σε αυτόν τον τομέα. Αυτό περιλαμβάνει έρευνα που στοχεύει στη βελτίωση της σταθερότητας και της ποικιλομορφίας της εκπαίδευσης, ενσωματώνοντας πρόσθετες πληροφορίες όπως σημασιολογικές πληροφορίες για πιο ρεαλιστικά αποτελέσματα και εφαρμογή GAN σε νέες εφαρμογές και τομείς. Συμπερασματικά, τα GAN έχουν σημαντικό αντίκτυπο στο πεδίο της μηχανικής μάθησης και έχουν τη δυνατότητα να εφαρμοστούν σε ένα ευρύ φάσμα εφαρμογών.

9. Επίλογος

Τα Generative Adversarial Networks (GAN) είναι ένας τύπος μοντέλου βαθιάς μάθησης που έχει σχεδιαστεί για τη δημιουργία δεδομένων που είναι παρόμοια με ένα σύνολο δεδομένων εκπαίδευσης. Τα GAN αποτελούνται από δύο μέρη: ένα δίκτυο Δημιουργού και ένα δίκτυο Διευκρινιστή. Το δίκτυο Generator παράγει πλαστά δεδομένα, ενώ το δίκτυο Discriminator προσπαθεί να διακρίνει τα πλαστά δεδομένα από τα πραγματικά δεδομένα. Τα δύο δίκτυα εκπαιδεύονται μαζί, με τον Generator να μαθαίνει να παράγει δεδομένα που είναι όλο και πιο δύσκολο για τον Discriminator να αναγνωρίσει ως πλαστά. Τα GAN έχουν χρησιμοποιηθεί για ένα ευρύ φάσμα εφαρμογών, συμπεριλαμβανομένης της δημιουργίας εικόνων, κειμένου και μουσικής. Συνολικά, τα GAN είναι ένα ισχυρό εργαλείο για τη δημιουργία συνθετικών δεδομένων, αλλά μπορεί να είναι δύσκολο να εκπαιδεύονται και μπορεί να απαιτούν σημαντικούς υπολογιστικούς πόρους. Υπάρχουν μερικές πιθανές ηθικές ανησυχίες σχετικά με τη χρήση των GAN. Μια ανησυχία είναι ότι τα GAN χρησιμοποιούνται για τη δημιουργία ψεύτικων εικόνων ή βίντεο, τα οποία ύστερα να χρησιμοποιηθούν για τη διάδοση παραπληροφόρησης ή την εξαπάτηση ανθρώπων. Μια άλλη ανησυχία είναι ότι τα GAN θα μπορούσαν να χρησιμοποιηθούν για τη δημιουργία συνθετικών δεδομένων που είναι μεροληπτικά ή άδικα. Μια ακόμη ανησυχία είναι ότι τα GAN θα μπορούσαν να χρησιμοποιηθούν για τη δημιουργία περιεχομένου που είναι προσβλητικό ή ενοχλητικό, όπως βίαιες ή πορνογραφικές σκηνές. Συνολικά, είναι σημαντικό να ληφθούν υπόψη οι πιθανές ηθικές επιπτώσεις και το περιβάλλον της χρήσης των GAN. Είναι περίπλοκα και απαιτούν προσεκτική σκέψη και εξέταση. Είναι σημαντικό να σταθμίσουμε προσεκτικά τα πιθανά οφέλη και μειονεκτήματα από τη χρήση των GAN και να τεθούν σε εφαρμογή δικλείδες ασφαλείας για τον μετριασμό τυχόν αρνητικών επιπτώσεων. Είναι δύσκολο να πούμε με βεβαιότητα τι επιφυλάσσει το μέλλον για τα Generative Adversarial Networks (GAN), αλλά πολλοί ειδικοί είναι αισιόδοξοι για τις δυνατότητές τους. Καθώς τα GAN συνεχίζουν να προχωρούν, έχουν τη δυνατότητα να φέρουν επανάσταση σε πολλούς διαφορετικούς κλάδους δημιουργώντας νέα, υψηλής ποιότητας δεδομένα που δεν διακρίνονται από τα δεδομένα του πραγματικού κόσμου. Αυτό θα μπορούσε να έχει ένα ευρύ φάσμα εφαρμογών, από τη δημιουργία εξαιρετικά ρεαλιστικών εικονικών κόσμων έως τη δημιουργία νέου περιεχομένου για βιντεοπαιχνίδια και άλλες μορφές ψυχαγωγίας. Τελικά, το μέλλον για τα GAN φαίνεται πολλά υποσχόμενο και πολλοί άνθρωποι είναι ενθουσιασμένοι να δουν ποιες νέες εξελίξεις θα προκύψουν στον τομέα.

