Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Υπολογιστική Νοημοσύνη

8ο Εξάμηνο

Multilayer Perceptron

Σφυράκης Εμμανουήλ
AEM:9507
sfyrakise@ece.auth.gr

17 Ιανουαρίου 2021

Περιεχόμενα

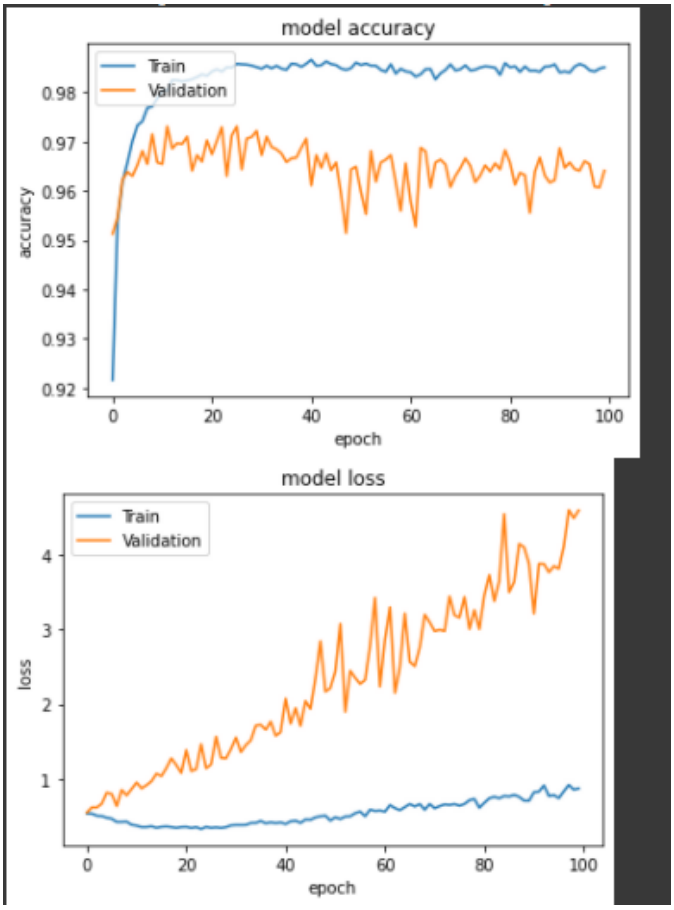
1 Διερεύνηση απόδοσης μοντέλου

Στην παρούσα εργασία πραγματοποιείται η εκπαίδευση ενός πολυστρωματικού Perceptron με χρήση πληθώρας διαφορετικών παραμέτρων batch_size, optimizer, L1 & L2 norm κ.α. Όσον αφορά την εκπαίδευση του δικτύου, εξάγουμε τα παρακάτω συμπεράσματα:

