

Νευρωνικά δίκτυα βαθιάς μάθησης για τη μη επιβλεπόμενη ανίχνευση ανωμαλιών σε δεδομένα

**Παρασκευόυπουλος Παναγιώτης**

Επιβλέπων καθηγητής: **Ιωάννης Μπούταλης**

**Περίληψη**

Περιεχόμενα

[1.Εισαγωγή 4](#_Toc71572545)

[Ανίχνευση ανωμαλιών 4](#_Toc71572546)

[Ιστορική αναδρομή 4](#_Toc71572547)

[Σύγχρονες εφαρμογές 5](#_Toc71572548)

[Νευρωνικά δίκτυα βαθιάς μάθησης 5](#_Toc71572549)

[Περιγραφή Δεδομένων 7](#_Toc71572550)

[Περιγραφή Μοντέλων Αυτοκωδικοποιητών 8](#_Toc71572551)

[Απλός Deep Autoencoder 8](#_Toc71572552)

[Sparse Autoencoder 9](#_Toc71572553)

[Denoising Autoencoder 9](#_Toc71572554)

[Contractive autoencoder 10](#_Toc71572555)

[Πειραματική Ανάλυση 11](#_Toc71572556)

[Προετοιμασία Δεδομένων 11](#_Toc71572557)

[Προεπεξεργασία Δεδομένων 11](#_Toc71572558)

[Διαδικασία Αρχικοποίησης και Εκπαίδευσης 11](#_Toc71572559)

[Συνάρτηση Απωλειών 12](#_Toc71572560)

[Διαδικασία Πρόβλεψης και Παραγωγής Αποτελεσμάτων 12](#_Toc71572561)

[Επιλογή Κατωφλίου (Threshold) 12](#_Toc71572562)

[Διαδικασία Ψηφοφορίας Συνόλου 13](#_Toc71572563)

[Μέθοδος Hard Voting 13](#_Toc71572564)

[Μέθοδος Soft Voting 14](#_Toc71572565)

[Τελικά Αποτελέσματα 15](#_Toc71572566)

# 1.Εισαγωγή

## Ανίχνευση ανωμαλιών

Η ανίχνευση ανωμαλιών είναι η αναγνώριση έκτροπων δεδομένων ή ακραίων τιμών (outliers) σε ένα σύνολο δεδομένων. Οι ανωμαλίες αυτές μπορεί να εμφανίζονται σπάνια μέσα στα δεδομένα ή τα χαρακτηριστικά τους να διαφέρουν αρκετά από το μέσο όρο των υπολοίπων δεδομένων.

Η ανίχνευση ανωμαλιών χωρίζεται σε τρείς κατηγορίες με βάση την τεχνική εκμάθησης των μοντέλων που χρησιμοποιούμε:

Την επιβλεπόμενη ανίχνευση ανωμαλιών κατα την οποία το σύνολο των δεδομένων μας περιέχει ετικέτες που υποδεικνύουν ποιά δεδομένα είναι κανονικά (inliers) και ποιά είναι έκτροπα (outliers) και η οποία αντιμετωπίζεται σαν πρόβλημα ταξινόμισης (classification).

Την ημι-επιβλεπόμενη ανίχνευση ανωμαλιών κατα την οποία θεωρούμε ότι ένα σύνολο δεδομένων είναι κανονικά και εκπαιδεύουμε το μοντέλο μας πάνω σε αυτά και έπειτα το μοντέλο μας εμφανίζει μεγάλη απόκλιση στα έκτροπα δεδομένα στη φάση της δοκιμής (test).

Την μη επιβλεπόμενη ανίχνευση ανωμαλιών κατα την οποία δεν έχουμε καθόλου ετικέτες στα δεδομένα και το μοντέλο μας προσπαθεί να τα κατηγοριοποιήσει έχοντας ως υπόθεση ότι τα έκτροπα δεδομένα είναι πολύ λιγότερα από τα κανονικά.

## Ιστορική αναδρομή

Η χρήση τεχνικών ανίχνευση ανωμαλιών είχε προταθεί ως ανίχνευση εισβολής (intrusion detection) από την Dorothy Denning το 1986

«»»»Να γραψω και άλλα ή να φύγει»»»»

<https://en.wikipedia.org/wiki/Anomaly_detection>

## Σύγχρονες εφαρμογές

Η ανίχνευση ανωμαλιών εφαμόζεται σε ένα ευρύ φάσμα πεδίων όπως η ανίχνευση εισβολής ή κακής χρήσης σε δίκτυα υπολογιστών, η ανίχνευση απάτης σε τραπεζικές συναλλαγές, η παρακολούθηση σωστής, λειτουργίας κάποιου συστήματος, η ανίχνευση συμβάντων σε δίκτυα αισθητήρων ή η ανίχνευση διαταρραχών σε ένα οικοσύστημα και η προεπεξεργασία δεδομένων με σκοπό την αφαίρεση έκτροπων δεδομένων πριν χρησιμοποιηθούν από κάποιο άλλο μοντέλο μηχανικής μάθησης ωστέ να αυξηθεί η τελική του απόδοση.

