Υπολογιστική Νοημοσύνη

Εργασία Μέρος Α'

Vol. VI, No. 2, Summer 1958

research trends

CORNELL AERONAUTICAL LABORATORY, INC., BUFFALO 21, NEW YORK

The Design of an

Intelligent AUTOMATON

by FRANK ROSENBLATT

Introducing the perceptron — A machine which senses, recognizes, remembers, and responds like the human mind.

STORIES about the creation of machines having human qualities have long been a fascinating province in the realm of science fiction. Yet we are now about to witness the birth of such a machine — a machine capable of perceiving, recognizing, and identifying its surroundings without any human training or control.

Development of that machine has stemmed from a search for an understanding of the physical mechanisms which underlie human experience and intelligence. The question of the nature of these processes is at least as ancient as any other question in western science and philosophy, and, indeed, ranks as one of the greatest scientific challenges of our time.

Our understanding of this problem has gone perhaps as far as had the development of physics before Newton. We have some excellent descriptions of the phenomena to be explained, a number of interesting hypotheses, and a little detailed knowledge about events in the nervous system. But we lack agreement on any integrated set of principles by which the functioning of the nervous system can be understood.

We believe now that this ancient problem is about to yield to our theoretical investigation for three reasons:

First, in recent years our knowledge of the functioning of individual cells in the central nervous system has vastly increased.

Second, large numbers of engineers and mathematicians are, for the first time, undertaking serious study of the mathematical basis for thinking, perception, and the handling of information by the central nervous system, thus providing the hope that these problems may be within our intellectual grasp.

Third, recent developments in probability theory and in the mathematics of random processes provide new tools for the study of events in the nervous system, where only the gross statistical organization is known and the precise cell-by-cell "wiring diagram" may never be obtained.

Receives Navy Support

In July, 1957, Project PARA (Perceiving and Recognizing Automaton), an internal research program which had been in progress for over a year at Cornell Aeronautical Laboratory, received the support of the Office of Naval Research. The program had been concerned primarily with the application of probability theory to

Τσικέλης Ιωάννης

1067407

Τμήμα Μηχανικών Η/Υ και Πληροφορικής

Πανεπιστήμιο Πατρών

2021-2022

Table Of Contents

Σύντομη περιγραφή

Στο πρώτο μέρος της εργασίας του μαθήματος Υπολογιστική Νοημοσύνη, κληθήκαμε να υλοποιήσουμε ένα πλήθος τεχνητών νευρωνικών δικτύων, τα οποία κατατάσσουν έναν αριθμό κειμένων σε 20 διαφορετικές κατηγορίες. Για τον έλεγχο των υλοποιήσεων χρησιμοποιήθκε το dataset DeliciousMIL. Με πάνω από 12.000 διαφορετικές εγγραφές για training και testing. Ο κώδικας γράφτηκε σε γλώσσα Python και έκανε χρήση του πακέτου Tensorflow για την ανάπτυξη των νευρωνικών δικτύων που απαιτούνταν.

Κώδικας Υλοποίησης:

Ο κώδικας της άσκησης είναι διαθέσιμος στο GitHub repository: itsikelis/ceid-comp-int-project.

A1. Προεπεξεργασία και Προετοιμασία δεδομένων

Α) Κωδικοποίηση Εισόδων:

Για την κωδικοποίηση της εισόδου στο νευρωνικό χρησιμοποιήθηκε το μοντέλο Bag Of Words (BoW). Σύμφωνα με αυτό, για κάθε γραμμή του dataset (που αντιστοιχούσε σε ένα ξεχωριστό κείμενο) δημιουργήθηκε ένα διάνυσμα 8520 θέσεων, με κάθε θέση να αντιστοιχεί στις 8520 διαφορετικές λέξεις που βρίσκονται συνολικά στα κείμενα, και σε κάθε θέση τοποθετήθηκε ένας ακέραιος αριθμός που αντιστοιχεί στο πλήθος εμφανίσεων της εκάστοτε λέξης στο συγκεκριμένο κείμενο.

Β) Κεντράρισμα – Κανονικοποίηση – Τυποποίηση Δεδομένων:

Κατά το στάδιο της προεπεξεργασίας των δεδομένων, αποφασίστηκε να γίνει τυποποίηση (standardisation) στις στήλες του μητρώου εισόδου. Η τυποποίηση, αφαιρεί τον μέσο όρο των στηλών από την τιμή της κάθε στήλης (νέο κέντρο στο 0) και θέτει την τυπική απόκλιση (σ) στο 1. Αν και θεωρητικά η διακύμανση των διαφρετικών τιμών μπορεί να διορθωθεί κατά την διάρκεια της εκπαίδευσης, με τη σωστή εναλλαγή των βαρών, το βήμα αυτό ορισμένες φορές, μεταξύ άλλων, βοηθάει το μοντέλο να συγκλίνει γρηγορότερα.

Η επιλογή του βήματος τυποποίησης έγινε με βάση το παρόν άρθρο και μετά από προσωπικό πειραματισμό και σύγκριση αποτελεσμάτων.