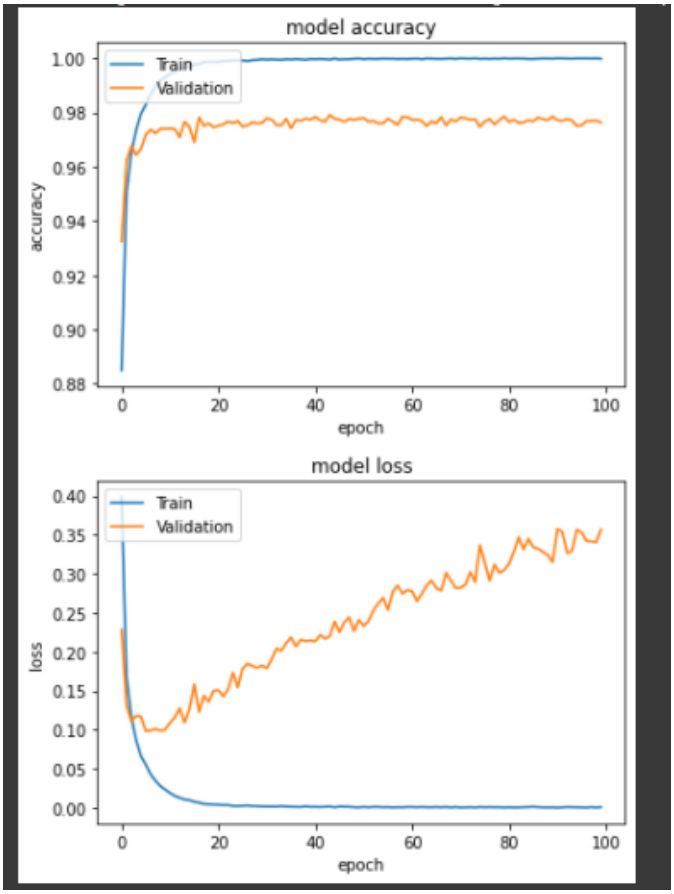
1.1 Batch_Size

Όσον αφορά την εκπαίδευση ενός default δικτύου με διαφορετικά μεγέθη batch , παρατηρούμε ότι για τον χρόνο εκπαίδευσης ισχύει: $t_{bs=1} > t_{bs=256} > t_{bs=60000}$. Η σχέση αυτή μεταξύ των χρόνων εκπαίδευσης είναι λογική καθώς όσο μικρότερο είναι το batch size, τόσες περισσότερες φορές θα υπάρξει update στις παραμέτρους του δικτύου άρα θα χρειαστεί και περισότερος χρόνος εκπαίδευσης. Στα σχήματα 1.1, 1.2 και 1.3 βλέπουμε την ακρίβεια του μοντέλου στα δεδομένα εκπαίδευσης και επικύρωσης, για batch_size=1, 256, 60000 αντίστοιχα.