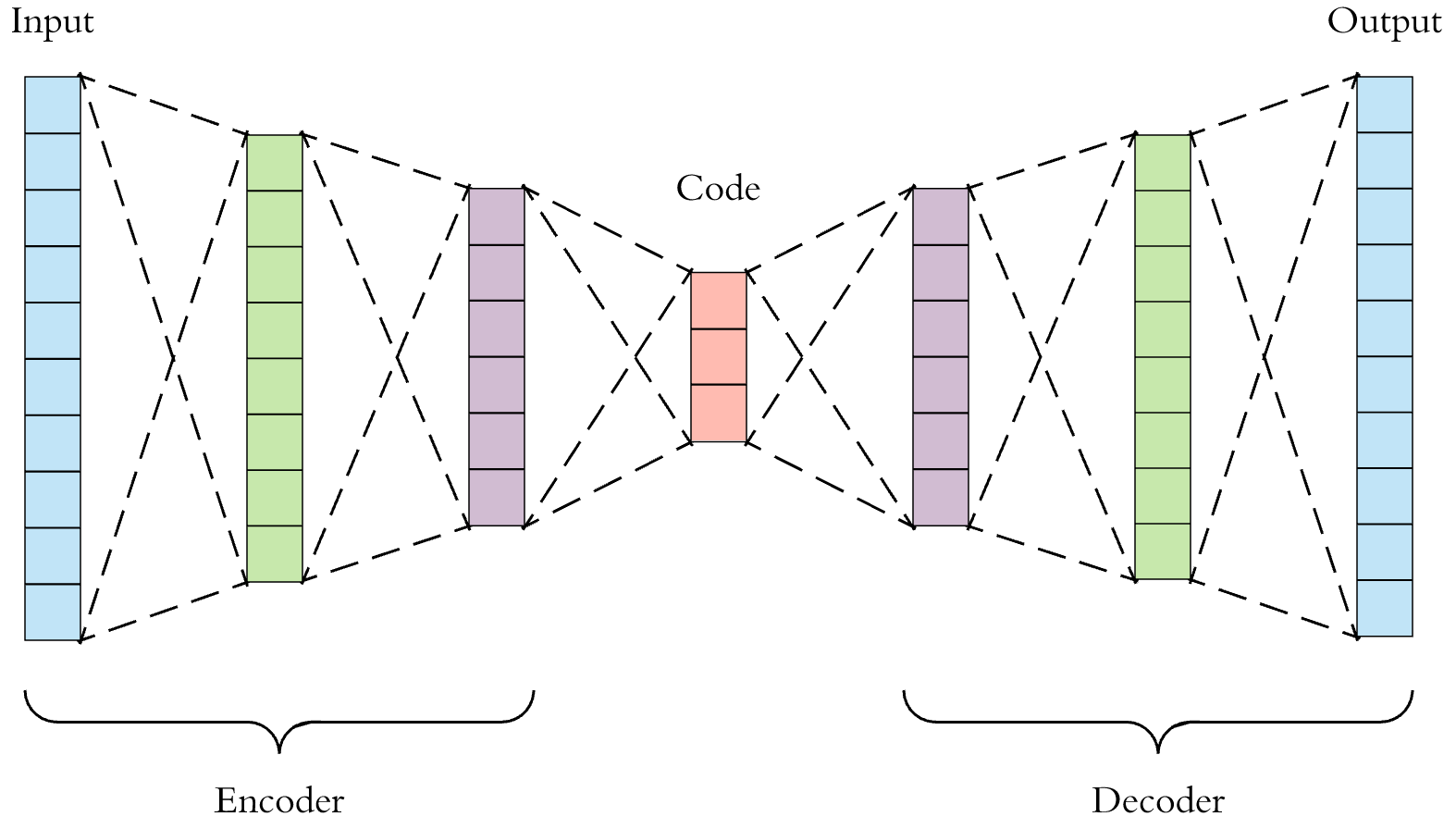
https://en.wikipedia.org/wiki/Anomaly\_detection

## Νευρωνικά δίκτυα βαθιάς μάθησης

<http://neuralnetworksanddeeplearning.com/chap6.html>

Τα νευρωνικά δίκτυα βαθιάς μάθησης αποτελούν κομμάτι της μηχανικής μάθησης και βασίζονται στα τεχνιτά νευρωνικά δίκτυα. Τα τεχνικά νευρωνικά δίκτυα αποτελούνται συνήθως από ένα στρώμα εισαγωγής, ένα κρυφό στρώμα και ένα στρώμα εξόδου. Τα νευρωνικά δίκτυα βαθιάς μάθησης επεκτείνουν τα κρυφά στρώματα σε περισσότερα του ενός, πετυχαίνοντας έτσι καλύτερη πρακτική απόδοση από τα τεχνιτά νευρωνικά δίκτυα ειδικότερα σε μεγάλα σύνολα δεδομένων καθώς και σε σύνθετα προβλήματα όπως η αναγνώριση εικόνας και ήχου. Στην συγκεκριμένη εργασία χρησιμοποιείται ένας συγκεκριμένος τύπους νευρωνικού δικτύου βαθιάς μάθησης που ονομάζεται Αυτοκωδικοποιητής (Autoencoder).

Ο Αυτοκωδικοποιητής είναι ένα είδος νευρωνικού δικτύου το οποίο εκπαιδεύεται ώστε να αντιγράφει τα δεδομένα που δέχεται στην είσοδο, στην έξοδό του.



<https://www.google.com/url?sa=i&url=https%3A%2F%2Ftowardsdatascience.com%2Fapplied-deep-learning-part-3-autoencoders-1c083af4d798&psig=AOvVaw0djFuovrbnPJlqw0ZDSefw&ust=1619796752020000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCND0j6Hjo_ACFQAAAAAdAAAAABAD>

Συγκεκριμένα αποτελείται απο δύο τμήματα: τον κωδικοποιητή (encoder) ο οποίος κωδικοποιεί την είσοδο σε ένα χώρο προβολής (latent space) και τον αποκωδικοποιητή ο οποίος προσπαθεί να ανακατασκευάσει την είδοδο απο τον χώρο προβολής.

Οπου Χ το τμήμα του encoder, η έξοδος του decoder και Χ ο χώρος προβολής. Η τρίτη εξίσωση την συνάρτηση κόστους με την οποία εκπαιδέυεται ο Αυτοκωδικοποιητής και εκφράζει την τετραγωνική διαφορά της εισόδου από την έξοδο μετά τις μεταβάζεις απο τον encoder και τον decoder το οποίο είναι το σφάλμα ανακατασκευής του Αυτοκωδικοποιητή.

Στην πρώτη εξίσωση το h είναι η έξοδος ενος κρυφού στρώματος του Αυτοκωδικοποιητή και τα W και b είναι το weight και το bias που εφαρμόζεται στους αντίστοιχους νευρώνες του δικτύου και το σ είναι η συνάρτηση ενεργοποίησης.