Γ) Διασταυρούμενη Επικύρωση (Cross-Validation):

Κατά το στάδιο του cross validation, το dataset διασπάται σε προκαθορισμένα κομμάτια, από τα οποία τυχαία, ορισμένα επιλέγονται για την εκπαίδευση και τα υπόλοιπα για την επικύρωση του μοντέλου. Η διαδικασία αυτή επαναλαμβάνεται k φορές (στην περίπτωης της άσκησης 5). Για την παρουσίαση των διαγραμμάτων παρακάτω, επιλέχθηκε να απεικονιστεί ο καλύτερος (κατά cross entropy) κύκλος εκπαίδευσης, με διαδικασία που γινόταν κατά το runtime.

Α2. Επιλογή αρχιτεκτονικής

Α) Cross Entropy – MSE – Accuracy:

Για την αξιολόγηση των μονέλων χρησιμοποιήθηκαν οι εξής μετρικές:

- **Categorical Cross Entropy**: Η μορφή του cross entropy που χρησιμοποιείται στα multiclass multilabel classification προβήματα, λαμβάνει υπ' όψην της την απόκλιση του κάθε στοιχείου του διανύσματος εξόδου ξεχωριστά συγκρίνοντάς την επιθυμητή έξοδο.
- **Mean Square Error**: Η μέση τιμή των αποκλίσεων των τιμών του διανύσματος εξόδου με την επιθυμητή έξοδο, υψωμένη στο τετράγωνο, για την αποφυγή αρνητικών τιμών που μπορεί εσφαλμένα να μειώνουν το αποτέλεσμα. Όσο πιο μικρό το MSE, τόσο πιο κοντά είναι η έξοδος του μοντέλου στο πραγματικό αποτέλεσμα.
- **Accuracy**: Το πλήθος των σωστών ππροβλέψεων, προς το πλήθος των συνολικών προβλέψεων του μοντέλου. Όσο πλησιάζει 1, τόσο πιο ακριβές είναι ένα μοντέλο για το δοθέν dataset.

Β) Νευρώνες Εισόδου Και Εξόδου:

Το μοντέλο δέχεται ως είσοδο διανύσματα διαστάσεων 1x8520 και ελέγχει σε ποιες από τις 20 κατηγορίες εμπίπτει. Επομένως για την είσοδο και την έξοδο του νευρωνικού χρησιμοποιήθηκαν 8520 και 20 νευρώνες αντίστοιχα.

Γ) Συνάρτηση Ενεργοποίησης Κρυφών Κόμβων:

Για την ενεργοποίηση των κρυφών κόμβων, επιλέχθηκε η συνάρτηση rectifier (ReLU) κυρίως λόγω της ευκολίας υπολογισμού της καθώς και της ευρείας χρήσης της στην βιομηχανία και την έρευνα.

Δ) Συνάρτηση Ενεργοποίησης Εξόδου:

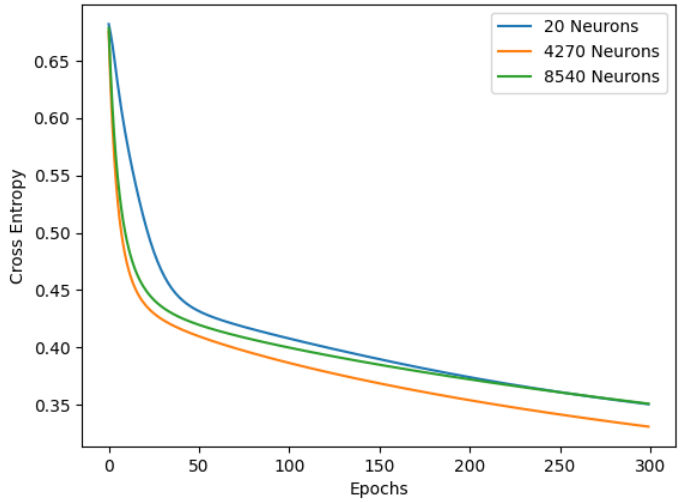
Για το επίπεδο εξόδου, χρησιμοποιήθηκε η σιγμοειδής (sigmoid) συνάρτηση, καθώς ενδείκνειται για προβλήματα ταξινόμησης multiclass multilabel. Η έξοδός της εκφράζει την πιθανότητα ένα κείμενο να ανήκει στην κατηγορία που συμβολίζει ο νευρώνας που ενεργοποιήθηκε.