Βιβλιογραφία

[1] Ahirwar, K. (2019). Generative Adversarial Networks Projects'. Birmingham: Packt Publishing

[2] Arjovsky, M., Chintala, S., Bottou, L. (2017). Wasserstein GAN

[3] Berasategi, A. (2019). Earth mover's distance. Ανακτήθηκε από: https://towardsdatascience.com/earth-movers-distance-68fff0363ef2

[4] Barnett, S. (2018). Convergence Problems with Generative Adversarial Networks (GANs). Mathematical Institute University of Oxford

[5] Berthelot, D., Schumm, T., Metz, L. (2017). BEGAN: Boundary Equilibrium Generative Adversarial Networks. Google

[6] Bowles, C., Chen, L., Ricardo, G., Bentley, P., Gunn, R., Hammers, A., Dickie, D., Hernandez, M., Wardlaw, J., Rueckert, D. (2018). GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks. Imperial College London

[7] Brownlee, J. (2019). Generative Adversarial Networks with Python. Συγγραφέας
Chavdarova, T., Fleuret, F. (2017). SGAN: An Alternative Training of Generative Adversarial Networks. Idiap Research Institute

[8] Ciresan, D., Meier, U., Masci, J., Gambardella, L., Schmidhuber, J. (2011). Flexible, High Performance Convolutional Neural Networks for Image Classification, IDSIA, USI and SUPSI, Switzerland

[9] Chollet, F. (2018). Deep Learning with Python. New York: Manning Publications

[10] Chong, P., Ruff, L., Kloft, M., Binder, A. (2021). Simple and Effective Prevention of Mode Collapse in Deep One-Class Classification

[11] Deep Convolutional Generative Adversarial Network (χ.χ). Στο tensorflow online. Ανακτήθηκε από: https://www.tensorflow.org/tutorials/generative/dcgan#next_steps

[12] Deng, J., Dong, W., Socher, R., Li L., Li, K., Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image dataset

[13] Donahue, J., Krahenbuhl, P., Darell, T. (2017). Adversarial Feature Learning. ICLR

[14] Dumoulin, V., Visin, F., (2018). A guide to convolution arithmetic for deep learning
Srivastava, N., Hinton, G., Krizhevsky, A., Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Department of Computer Science University of Toronto

[15] Dwivedi, H. (2022). Understanding GAN Loss Functions. Ανακτήθηκε από: https://neptune.ai/blog/gan-loss-functions

[16] Fedus, W., Rosca, M., Lakshminarayanan, B., Dai, A., Mohamed, S., Goodfellow, I. (2018). Many Paths to Equilibrium: GANs do not Need to Decrease a Divergence. Google Brain, DeepMind

[17] Foster, D. (2019). Generative Deep Learning, Teaching Machines to Paint, Write, Compose and Play. California: O'Reilly Media, Inc.

[18] Goodfellow, I. (2016). NIPS 2016 Tutorial Generative Adversarial Networks

[19] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning, Chapter 2 Linear Algebra (σ. 29-50). MIT Press. Ανακτήθηκε από: https://www.deeplearningbook.org/contents/linear_algebra.html

[20] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2014). Generative Adversarial Networks. University of Montreal

[21] Hachcham, A. (2022). Training and Debugging Deep Convolutional Generative Adversarial Networks. Ανακτήθηκε από: https://neptune.ai/blog/deep-convolutional-generative-adversarial-networks

[22] Hui, J. (2018). GAN - A comprehensive review into the gangsters of GANS (Part 1). Ανακτήθηκε από: https://jonathan-hui.medium.com/gan-a-comprehensive-review-into-the-gangsters-of-gans-part-1-95ff52455672

[23] Hui, J. (2018). GAN - Why it is so hard to train Generative Adversarial Networks! Ανακτήθηκε από: https://jonathan-hui.medium.com/gan-why-it-is-so-hard-to-train-generative-advisory-networks-819a86b3750b#4987