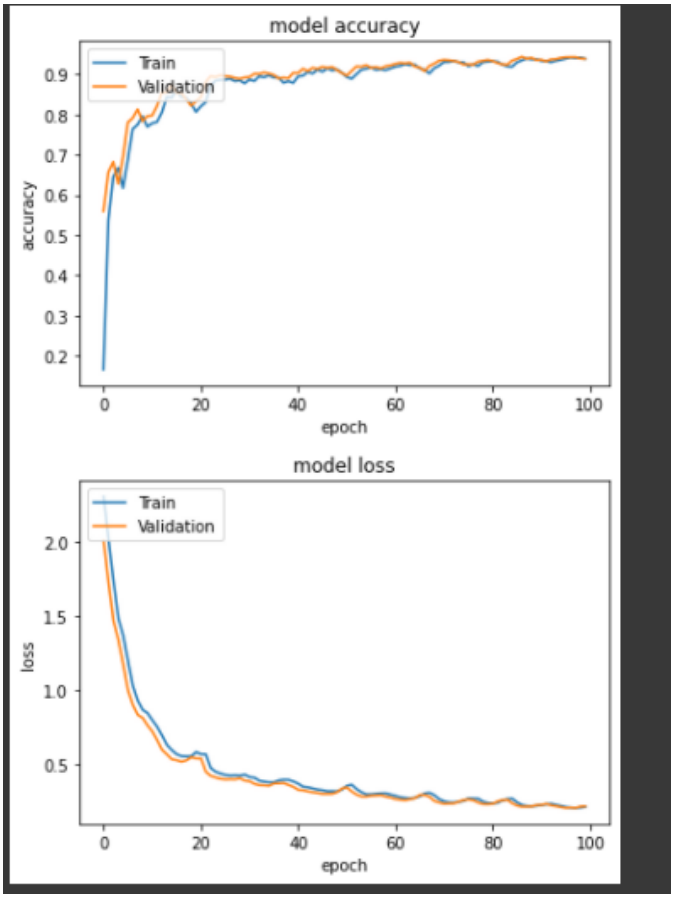
1.1.1 Καμπύλες ακριβείας και κόστους



Σχήμα 1.1: Καμπύλες ακριβείας και κόστους για batch_size=1



Σχήμα 1.2: Καμπύλες ακριβείας και κόστους για batch_size=256



Σχήμα 1.3: Καμπύλες ακριβείας και κόστους για batch_size=60000

1.1.2 Σχολιασμός αποτελεσμάτων

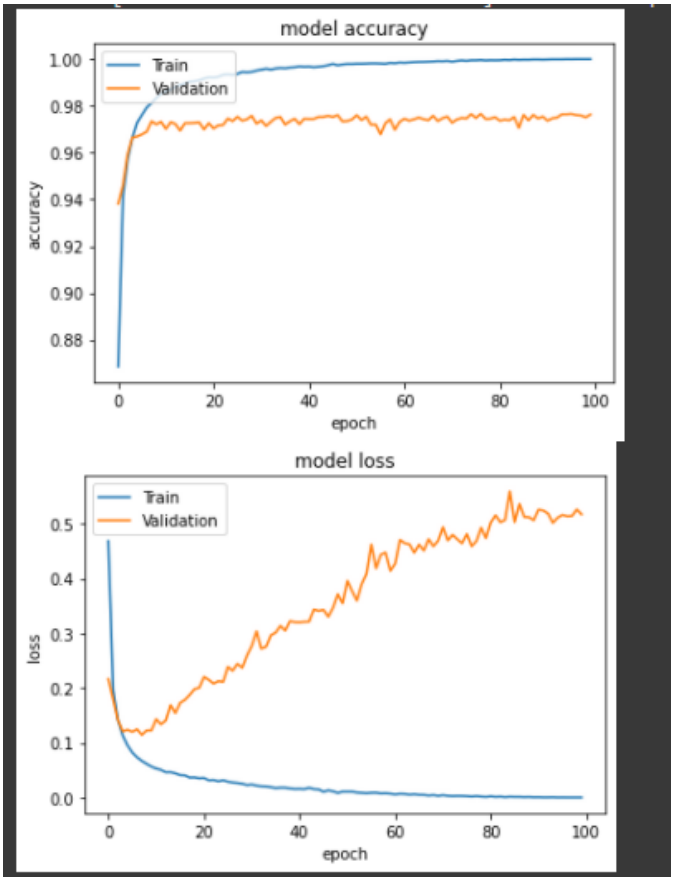
Παρατηρούμε ότι για **batch_size=1** το μοντέλο μας έχει αρκετά υψηλό loss στα δεδομένα επικύρωσης, κάτι που ωστόσο δεν ισχύει για τα δεδομένα εκπαίδευσης(φαινόμενο **overfitting**).

Για **batch_size=256** και **batch_size=60000** δεν αντιμετωπίζουμε τέτοιο πρόβλημα και σε γενικές γραμμές τα ποσοστά ακριβείας και κόστους είναι ικανοποιητικά τόσο για το σετ εκπαίδευσης όσο και για το σετ επικύρωσης.