Ε) Αριθμοί Νευρώνων Πρώτου Κρυφού Επιπέδου:

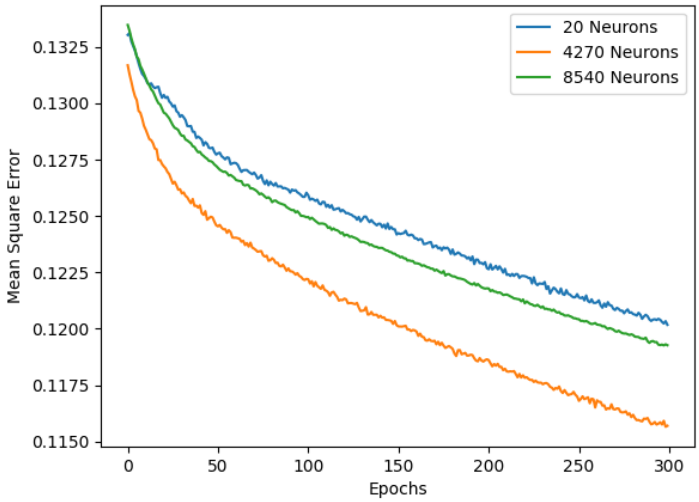
Για το πρώτο κρυφό επίπεδο, εκτελέστηκαν τρία πειράματα με διαφορετικό αριθμό κόμβων κρυφού επιπέδου. Διατηρήθηκαν τα αποτελέσματα του καλύτερου fold, με βάση το χαμηλότερο Cross-Entropy:

# of Neurons	Cross Entropy	Mean Square Error	Accuracy
20	0.3549	0.1167	0.3109
4720	0.3529	0.1210	0.3618
8540	0.3684	0.1198	0.3529

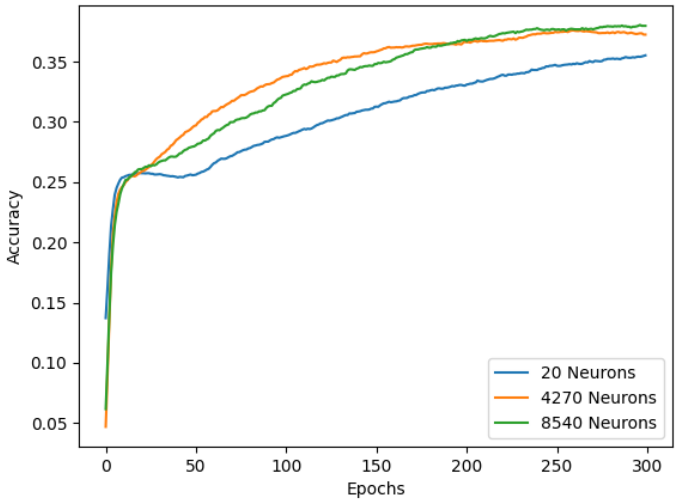
Κατά την εκπαίδευση, τα διαγράμματα για τις μετρικές προέκυψαν ως εξής:



Εικόνα 1: Cross Entropy - Epochs



Εικόνα 2: MSE - Epochs



Εικόνα 3: Accuracy – Epochs

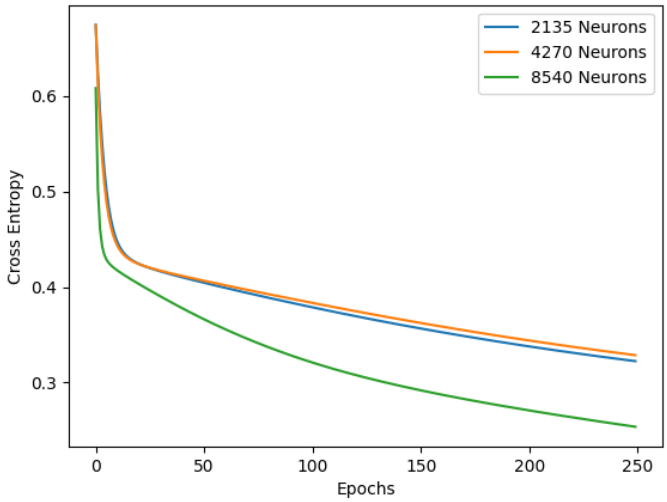
Παρατηρείται πως το μοντέλο με τους 4270 νευρώνες στο ενδιάμεσο επίπεδο πετυχαίνει την χαμηλότερη τιμή της συνάρτησης κόστους(categorical cross entropy) και πολύ γρηγορότερο μηδενισμό του μέσου τετραγωνικού σφάλματος. Το μοντέλο με τους 8540 νευρώνες στο ενδιάμεσο επίπεδο πέτυχε καλύτερη ακρίβεια. Όλα τα μοντέλα όμως παρουσιάζουν προβληματικά αποτελέσματα σχετικά με το overfitting, όπως θα φανεί και παρακάτω.

