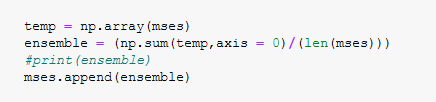
Είδη autoencoders:

1. Deep Autoencoder
2. Denoising Autoencoder (με χρήση Gaussian noise στο dataset και μεγέθους 200% των τιμών του dataset)
3. Sparse Autoencoder (με χρήση activity regularization)
4. Sparse Autoencoder2 (με χρήση activity regularization και dropout layers)
5. Contractive Autoencoder (με χρήση L2 regularization)
6. Ensemble (απλός μέσος όρος των mean squared error των παραπάνω)
7. Ensemble2 (Average layer όλων των δικτύων εκτός του Ensemble που ακολουθείται απο Dense layer και ειναι trainable)

Χρήση τριών datasets: cardio(1831 rows,21 cols), ionosphere(351 rows, 33 cols) και satellite(6435 rows, 36 cols)

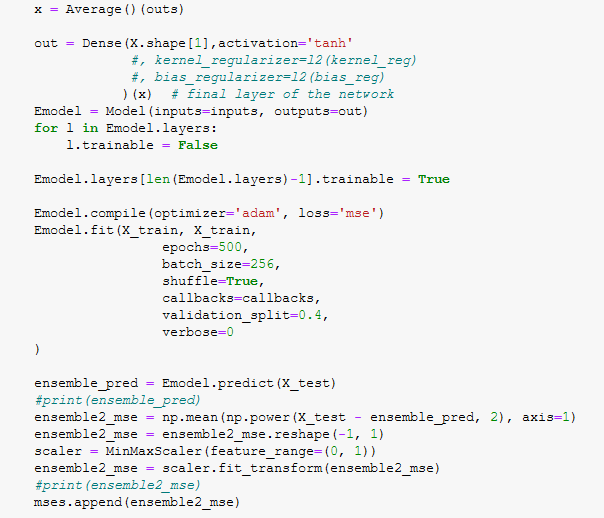
Δημιουργία του Ensemble απο το μέσο όρο των mean squared error των μοντέλων:



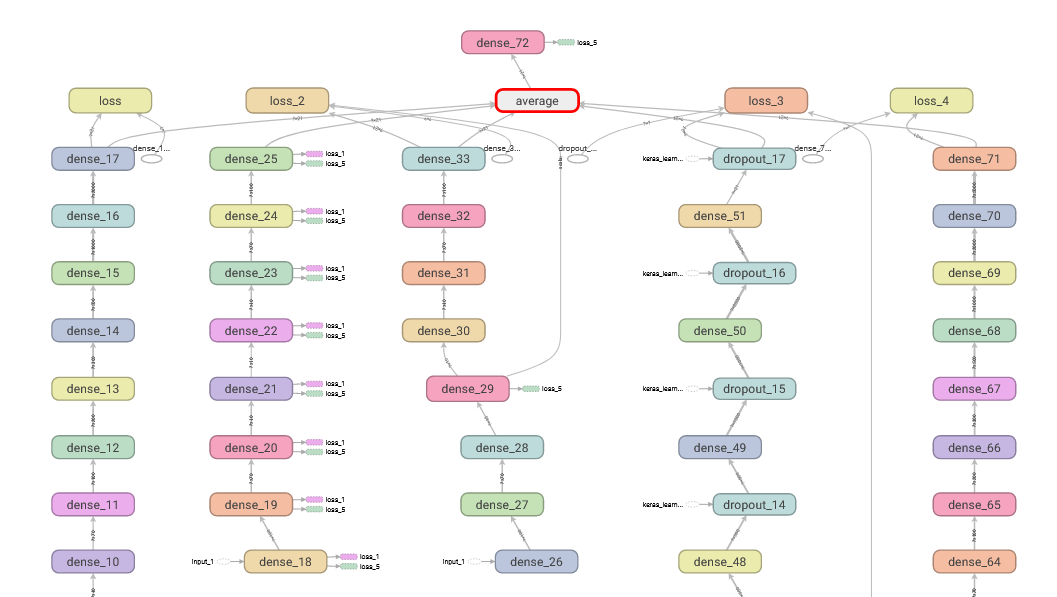
Όπου mses είναι η λίστα με όλα τα mean squared errors των μοντέλων της κάθε δοκιμής, δηλαδή των μοντέλων με διαφορετικό βάθος και ίδια αρχιτεκτονική για τα πρώτα γραφήματα και όλων των μοντέλων (deep, denoising, sparse, sparse2, contractive) για τα δεύτερα γραφήματα.

Ουσιαστικά το Ensemble δεν χρειάζεται να υπάρχει σαν μοντέλο διότι αντί για δίκτυο χρησιμοποιούμε τα αποτελέσματα των υπολοίπων μοντέλων.

Δημιουργία του Ensemble2 συνδιάζοντας τα μοντέλα. Οι επιλογές αρχιτεκτονικών γίνονται όπως και στον Ensemble παραπάνω:



Η μεταβλική outs περιέχει το τελευταίο Layer του κάθε μοντέλου οπότε χρησιμοποιώντας ως input του Ensemble2 το inputs το οποίο χρησιμοποιείται απο κοινού σε όλα τα μοντέλα και σαν output το συνδυασμό των outs σε ένα Average Layer το οποίο έχει στη συνέχεια ένα Dense Layer που εκπαιδεύεται και δίνει το τελικό output του Ensemble2. Ενα γράφημα του Ensemble2 είναι το παρακάτω:



Το γράφημα είναι αντεστραμένο (η έξοδος είναι πάνω και οι είσοδοι απο κάτω). Το Layer dense\_72 είναι το τελικό Layer που εκπαιδεύεται για τις επιλογές του Ensemble2. Ολα τα υπόλοιπα layers δεν ειναι trainable έτσι ώστε να διατηρήσουν τη δομή και τα αποτελέσματα που είχαν.