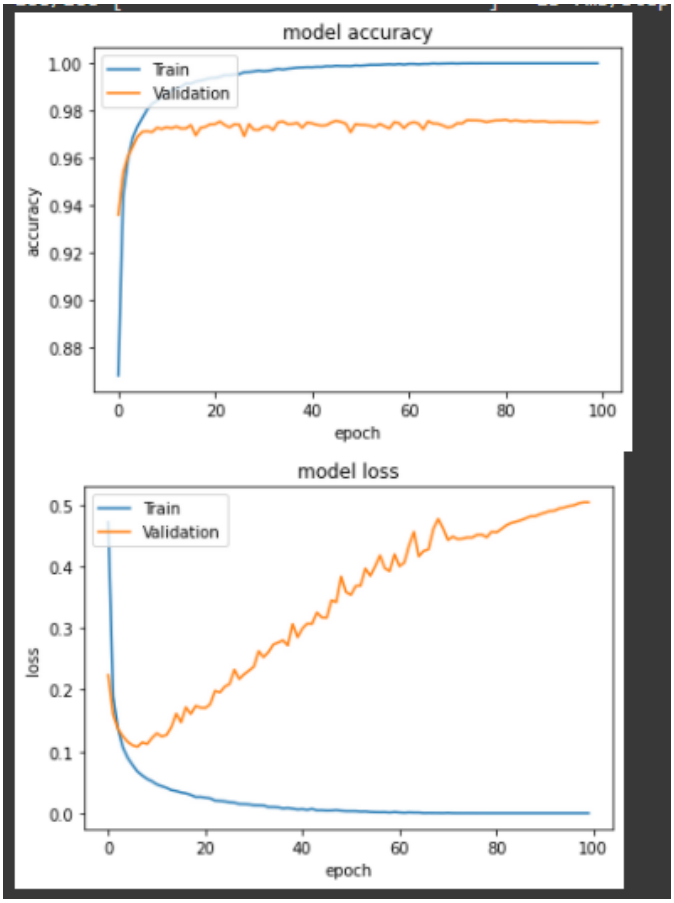
1.2 RMSProp

Η δοκιμή διαφορετικών τιμών rho για τον ήδη υπάρχων βελτιστοποιητή (ο RMSProp είναι ο default optimizer) και $lr = 0.001$ δίνει τις καμπύλες ακριβείας και κόστους των σχημάτων **??**, 1.5, 1.6, 1.7, 1.8.