ΣΤ) Δεύτερο Κρυφό Επίπεδο:

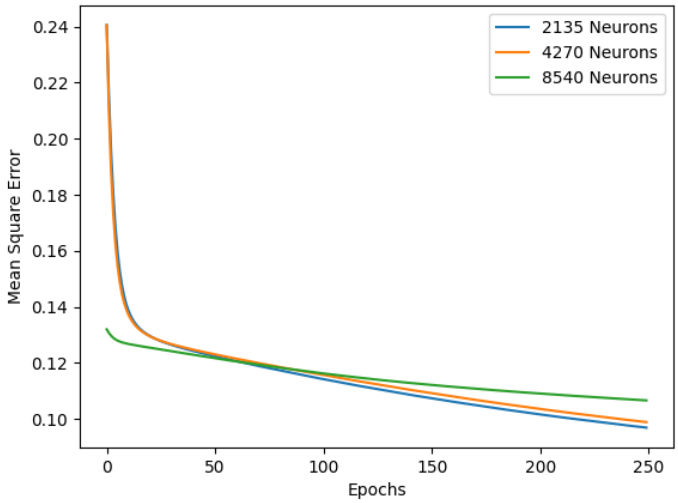
Για την παρούσα υλοποίηση, αποφασίστηκε η χρήση 4270 νευρώνων στο πρώτο κρυφό επίπεδο, με βάση την προηγούμενη επίδοση των μοντέλων. Έπειτα, πραγματοποιήθηκαν τα εξής πειράματα για τον καθορισμό των νευρώνων του δεύτερου κρυφού επιπέδου (πάλι, διατηρήθηκαν τα αποτελέσματα του καλύτερου fold, με βάση το χαμηλότερο Cross-Entropy):

# of Neurons	Loss	MSE	Accuracy
2135	0.3391	0.1015	0.3622
4270	0.3422	0.1029	0.3552
8540	0.3006	0.1123	0.3188

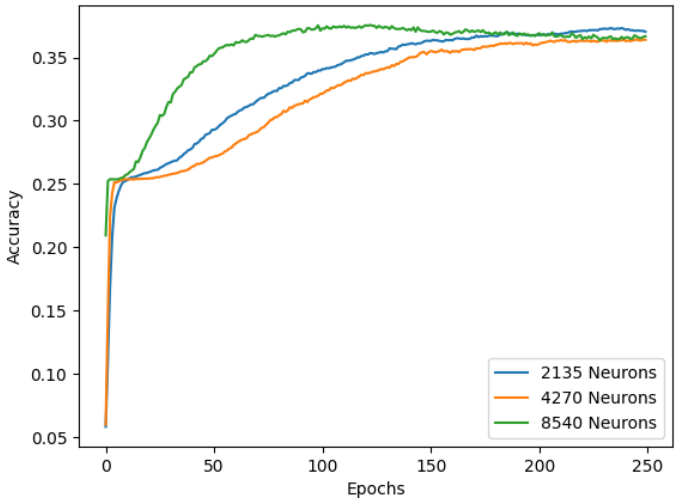
Κατά την εκπαίδευση, τα διαγράμματα για τις μετρικές προέκυψαν ως εξής:



Εικόνα 4: Cross-Entropy - Epochs



Εικόνα 5: MSE - Epochs



Εικόνα 6: Accuracy - Epochs

Από τα διαγράμματα, παρατηρείται πως το μοντέλο με τους 8540 νευρώνες στο δεύτερο κρυφό δίκτυο μειώνει πολύ γρηγορότερα την συνάρτηση κόστους. Το μοντέλο με τους 2135 όμως πετυχαίνει χαμηλότερο μέσο τετραγωνικό σφάλμα και καλύτερη ακρίβεια στο βάθος του χρόνου. Φαίνεται πως η επιλογή μίας μέσης επιλογής πλήθους νευρώνων μεταξύ των απιπέδων λειτουργεί καλύτερα. Για την συνέχεια των πειραμάτων θα χρησιμοποιηθεί το μοντέλο με τους 4270 και 2135 νευρώνες στο πρώτο και στο δεύτερο κρυφό επίπεδο αντίστοιχα.

Ζ) Κριτήριο Τερματισμού:

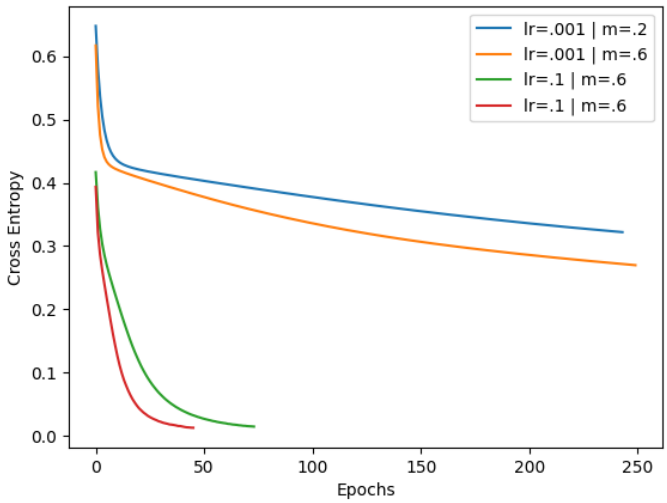
Από τα διαγράμματα, φαίνεται πως τα μοντέλα που υλοποιήθηκαν συνήθως συγκλίνουν ικανοποιητικά μετά τις 100-150 πρώτες εποχές. Για την μείωση του χρόνου υπολογισμού των μοντέλων, καθώς και για την αποφυγή φαινομένων υπερεκπαίδευσης, μπορεί να χρησιμοποιηθεί η τεχνική του early stopping. Στις παρούσες υλοποιήσεις, δημιουργήθηκε μία αρκετά αισιόδοξη συνθήκη που ήλεγχε την loss function(cross entropy) του μοντέλου για να παρατηρήσει πότε αυτή έμενε στάσιμη με διαφορές μικρότερες το 0.0006. Αυτή η συνθήκη δεν ικανοποιούταν σχεδόν ποτέ. Έτσι, σε κάποια ερωτήματα χρησιμοποιήθηκε και early stopping όταν το Accuracy άρχιζε να συγκλίνει με διαφορές μικρότερες του 0.0008. Η τιμή αυτή θεωρήθηκε επίσης, μετά από παρατήρηση των γραφικών παραστάσεων που παρατέθηκαν παραπάνω.