Η δεύτερη εξίσωση πειργράφει την ανακτασκευή στην έξοδο του Αυτοκωδικοποιητή.

Η τρίτη εξίσωση περιγράφει συνολικά το σφάλμα ανακατασκευής της εκπαίδευσης του Αυτοκωδικοποιητή που περγράφηκε παραπάνω.

https://en.wikipedia.org/wiki/Autoencoder

Ο κωδικοποιητής και ο αποκωδικοποιητής αποτελούνται συνήθως από ίσο αριθμό κρυφών στρωμάτων τα οποία μειώνονται σε μέγεθος όσο πλησιάζουν το χώρο προβολής, ο οποίος έχει συνήθως μέγεθος μικρότερο της εισόδου. Στην περίπτωση αυτή ο Αυτοκωδικοποιητής ονομάζεται undercomplete ενώ στην αντίθετη περίπτωση που ο χώρος προβολής είναι μεγαλύτερος της εισόδου και της εξόδου ονομάζεται overcomplete.

Ορίζοντας το χώρο προβολής μικρότερο της εισόδου και της εξόδου περιορίζουμε την ικανότητα του δικτύου να αντιγράφει ακριβώς την είσοδο στην έξοδο κατά την εκπαίδευση και αρα να γίνει ειδικό στο να αναγνωρίζει μόνο τα δεδομένα εκπαίδευσης (το λεγόμενο overfitting, φαινόμενο που επιρεάζει αρνητικά τα αποτελέσματά μας) και επίσης πετυχαίνουμε την αποτύπωση των κύριων χαρακτηριστικών του συνόλου των δεδομένων κατά την ανακατασκευή τους.

# Περιγραφή Δεδομένων

Τα σύνολα δεδομένων που χρησιμοποιήθηκαν προέρχονται απο τη συλλογή Outlier Detection DataSets (ODDS) η οποία περιλαμβάνει δεδομένα για ανίχνευση ανωμαλιών τα οποία αποτελούνται απο δύο κλάσεις που χαρακτηρίζουν τα κανονικά και τα έκτροπα δεδομένα. Απο αυτά επιλέχθηκαν τα παρακάτω σύνολα:

* Cardiotocogrpahy dataset: πρόκειται για δεδομένα απο μετρήσεις καρδιοτογραφίας εμβρύων στο οποίο οι δύο κλάσεις αντισοιχούν σε κανονικές και παθολογικές μετρήσεις. Αποτελείται απο 1831 γραμμές με 21 χαρακτηριστικά και ποσοστό έκτροπων δεδομένων 9.6%
* Satellite dataset: πρόκειται για δεδομένα απο κωδικοποιημένα τμήματα εικόνας δορυφόρου τα οποία καλύπτουν μια μικρή περιοχή της φωτογραφίας σε τέσσερα διαφορετικά φάσματα χωρισμένα ανα περιοχές 3Χ3. Ετσι έχουμε 36 διαφορετικά χαρακτηριστικά και 6435 γραμμές με ποσοστό έκτροπων δεδομένων 32%
* Breast Cancer Wisconsin: πρόκειται για δεδομένα απο ιατρικές μετρήσεις όγκων οι οποίες χαρακτηρίζονται ως καλοήθεις (τα κανονικά δεοδμένα) και κακοήθεις (τα έκτροπα δεδομένα). Εχουμε 683 γραμμές και 9 χαρακτηριστικά με ποσοστό έκτροπων δεδομένων 35%.
* Wine dataset: πρόκειται για δεδομένα απο χημικές αναλύσεις κρασιών απο τρείς διαφορετικές ποικιλίες και περιέχει 129 γραμμές με 13 χαρακτηριστικά και ποσοστό έκτροπων δεδομένων 7.7%. Τα έκτροπα δεδομένα περιγράφουν την μία απο τις τρείς ποικιλίες και τα κανονικά περιγράφουν τις άλλες δύο ποικιλίες.