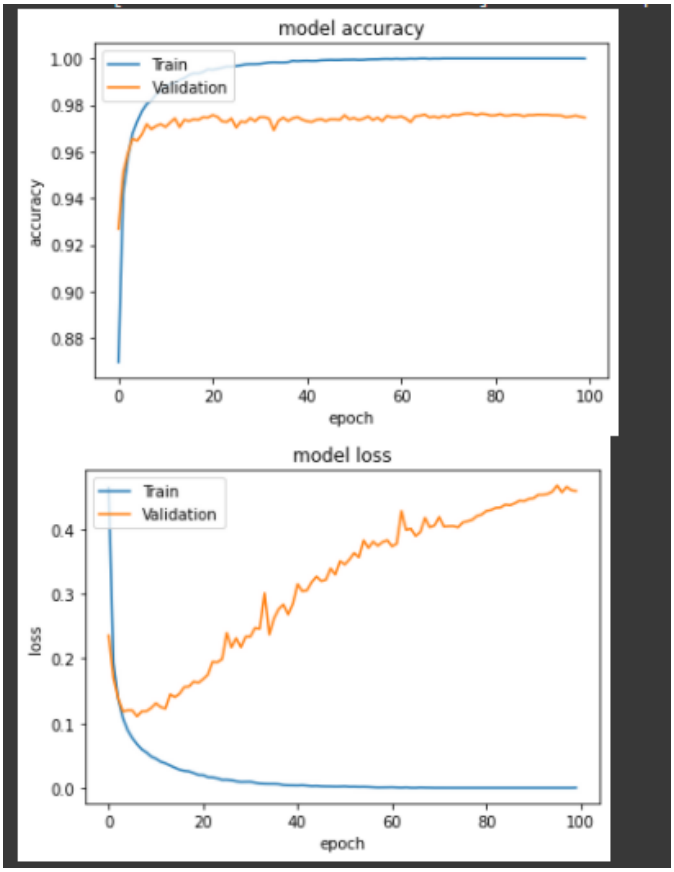
1.2.1 Καμπύλες ακριβείας και κόστους



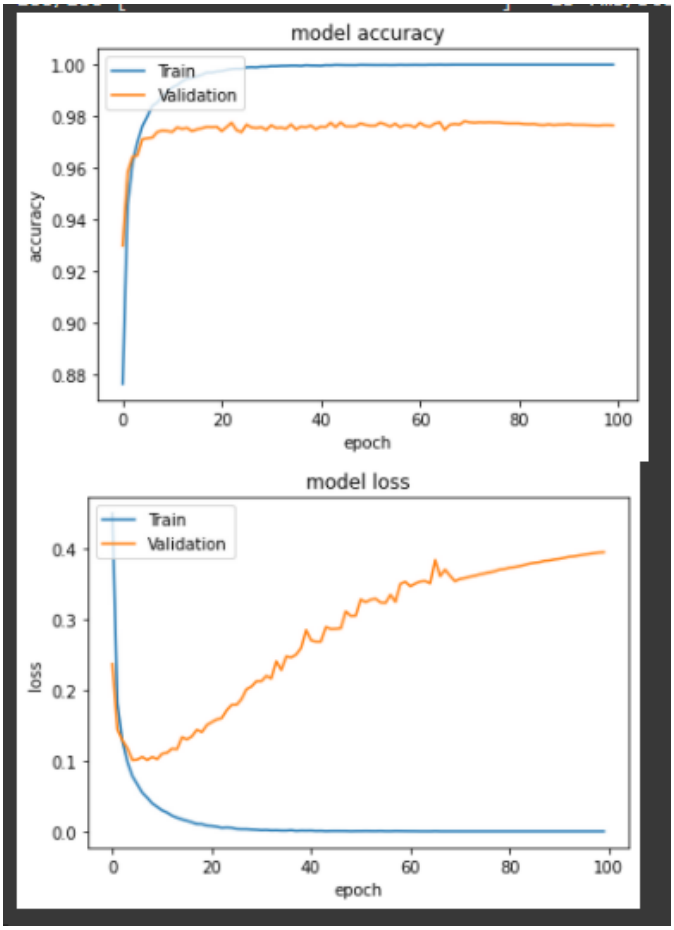
Σχήμα 1.4: Καμπύλες ακριβείας και κόστους για rho=0.01



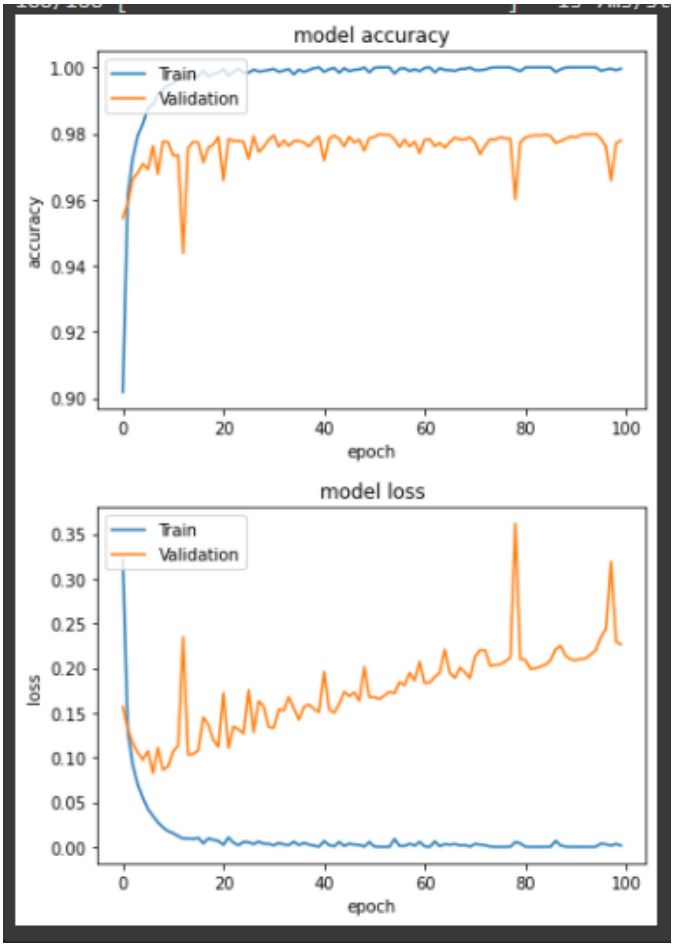
Σχήμα 1.5: Καμπύλες ακριβείας και κόστους για rho=0.05