[24] IBM Cloud Education. (2020). Natural Language Processing (NLP). Ανακτήθηκε από: https://www.ibm.com/cloud/learn/natural-language-processing

[25] Ioffe, S., Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training b y Reducing Internal Covariate Shift

[26] Jolicoeur-Martineau, A., Mitliagkas, I. (2020). Gradient Penalty from a Maximum Margin Perspective. Department of Computer Science University of Montreal

[27] Joshi, N. (2020). 5 applications of generative adversarial networks. Ανακτήθηκε από: https://www.allerin.com/blog/5-applications-of-generative-adversarial-networks

[28] Karras, T., Aila, T., Laine, S., Lehtinen, J. (2018). Progressive Growing of GANs for Improved Quality, Stability and Variation. NVIDIA

[29] Kingma, D., Ba, J. (2014). Adam: A Method for Stochastic Optimization

[30] Kingma, D., Welling, M. (2013). Auto-Encoding Variational Bayes

[31] Langr, J., Bok V. (2019). GANs in Action Deep Learning with Generative Adversarial Networks. New York: Manning Publications

[32] Machine's Creativity (2019). Artificial Art: How GANs are making machines creative. Ανακτήθηκε από: https://heartbeat.comet.ml/artificial-art-how-gans-are-making-machines-creative-b99105627198

[33] Mirza, M., Osindero, S. (2014). Conditional Generative Adversarial Nets

[34] Radford, A.., Metz, L., Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. ICLR

[35] Rectified Linear Units (ReLU) in Deep Learning. (χ.χ.). Ανακτήθηκε από: https://www.kaggle.com/code/dansbecker/rectified-linear-units-relu-in-deep-learning/notebook

[36] Rezende, D., Mohamed, S., Wierstra, D. (2014). Stochastic Backpropagation and Approximate Inference in Deep Generative Models. Google DeepMind, UK

[37] Rusu, A., Rabinowitz, N., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R., Hadsell, R.(2016). Progressive Neural Networks. Google DeepMind, UK

[38] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X. (2016). Improved Techniques for Training GANs

[39] Salimans, T., Kingma, D. (2016). Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks. OpenAI

[40] Shahriar, S. (2021). GAN Computers Generate Arts? A Survey on Visual Arts, Music, and Literary Text Generation using Generative Adversarial Network. Department of Computer Science and Engineering American University of Sharjah, UAE

[41] Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., Wang, F. (2017). Generative Adversarial Networks: Introduction and Outlook

[42] Wang, Z., She, Q., Ward, T. (2020). Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy

[43] Wiggers, K. (2021). Adversarial attacks in machine learning: What they are and how to stop them. Ανακτήθηκε από: https://venturebeat.com/security/adversarial-attacks-in-machine-learning-what-they-are-and-how-to-stop-them/

[44] Zeiler, M., Fergus, R. (2013). Visualizing and Understanding Convolutional Networkσ

[45] Zhang, H., Goodfellow, I., Metaxas, D., Odena, A. (2019). Self-Attention Generative Adversarial Networks

[46] Zhang, Wei., Sheng, Q., Alhazmi, A. (2019). Adversarial Attacks on Deep Learning Models in Natural Language Processing: A Survey

[47] Zhu, J., Park, T., Isola, P., Efros, A. (2020). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. BAIR laboratory, UC Berkeley

[48] Ανάλυση Πολυμεταβλητών Τεχνικών, Εφαρμογές Περιπτώσεων: Κεφάλαιο 3: Λογιστική Παλινδρόμηση. (χ.χ.). Ανακτήθηκε από: https://eclass.aegean.gr/modules/document/file.php/MATH110/%CE%A3%CE%A5%CE%9C%CE%A0%CE%9B%CE%97%CE%A1%CE%A9%CE%9C%CE%91%CE%A4%CE%99%CE%9A%CE%9F%20%CE%A5%CE%9B%CE%99%CE%9A%CE%9F/04_chapter03.pdf

[49] Αφελής ταξινόμηση Μπέυζ. (χ.χ.). Ανακτήθηκε από: https://course.elementsofai.com/el/3/3

[50] Κομηνέας, Σ., Χαρμανδάρης, Ε. (χ.χ). Μαθηματική Μοντελοποίηση: 7.1 Εισαγωγή: Στοχαστικά Συστήματα. Ανακτήθηκε από: http://repfiles.kallipos.gr/html_books/9863/Ch7.S1.html