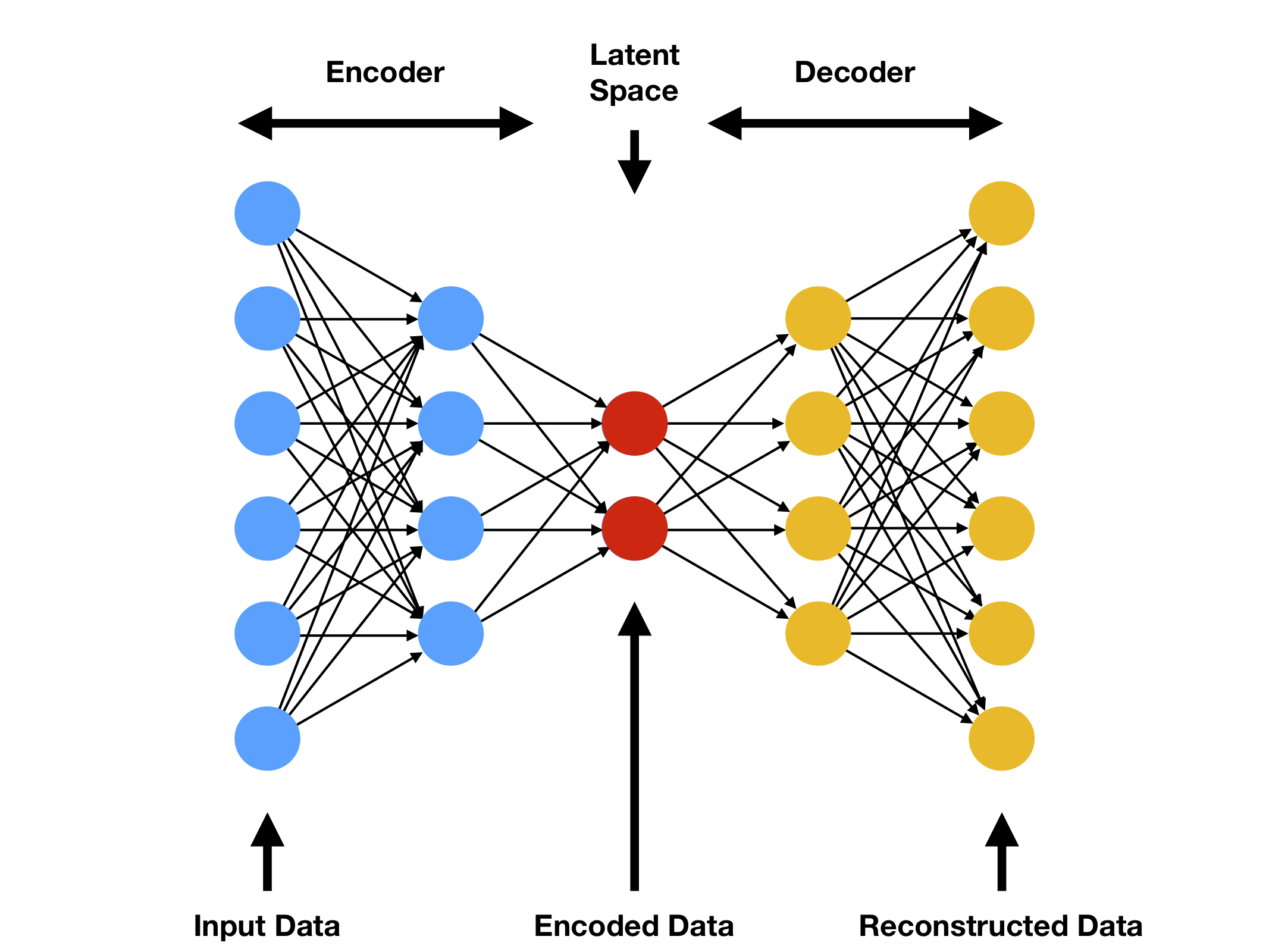
Για την χρήση των παραπάνω δεδομένων έγινε αρχικά κανονικοποίηση των τιμών τους απο -1 εως 1. Επειτα δημιουργήθηκαν τα δεδομένα εκπαίδευσης χρησιμοποιόντας τα μισά κανονικά δεδομένα του κάθε συνόλου και στη συνέχεια δημιουργήθηκαν τα δεδομένα τεστ που περιέχουν όλα τα έκτροπα δεδομένα και το υπόλοιπο μισό των κανονικών δεδομένων με τυχαία σειρά.

Ετσι δημιουργήθηκε το σύνολο εκπαίδευσης το οποίο δεν περιέχει ετικέτες για τα δεδομένα καθώς ειναι όλα κανονικά και το σύνολο τεστ που έχει ετικέτες για την ανίχνευση των ανωμαλιών.

# Περιγραφή Μοντέλων Αυτοκωδικοποιητών

## Απλός Deep Autoencoder

Ο απλός Deep Autoencoder είναι το βασικό μοντέλο που περιγράφηκε παραπάνω για τους αυτοκωδικοποιητές και αποτελείται απο πλήρως συνδεδεμένα στρώματα που έχουν ως συνάρτηση ενεργοποίησης την relu και στο στρώμα εξόδου έχουν ως συνάρτηση ενεργοποίησης την tanh.



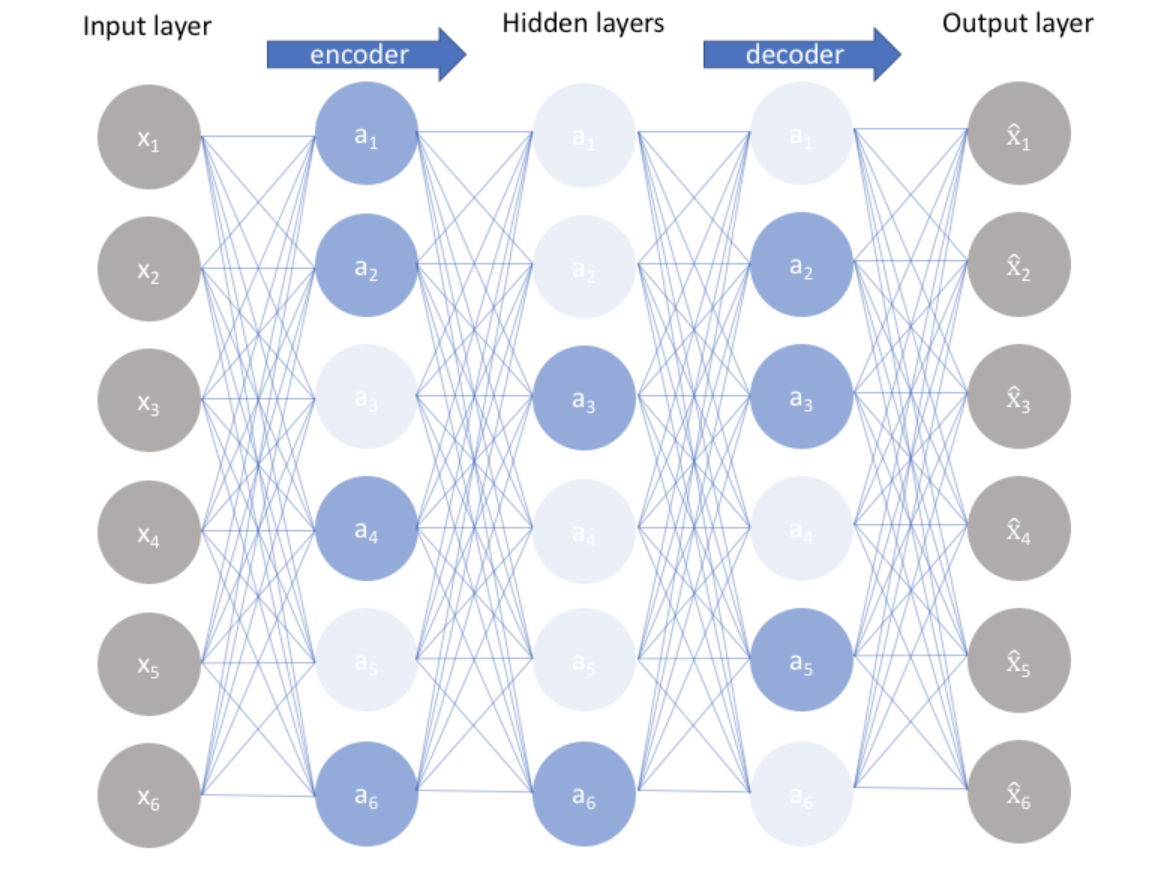
<https://www.compthree.com/images/blog/ae/ae.png>