Α3. Μεταβολές στον ρυθμό εκπαίδευσης και σταθεράς ορμής

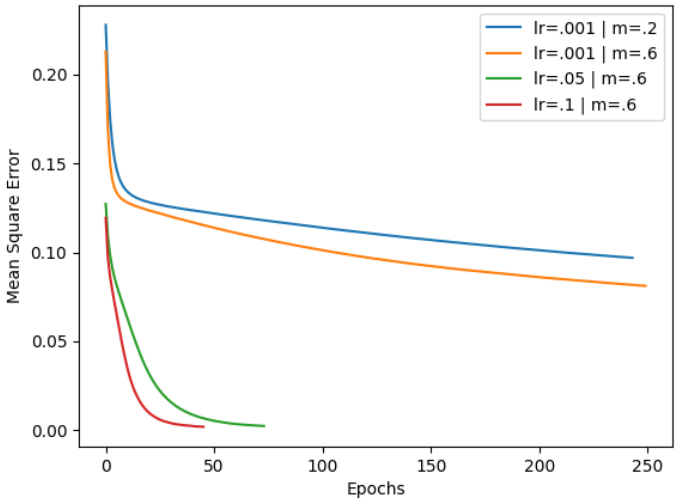
Με βάση την προηγούμενη βέλτιστη τοπολογία για το δίκτυο, πραγματοποιήθηκαν οι εξής πειραματισμοί για τον καθορισμό του ρυθμού μάθησης και της σταθεράς ορμής. Τα αποτελέσματα φαίνονται παρακάτω:

Learning Rate	Momentum	Loss	MSE	Accuracy
0.001	0.2	0.3378	0.1012	0.3605
0.001	0.6	0.3075	0.0924	0.3316
0.05	0.6	0.6737	0.1079	0.2939
0.1	0.6	0.7245	0.1087	0.2832

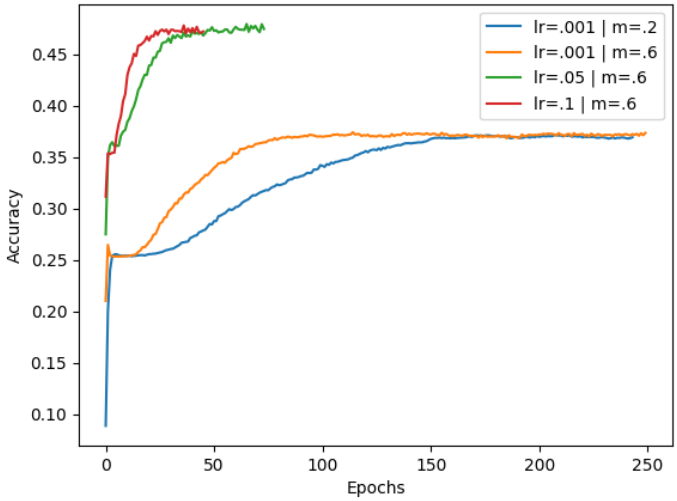
Κατά την εκπαίδευση, τα διαγράμματα για τις μετρικές προέκυψαν ως εξής:



Εικόνα 7:Cross-Entropy – Epochs



Εικόνα 8: MSE – Epochs



Εικόνα 9: Accuracy – Epochs

Στα διαγράμματα παρατηρείται το πώς η αύξηση του ρυθμού μάθησης επηρεάζει την ταχύτητα σύγκλισης του μοντέλου. Αυτό επαληθεύεται και από την θεωρία, σύμφωνα με την οποία ο ρυθμός μάθησης είναι η σταθερά που καθορίζει το ποσοστό αλλαγής των βαρών ανά εποχή εκπαίδευσης. Όσο μεγαλύτερη, τόσο περισσότερο το μοντέλο θα “μαθαίνει” από τα νέα δεδομένα και αντίστοιχα θα “ξεχνά” από τα παλιά.