Διαφορετικά βάθη απο Layers απο 1 εως 8 (π.χ. για βάθος 5 έχουμε 5 layer για τον encoder, 1 μεσαίο layer και 5 layers για τον decoder), έγινε μέσος όρος αποτελεσμάτων απο 20 δοκιμές.

Παρατηρούμε ότι για τον Deep και τον Denoising Autoencoder έχουμε καλύτερη απόδοση με πιο βαθύ δίκτυο, για τους υπόλοιπους το βάθος δεν επιρεάζει το αποτέλεσμα διότι έχουν κάποια μορφή regularization.

Τα Ensemble παίρνουν ως μέσο όρο ένα είδος Autoencoder τη φορά, δηλαδή κάνουν συνδυασμό πχ του Contracting model για τα διαφορετικά βάθη δικτύου, η σύγκριση των διαφορετικών τύπων γίνεται παρακάτω.

Παρατηρούμε ότι ο Deep Autoencoder έχει καλύτερη απόδοση στα 4 Layers και κάτω στο συγκεκριμένο dataset, επίσης και ο Denoising Autoencoder έχει αρκετά καλή απόδοση. Αυτό μπορεί να γίνεται λόγω του μικρού μεγέθους του dataset (351 points) που ευνοεί τα δίκτυα χωρίς κάποια μορφή regularization να αποδόσουν καλύτερα.

Παρατηρούμε ότι και πάλι το Deep model έχει καλύτερη απόδοση παρ όλο που συνολικά οι αποδόσεις δεν απέχουν πολύ (2% καλύτερο). Ενδέχεται να έχουμε φτάσει στα όρια του dataset.

Οσον αφορά τα 2 Ensemble δίκτυα παρατηρούμε ότι στο Cardio έχουν περίπου τις ίδιες αποδόσεις ενώ στα ionosphere και satellite το εκπαιδευμένο Ensemble υπερτερεί του απλού μέσου όρου.

Παρακάτω η σύγκριση των διαφορετικών αρχιτεκτονικών. Στη σύγκριση χρησιμοποιήθηκε το καλύτερο βάθος δικτύου για την κάθε αρχιτεκτονική σύμφωνα με τα παρπάνω και έγινε μέσος όρος αποτελεσμάτων απο 20 δοκιμές.

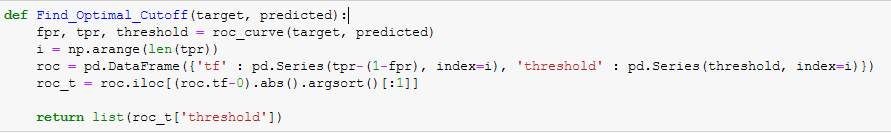
Στο cardio dataset παρατηρούμε πολύ μικρή διαφορά ανάμεσα στις αρχιτεκτονικές.

Στο ionosphere ο Deep Autoencoder έχει τη μέγιστη απόδοση και ακολουθεί το εκπαιδευμένο Ensemble

Στo satellite dataset βλέπουμε ότι παρ’ όλο που έχουμε πολύ μικρές διαφορές στις αποδόσεις (περίπου 1-2%) το εκπαιδευμένο Ensemble έχει την καλύτερη απόδοση απο όλα τα υπόλοιπα δίκτυα.

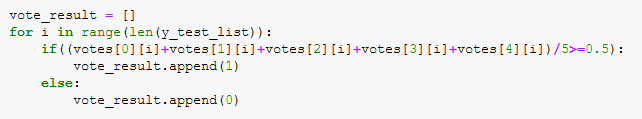
**Voting Ensemble**

Δημιούργησα μια συνάρτηση που επιλέγει ως threshold την τιμή με το καλύτερο δυνατό balance απο true positive και false positive:



Η συνάρτηση αυτή χρησιμοποιείται για την επιλογή του threshold κάθε μοντέλου και στη συνέχει κατηγοριοποιούνται τα δεδομένα σε outliers και inliers και συγκρίνονται με το y\_test που είναι το επιθυμητό.

Το Voting Ensemble παίρνει τις τιμές εξόδου των μοντέλων μετά την εφαρμογή του threshold, υπολογίζει το μέσο όρο τους και αν ο μέσος όρος είναι μεγαλύτερος ή ίσος του 0.5 τότε το αποτέλεσμα ορίζεται σαν outlier αλλιώς ορίζεται σαν inlier. (οι outliers έχουν τιμή 1 και οι inliers τιμή 0 στο dataset).

Στο Voting Ensemble δεν λαμβάνονται υπόψη τα άλλα 2 Ensembles (Ensemble1 και Ensemble2). Εκανα δοκιμές στις οποίες τα λάμβανα υπόψη και τα αποτελέσματα ήταν ίδια.  


Τα παρακάτω γραφήματα δείχνουν την πραγματική απόδοση πρόβλεψης των μοντέλων με το επιλεγμένο threshold και το αποτέλεσμα του Voting Ensemble:

Το cardio dataset είναι το μόνο στο οποίο το Voting Ensemble ξεπερνάει το Ensemble2 στο οποίο και απο τα παραπάνω tests το Ensemble2 δεν ειχε τόσο καλή απόδοση.

Επίσης είναι ορατή και εδώ η ανικανότητα οποιουδήποτε μοντέλουν να ξεπεράσει το 70% στην απόδοση του satellite dataset