<https://mathworld.wolfram.com/images/interactive/TanhReal.gif>

## Sparse Autoencoder

Ο Sparse Autoencoder είναι μια αρχιτεκτονική αυτοκωδικοποιητή στην οποία εισάγουμε επιπλέον περιορισμό στα στρώματα του αυτοκωδικοποιητή ωστε να μην ενεργοποιούνται όλοι οι κόμβοι του κάθε στρώματος και έτσι ο αυτοκωδικοποιητής να στηρίζεται σε λιγότερους κόμβους ανα στρώμα για να πετύχει το επιθυμητό αποτέλεσμα. Ετσι ο Sparse Autoencoder αποτελείται απο πλήρως συνδεδεμένα στρώματα με συνάρτηση ενεργοποίηση relu στα οποία εφαρμόζουμε L1 ομαλοποίηση εκτός απο τα στρώματα εισόδου και εξόδου. Η L1 ομαλοποίηση προσθέτει εναν επιπλέον συντελεστή για την ελαχιστοποίηση της συνάρτησης κόστους κατά την εκπαίδευση στον οποίο ορίζουμε την βαρύτητα ως 1 έτσι ώστε να αποκλείει τελείως κάπους κόμβους απο το δίκτυο.



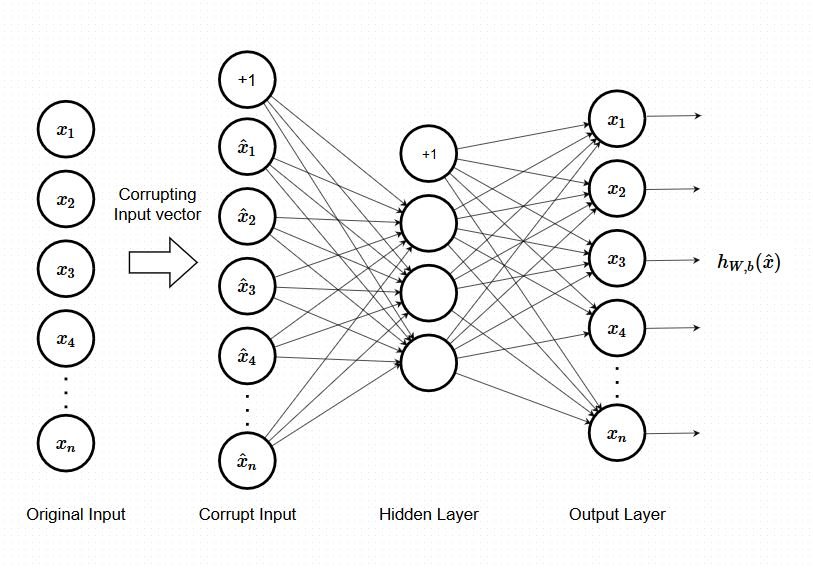
<https://paperswithcode.com/media/methods/Screen_Shot_2020-06-28_at_3.36.11_PM_wfLA8dB.png>

Η εξίσωση της συνάρτησης κόστους με την πρόσθεση της L1 ομαλοποίησης.

Επιπλέον στον Sparse Autoencoder έχει εφαρμοστεί σε ξεχωριστό μοντέλο και μια δεύτερη προσέγγιση χρησιμοποιόντας στρώματα Dropout τα οποία αποκλείουν ένα τυχαίο αριθμό κόμβων σε κάθε βήμα της εκπαίδευσης.

## Denoising Autoencoder

Ο Denoising Autoencoder είναι μια αρχιτεκτονική που έχει ίδια δομή με τον απλό Deep Autoencoder όσον αφορά τα στρώματα και την υλοποίηση. Η διαφορά εντοπίζεται στα δεδομένα εκπαίδευσης. Στα δεδομένα εκπαίδευσης προσθέτουμε ένα ποσοστό γκαουσιανού θορύβου αλλά στην έξοδο συγκρίνουμε την ανακατασκευή με τα αρχικά δεδομένα χωρίς θόρυβο. Το δίκτυο έτσι γίνεται λιγότερο ευαίσθητο στην απομνημόνευση των δεδομένων εκπαίδευσης και αποκτά μια πιο γενική εικόνα των χρήσιμων χαρακτηριστικών των δεδομένων στο χώρο προβολής. Οι τακτικές που εφαρμόζουμε στον Denoising Autoencoder έχουν αντίκτυπο κυρίως στο τμήμα του decoder αφού εκεί γίνεται η ανακατασκευή των δεδομένων.



https://www.researchgate.net/profile/Varun-Kumar-52/publication/272086159/figure/fig2/AS:573504648998912@1513745571839/Denoising-Autoencoder.png

## Contractive autoencoder

Ο Contractive Autoencoder είναι μια αρχιτεκτονική στην οποία προσπαθούμε όπως και στον Denoising Autoencoder να βελτιώσουμε την εκπαίδευση του autoencoder περιορίζοντας την απόκλιση της συνάρτησης ενεργοποίησης για παρόμοια δεδομένα εισόδου. Αυτό επιτυγχάνεται χρησιμοποιόντας L2 ομαλοποίηση σε όλα τα στρώματα εκτός της εισόδου και εξόδου. Η L2 ομαλοποίηση λειτουργεί εφαρμόζοντας επιπλέον συντελεστή στη συνάρτηση κόστους των κόμβων όμως δεν έχει την ιδιότητα να αποκλείει ή να μηδενίζει κόμβους όπως η L1, όμως πετυχαίνει την ελαχιστοποίηση μεγάλων αποκλίσεων κατα την εκπαίδευση. Ετσι εφαρμόζοντας L2 ομαλοποίηση σε όλα τα στρώματα εκτός της εισόδου και της εξόδου πετυχαίνουμε το επιθυμητό αποτέλεσμα. Σε αντίθεση με τον Denoising Autoencoder, ο Contracting Autoencoder επιδρά κυρίως στο τμήμα encoder του αυτοκωδικοποιητή.