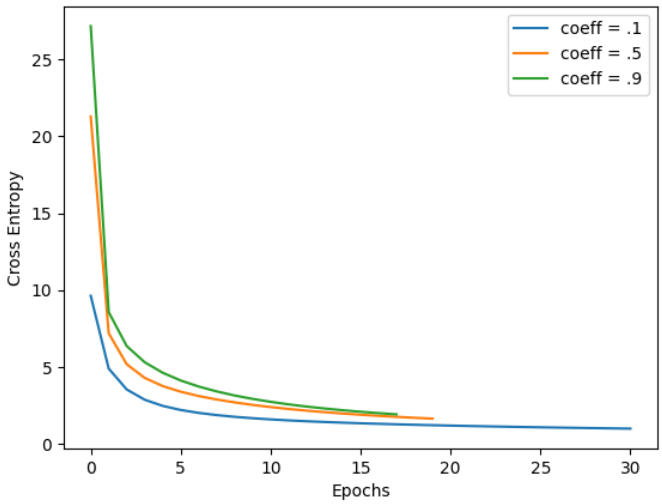
Επιπρόσθετα, παρατηρείται και πως η αυηξημένη σταθερά ορμής συμβάλει στην ταχύτερη σύγκλιση του μοντέλου. Η σταθερά ορμής εκμεταλλεύεται την κατεύθυνση της συνάρτησης κόστους και μεταπηδά όλο και πιο μακριά σε κάθε επανάληψη, με βάση την προηγούμενη αποκτειθήσα “ορμή”, με σκοπό την αποφυγή περιπτώσεων τοπικών ελαχίστων.

Α4. Ομαλοποίηση

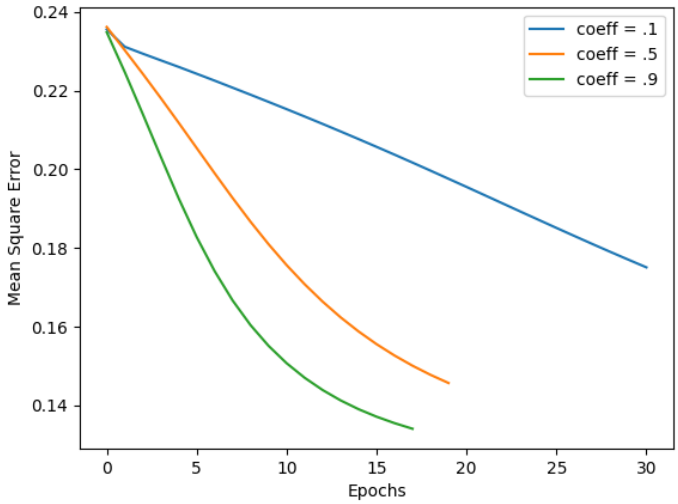
Για την αποφυγή υπερ-εκπαίδευσης του μοντέλου, χρησιμοποιήθηκε η μέθοδος της l2 ομαλοποίησης των βαρών του κάθε επιπέδου(επιλογή εφαρμογής στην έξοδο του κάθε επιπέδου). Τα αποτελέσματα έχουν ως εξής:

L2 Coefficient	Loss	MSE	Accuracy
0.1	1.1951	0.1736	0.2568
0.5	2.425	0.1446	0.2555
0.9	3.1741	0.1335	0.2545

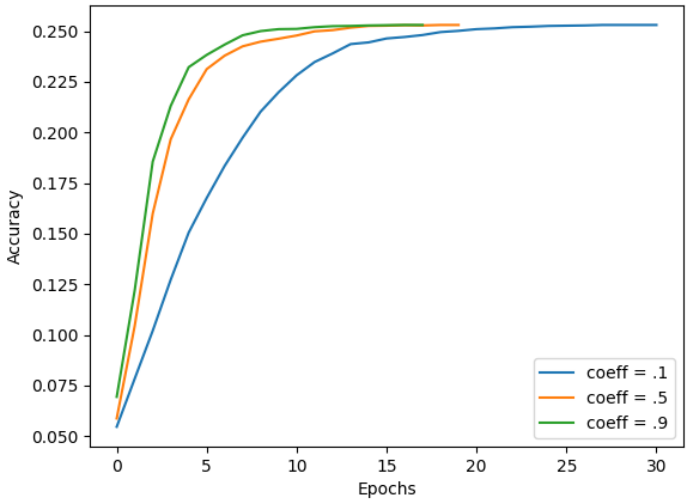
Κατά την εκπαίδευση, τα διαγράμματα για τις μετρικές προέκυψαν ως εξής:



Εικόνα 10: Cross-Entropy – Epochs



Εικόνα: MSE – Epochs



Εικόνα 11: Accuracy – Epochs

Αυτό που παρατηρείται απ' ευθείας είναι η κατά τουλάχιστον 0.05 μονάδες μικρότερη ακρίβεια που πετυχαίνουν τα νέα μοντέλα. Εφόσον η μέθοδος της ομαλοποίησης φροντίζει να αποφεύγεται η υπερ-εκπαίδευση ενός συνόλου δεδομένων, είναι φυσικό πλέον το μοντέλο να πετυχαίνει καλύτερη γενίκευση, με λιγότερη ακρίβεια. Ιδιαίτερη προσοχή όμως πρέπει να δοθεί στις περιπτώσεις που η σταθερά είναι πολύ μεγάλη, με αποτέλεσμα το μοντέλο να κάνει underfit.