<https://www.jeremyjordan.me/autoencoders/>

# Πειραματική Ανάλυση

## Προετοιμασία Δεδομένων

Για να γίνει χρήση των δεδομένων απο τους αυτοκωδικοποιητές απαιτείται η προετοιμασία και η προεπεξεργασία τους. Στην προαιτημασία κάνουμε εισαγωγή του συνόλου δεδομένων που θέλουμε να χρησιμοποιήσουμε και διαχωρίζουμε τις στήλες των χαρακτηριστικών και τη στήλη με τις ετικέτες και σχηματίζονται δύο πίνακες Χ και Υ. Στη συνέχεια διαχωρίζουμε τα έκτροπα δεδομένα και τα κανονικά δεδομένα σε δύο πίνακες Normal Data και Outliers. Στη συνέχεια χωρίζουμε με τυχαίο τρόπο στα δυο τον πίνακα των κανονικών δεδομένων, το πρώτο μισό είναι τα δεδομένα εκπαίδευσής μας και το δεύτερο μισό θα χρησιμοποιηθεί στα δεδομένα τεστ. Για να ολοκλρωθούν τα δεδομένα τεστ προσθέτουμε σε αυτά το δεύτερο μισό των κανονικών δεδομένων και έτσι έχουμε τα δεδομένα τεστ που αποτελούνται και απο κανονικά και απο έκτροπα δεδομένα μαζί με τις ετικέτες τους. Οι τελικοί πίνακες που θα χρησιμοποιηθούν και παρακάτω είναι οι X\_train, X\_test, Y\_test.

## Προεπεξεργασία Δεδομένων

Στην προεπεξεργασία των δεδομένων πρεπει αρχικά να δημιουργήσουμε τα δεδομένα με θόρυβο για τον Denoising Autoencoder. Αυτά δημιουργούνται με προσθήκη Γκαουσιανού θορύβου στο X\_train για τη δημιουργία του X\_train\_noisy. Η προσθήκη θορύβου γίνεται υπολογίζοντας τη μέση τιμή και τη διασπορά του κάθε χαρακτηριστικού του συνόλου δεδομένων (δηλαδή της κάθε στήλης του πίνακα) και προσθέτοντας τυχαίες τιμές της συγκεκριμένης κατανομής σε όλα τα στοιχεία της στήλης.

Τελικό βήμα της προεπεξεργασίας είναι η αναγωγή όλων των δεδομένων (εκπαίδευσης και τεστ) στο εύρος τιμών απο -1 εως 1. Αυτό το βήμα χρησιμοποιείται ευρέως στη βιβλιογραφία και είναι κοινή πρακτική της μηχανικής μάθησης και της χρήσης νευρωνικών δικτύων. Η χρησιμότητα αυτής της τεχνικής έχει δύο πλεονεκτήματα. Αρχικά βελτιώνει την διαδικασία της ελάττωσης της παραγώγου (gradient descent) η οποία χρησιμοποιήται κατά την εκπαίδευτη του νευρωνικού δικτύου κάνοντας τη σύγκλιση στο ελάχιστο πιο γρήγορη και αποφεύγοντας περιπτώσης που η ελλάτωση της παραγώγου δεν συγκλίνει. «»πηγή»» Επίσης στα σύνολα δεδομένων που χρησιμοποιούνται στην εργασία τα διάφορα χαρακτηριστικά αποτελούν μετρήσεις ή στοιχεία που δεν συσχετίζονται και αρα έχουν πολύ διαφορετικά εύρη τιμών, οπότε εαν δεν γινόταν αυτή η κανονικοποίηση, τα χαρακτηριστικά με τις μεγαλύτερες τιμές θα θεωρούνταν πιο σημαντικά στην εκπαίδευση χωρίς αυτό να είναι επιθυμητό. Ετσι με την κανονικοποίηση τους εξασφαλίζουμε την ίση μεταχείριση όλων των χαρακτηριστικών κατα την εκπαίδευση.

## Διαδικασία Αρχικοποίησης και Εκπαίδευσης

Μετά την προετοιμασία και προεπεξεργασία των δεδομένων αρχικοποιούμε τα πέντε διαφορετικά μοντέλα Αυτοκωδικοποιητών με τα χαρακτηριστικά που περιγράφηκαν στην προηγούμενη ενότητα.

Στην διαδικασία της εκπαίδευσης όπως αναφέρθηκε και παραπάνω χρησιμοποιούμε και σαν είσοδο και σαν απαιτούμενη έξοδο τα ίδια δεδομένα δηλαδή το X\_train που περιέχει μόνο κανονικά δεδομένα. Στην εκπαίδευση χρησιμοποιείται ενα τμήμα των δεδομένων εκπαίδευσης ως δεδομένα επιβεβαίωσης (validation data) τα οποία χρησιμοποιούνται για την παρακολούθηση της ακρίβειας του νευρωνικού δικτύου κατα την εκπαίδευση. Αυτό γίνεται με εισαγωγή των δεδομένων στην είσοδο του δικτύου και η μέτρηση της ακρίβειας στην έξοδο κατα τη διάρκεια της εκπαίδευσης ωστε να έχουμε εικόνα της προόδου της εκπαίδευσης. Το ποσοστό των δεδομένων εκπαίδευσης που χρησιμοποιούνται ως δεδομένα επιβεβαίωσης είναι 30%.

Στην εκπαίδευση χρησιμοποιείται η τεχνική που ονομάζεται Πρόορο Σταμάτημα (Early Stopping). Η τεχνική αυτή παρακολουθεί την ακρίβεια των δεδομένων επιβεβαίωσης στην πορεία της εκπαίδευσης και εαν διαπιστωθεί ότι η ακρίβεια δεν βελτιώνεται για πολλές συνεχόμενες επαναλήψεις της εκπαίδευση η εκπαίδευση σταματάει. Ετσι αποφεύγουμε την υποερβολική εκπαίδευση του νευρωνικού δικτύου στα δεδομένα εκπαίδευσης που ονομάζεται Υπερπροσαρμογή (Overfitting).

## Συνάρτηση Απωλειών

Η συνάρτηση απωλειών είναι ο τρόπος με τον οποίο μετράται το σφάλμα μεταξύ της εξόδου του νευρωνικού δικτύου και της επιθυμητής εξόδου. Στην συγκεκριμένη περίπτωση χρησιμοποιούμε το μέσο τετραγωνικό σφάλμα (Mean Squared Error) οπως περιγράφηκε στον ορισμό του Αυτοκωδικοποιητή παραπάνω

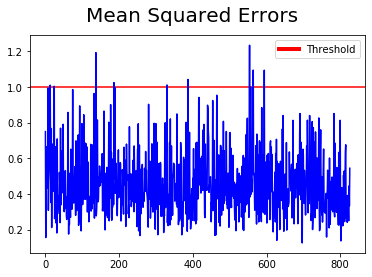
## Διαδικασία Πρόβλεψης και Παραγωγής Αποτελεσμάτων

Μετά το τέλος της διαδικασίας της εκπαίδευσης ακολουθεί η πρόβλεψη πάνω στα δεδομένα τεστ. Στο στάδιο αυτό δίνουμε ως είσοδο στο νευρωνικό δίκτυο τα δεδομένα τεστ και παίρνουμε στην έξοδο την ανακατασκευή τους. Επειτα υπολογίζουμε την τετραγωνική διαφορά τους. Την διαδικασία αυτή την εφαρμόζουμε και στα δεδομένα εκπαίδευσης, δηλαδή βάζουμε ως είσοδο τα δεδομένα εκπαίδευσης και στη συνέχεια υπολογίζουμε την τετραγωνική διαφορά της εξόδου απο την είσοδο. Ετσι σχηματίζονται δύο πίνακες error\_train και error\_test.

«»maths squared diff»»

## Επιλογή Κατωφλίου (Threshold)

Η χρήση Αυτοκωδικοποιητών απαιτεί τον ορισμό ενός κατωφλίου σφάλματος ανακατασκευής. Το κατώφλι αυτό θα ξεχωρίζει τα κανονικά απο τα έκτροπα δεδομένα. Ο υπολογισμός του κατωφλίου γίνεται με βάση τα δεδομένα εκπαίδευσης καθώς είναι τα μόνα στα οποία έχουμε πρακτικά πρόσβαση σε μια μη επιβλεπόμενη μέθοδο ανίχνευσης ανωμαλιών. Γνωρίζοντας ότι τα δεδομένα αυτά είναι κανονικά ξεχωρίζουμε ένα μικρό ποσοστό (για παράδειγμα το 1%) απο αυτά που έχουν το μεγαλύτερο σφάλμα ανακατασκευής και τα θεωρούμε ως έκτροπα. Ετσι το κατώφλι ορίζεται στο 1% των μεγιστων τιμών σφάλματος εκπαίδευσης (error\_train).



«»πινακες/σχηματα»»

Εχοντας υπολογίσει το κατώφλι για κάθε τύπο Αυτοκωδικοποιητή προχωράμε στον υπολογισμό των πρώτων αποτελεσμάτων. Αυτό γίνεται συγκρίνοντας απλά τα error\_test με την τιμή του κατωφλίου. Εαν το σφάλμα είναι μικρότερο απο το κατώφλι αυτό σημαίνει ότι είναι κανονικό δεδομένο ενώ αν είναι μεγαλύτερο είναι έκτροπο δεδομένο. Σχηματίζονται έτσι πίνακες με ετικέτες πρόβλεψης για κάθε Αυτοκωδικοποιητή. Για τη μέτρηση της τελικής απόδοσης του κάθε Αυτοκωδικοποιητή συγκρίνουμε τις ετικέτες πρόβλεψη με το Y\_test δηλαδή τις πραγματικές ετικέτες των δεδομένων και υπολογίζουμε το ποσοστό επιτυχημένων προβλέψεων.

«»σχηματικη απεικόνιση»»

«»να γράψω για τα roc curves και σύγκριση μεταξύ των autonecoders»»

Ακολουθεί μια σύγκριση των πέντε διαφορετικών μοντέλων Αυτοκωδικοποιητών με χρήση μετρηκής AUC (Area Under Curve). Η μετρική αυτή βασίζεται στην καμπύλη ROC (Receiver Operating Characteristics) η οποία περγράφει την απόδοση της κατηγοριοποίησης των αποτελεσμάτων ως ψευδώς θετικά και αληθώς θετικά αποτελέσματα. Οσο μεγαλύτερη περιοχή του γραφήματος βρίσκεται κάτω απο την καμπύλη τόσο μεγαλύτερος είναι ο δείκτης AUC και τόσο καλύτερη απόδοση έχουμε στην κατηγοριοποίηση, πετυχαίνοντας χαμηλό ποσοστό ψευδώς θετικών αποτελεσμάτων.



<https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>

Για την εξαγωγή των αποτελεσμάτων χρησιμοποιήθηκε το σύνολο δεδομένων Cardio

Παρατηρούμε οτι έχουμε μια μικρή διακύμανση των αποδόσεων AUC της τάξης του 4%.