A5. Ενσωματώσεις Λέξεων (Bonus)

A) Προεπεξεργασία Δεδομένων (Padding):

Για τη μέθοδο κατηγοριοποίησης με ενσωματώσεις λέξεων, υλοποιήθηκε μία συνάρτηση η οποία κάνει padding στα δεδομένα εισόδου. Συγκεκριμένα, οι λέξεις του dataset φορτώνονται σε ένα μητρώο, το οποίο έχει μέγιστο αριθμό στηλών όσο το μεγαλύτερο κείμενο του dataset. Στις κενές θέσεις του μητρώου, μετά από στοίχιση των δεδομένων αριστερά, προστίθενται μηδενικά.

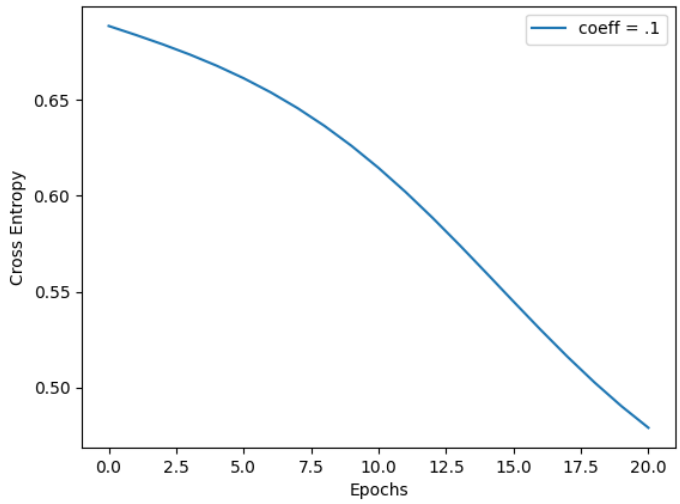
Το μητρώο αυτό εισάγεται σε ένα embedding layer, το οποίο παράγει έναν τένσορα, οποίος μεταφέρεται μέσα από ένα flatten layer το οποίο μετατρέπει τον βαθμό του σε 2 για να περάσει μετά από τα επόμενα επίπεδα του νευρωνικού.

B) Μοντέλο Word Embedding:

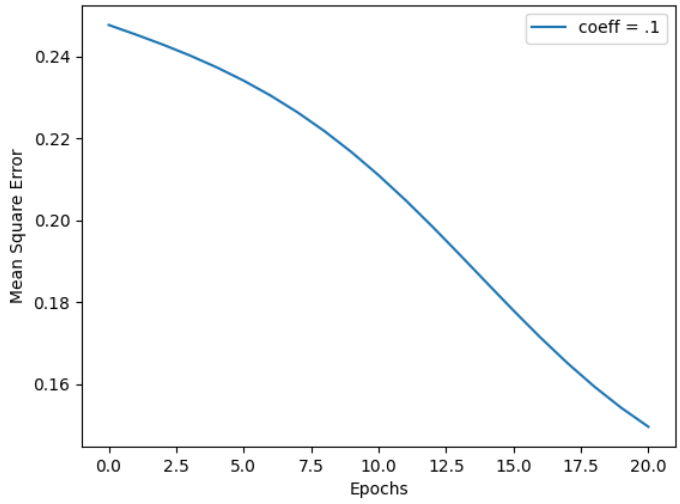
Τροποποιώντας το καλυτερο μοντέλο που προέκυψε από το ερώτημα Α2 (1 κρυφό δίκτυο, 4270 νευρώνες) για να δέχεται είσοδο από Word Embeddings, προέκυψαν τα εξής αποτελέσματα:

Loss	MSE	Accuracy
0.4732	0.1471	0.2552

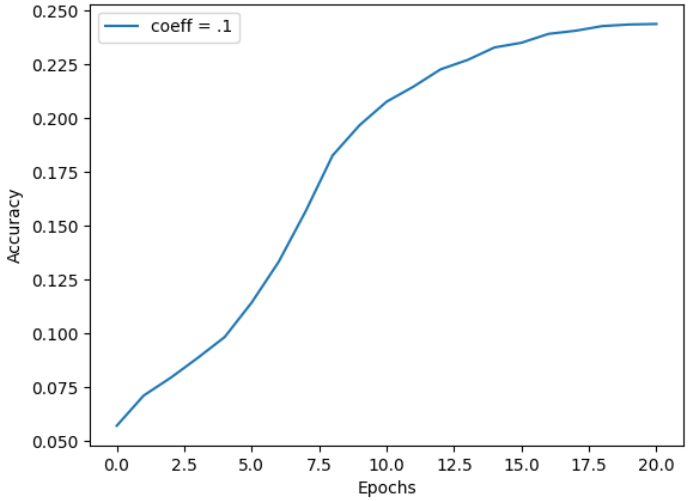
Κατά την εκπαίδευση, τα διαγράμματα για τις μετρικές προέκυψαν ως εξής:



Εικόνα 12: Cross-Entopy - Epochs



Εικόνα 13: MSE - Epochs



Εικόνα 14: Accuracy – MSE

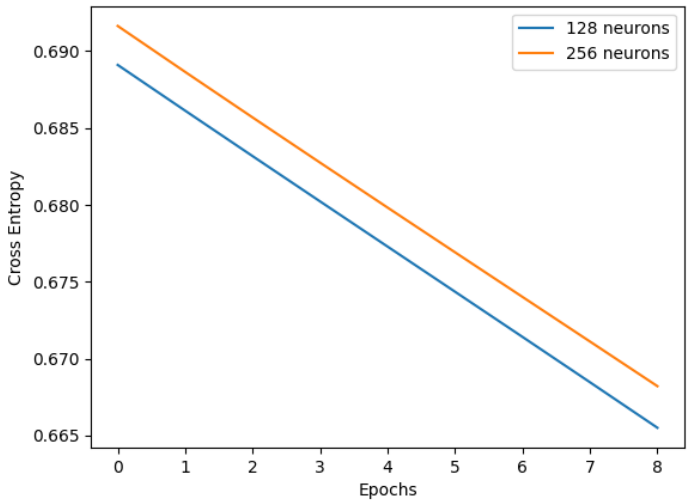
Εδώ παρατηρείται η μεγαλύτερη ταχύτητα σύγκλισης του μοντέλου σε τιμές περίπου ίδιες με την περίπτωση της L2 ομαλοποίησης, που σημαίνει πως μάλλον το μοντέλο δεν έχει υπερεκπαιδευτεί.

Γ) Long-Short Term Memory:

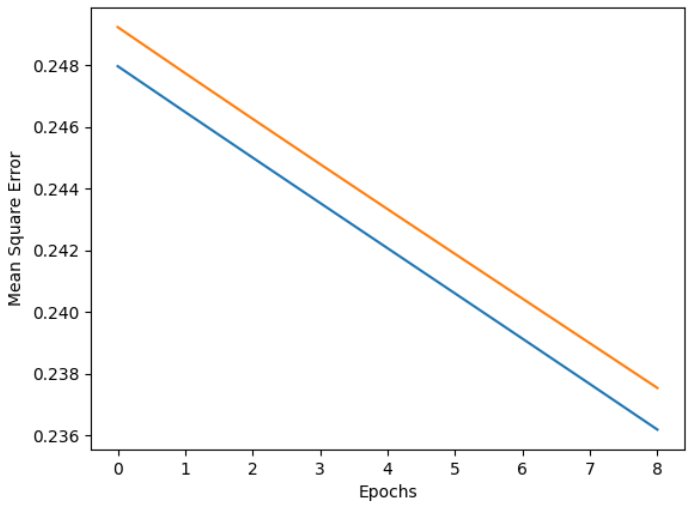
Για το παρόν ερώτημα, δημιουργήθηκαν δύο μοντέλα Long-Short Term Memory. Τα αποτελέσματά τους έχουν ως εξής:

Neurons (LSTM Layer)	Loss	MSE	Accuracy
128	0.664	0.2354	0.01
256	0.6667	0.2368	0.2619

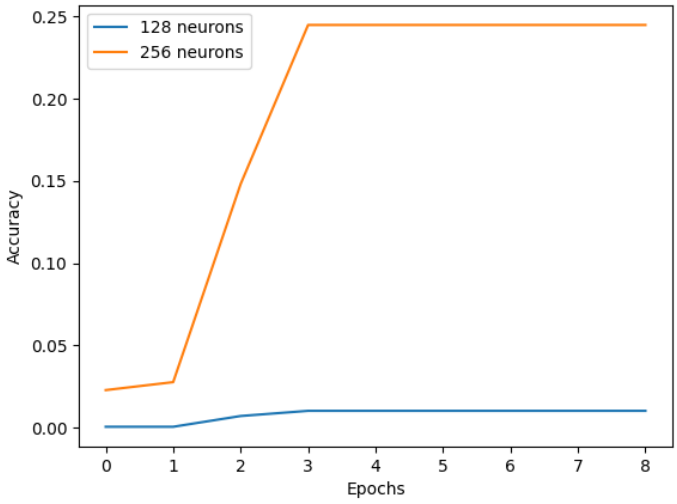
Κατά την εκπαίδευση, τα διαγράμματα για τις μετρικές προέκυψαν ως εξής:



Εικόνα 15: Cross-Entropy - Epochs



Εικόνα 16: MSE - Epochs



Εικόνα 17: Accuracy - Epochs

Εδώ, στην περίπτωση των 128 νευρώνων στο LSTM επίπεδο, παρατηρείται υπο-εκπαίδευση. Ενώ η περίπτωση των 256, εκαπιδεύεται πολύ γρήγορα, σε επίπεδο όπως προηγούμενα μοντέλα.