# Διαδικασία Ψηφοφορίας Συνόλου

Για την επίτευξη υψηλότερης απόδοσης στις προβλέψεις της ανίχνευσης ανωμαλιών έχουν υλοποιηθεί δύο μέθοδοι Ψηφοφορίας Συνόλου. Η ψηφοφορία συνόλου πραγματοποιείται χρησιμοποιόντας συνδυαστικά όλα τα αποτελέσματα που έχουμε διαθέσιμα απο τα διαφορετικά είδη Αυτοκωδικοποιητών που χρησιμοποιήσαμε παραπάνω.

## Μέθοδος Hard Voting

Η πρώτη μέθοδος ψηφοφορίας συνόλου ανήκει στον τύπο hard voting. Ο τύπος αυτός παίρνει ως ψήφους τις τελικές ετικέτες απο την πρόβλεψη όλων των Αυτοκωδικοποιητών και παράγει ως τελικό αποτέλεσμα τον μέσο όρο τους. Ετσι ανάλογα με το σύνολο δεδομένων και τα πλεονεκτήματα του κάθε είδους Αυτοκωδικοποιητή μπορούμε να ξεπεράσουμε την καλύτερη απόδοση του ενός Αυτοκωδικοποιητή καθώς θα κερδίσουμε πετυχημένες προβλέψεις απο κάποιους Αυτοκωδικοποιητές που οι υπόλοιποι δεν ανίχνευσαν σωστά.

Πιο συγκεκριμένα η υλοποίηση του hard voting που εφαρμόστηκε, λαμβάνει υπόψη το κατώφλι του κάθε Αυτοκωδικοποιητή και εαν βρισκόμαστε κάτω απο το κατώφλι καταχωρείται ψήφος -1 και εαν βρισκόμαστε πάνω απο το κατώφλι ψήφος +1. Αφού αποθηκευτούν όλες οι ψήφοι όλων των Αυτοκωδικοποιητών γίνεται πρόσθεση μεταξύ τους και εαν το αποτέλεσμα είναι μικρότερο του μηδενός η τελική πρόβλεψη είναι κανονική ενω εάν είναι μεγαλύτερο του μηδενός είναι έκτροπη.

«» γραφήματα/εικόνες για hard voting, αποτελέσματα negative voting»»

## Μέθοδος Soft Voting

Η δεύτερη μέθοδος ψηφοφορίας συνόλου ανήκει στον τύπο soft voting. Ο τύπος αυτός απαιτεί την χρήση πιθανοτήτων ή βαθμού σιγουριάς απο κάθε μέλος για τον τελικό υπολογισμό του αποτελέσματος. Με την μέθοδο αυτή μπορούμε να πετύχουμε καλύτερη τελικά απόδοση όταν οι αποδόσεις των Αυτοκωδικοποιητών βρίσκονται κοντά στο κατώφλι, καθώς παίρνουμε όλη την πληροφορία για το βαθμό σιγουριάς του κάθε Αυτοκωδικοποιητή.

Η μέθοδος αυτή είναι πιο πολύπλοκη στη διαδικασία απο την Hard Voting μέθοδο διοτι χρειάζεται να μετατρέψουμε τις τιμές εξόδου των αυτοκωδικοποιητών σε μια μορφή συγκρίσιμης πιθανότητας εκφράζοντας ετσι την απόσταση από το κατώφλι. Ετσι αρχικά παίρνουμε τις αποστάσεις απο το κατώφλι που εχουμε υπολογίσει στα προηγούμενα στάδια και εφαρμόζουμε τον παρακάτω τύπο ωστε να μετατραπούν σε πιθανότητες.

Η τιμή mean αναφέρεται στο μέσο όρο των αποστάσεων από το κατώφλι. Στο πρώτο σκέλος το αποτέλεσμα χαρακτηρίζεται ως έκτροπο και υπολογίζεται ως πιθανότητα ο λόγος της απόστασης του σφάλματος απο το κατώφλι δια την απόσταση της μέσης τιμής απο το κατώφλι και προστίθεται το 0.5, διότι το 0.5 θα χρησιμοποιηθεί ως το όριο στην διαδικασία ψηφοφορίας. Στο δεύτερο σκέλος το αποτέλεσμα χαρακτηρίζεται ως κανονικό και η πιθανότητα υπολογίζεται ως ο λόγος της απόστασης του σφάλματος απο τη μέση τιμή δια την απόσταση της μέσης τιμής απο το κατώφλι. Ετσι έχουμε εκφράσει τις αποστάσεις απο το κατώφλι ως πιθανότητες με κατώφλι το 0.5 και εχουμε εκφράσει μέσα απο τις πιθανότητες αυτές τον βαθμό σιγουριάς του κάθε Αυτοκωδικοποιητή για κάθε αποτέλεσμα.

Ακολουθεί η διαδικασία ψηφοφορίας στην οποία για κάθε δεδομένο απο το σύνολο των αποτελεσμάτων υπολογίζεται ο μέσος όρος των πιθανοτήτων που υπολογίσαμε στο προηγούμενο βήμα και συγκρίνεται με το κατώφλι των 0.5 ωστε να αποδωθεί η ψήφος για κανονικό ή έκτροπο αποτέλεσμα.

# Τελικά Αποτελέσματα

Για την εξαγωγή των τελικών αποτελεσμάτων έχει υλοποιηθεί επαναληπτική διαδικασία 10 επαναλήψεων της διαδικασίας που περιγράφηκε και έχει γίνει μέσος όρος όλων των αποτελεσμάτων για να αποφευχθούν μεμονομένα περιστατικά πολύ υψηλής ή πολύ χαμηλής απόδοσης που οφείλονται στην τυχαιότητα της διαδικασίας εκπαίδευσης των Αυτοκωδικοποιητών.

Στα γραφήματα αποδόσεων έχει χρησιμοποιηθεί η καλύτερη απόδοση απο τα πέντε μοντέλα Αυτοκωδικοποιητών σε σύγκριση με το αποτέλεσμα της μεθόδου Soft Voting και της μεθόδου Hard Voting.

# Συμπεράσματα