Σχήμα 1.6: Καμπύλες ακριβείας και κόστους για rho=0.1



Σχήμα 1.7: Καμπύλες ακριβείας και κόστους για rho=0.5



Σχήμα 1.8: Καμπύλες ακριβείας και κόστους για rho=0.99

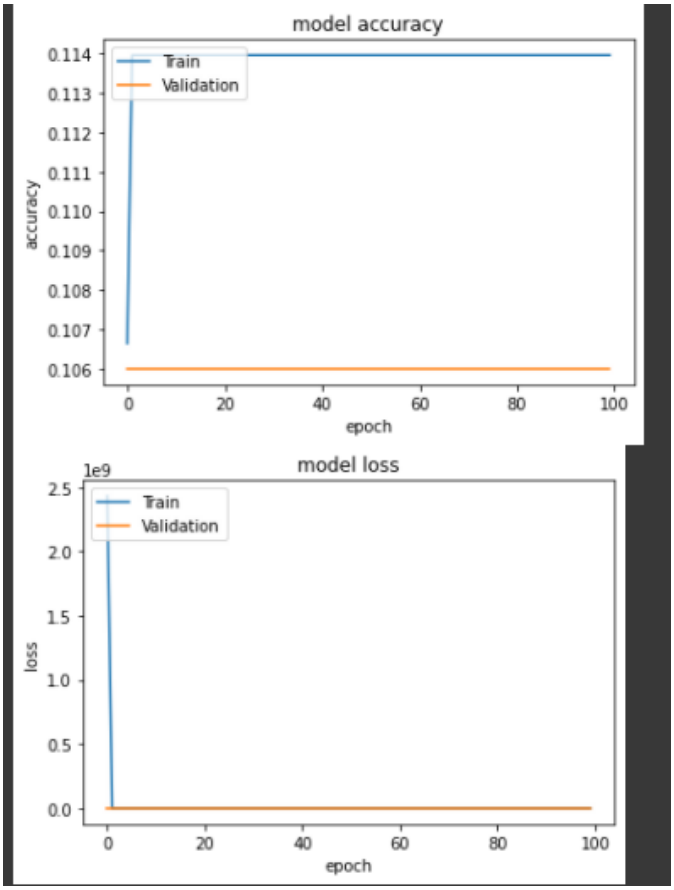
1.2.2 Σχολιασμός αποτελεσμάτων

Για όλες τις τιμές που δοκιμάστηκαν για το rho, πλην του 0.99, το μοντέλο μας αδυνατεί να γενικέυσει επιτυχώς σε άγνωστα δεδομένα. Αυτό φαίνεται από την μεγάλη απόκλιση μεταξύ training και validation loss(**overfitting**). Για rho=0.99, το μοντέλο παρουσιάζει έντονα το φαινόμενο της ταλάντωσης, γεγονός που μπορεί να οφείλεται σε διάφορους τιμές όπως αυτή του batch_size, του learning rate κ.λπ.

1.3 SGD Optimizer & Kernel Initializer

Με χρήση του βελτιστοποιητή stochastic gradient descent και ταυτόχρονη αρχιμοποίηση των βαρών κάθε στρώματος με βάση κανονική κατανομή με μ.ο. 10 λαμβάνουμε τις καμπύλες των σχημάτων 1.9

1.3.1 Καμπύλες ακριβείας και κόστους



Σχήμα 1.9: Καμπύλες ακριβείας και κόστους για SGD(0.01) και RandomNormal(mean=10)

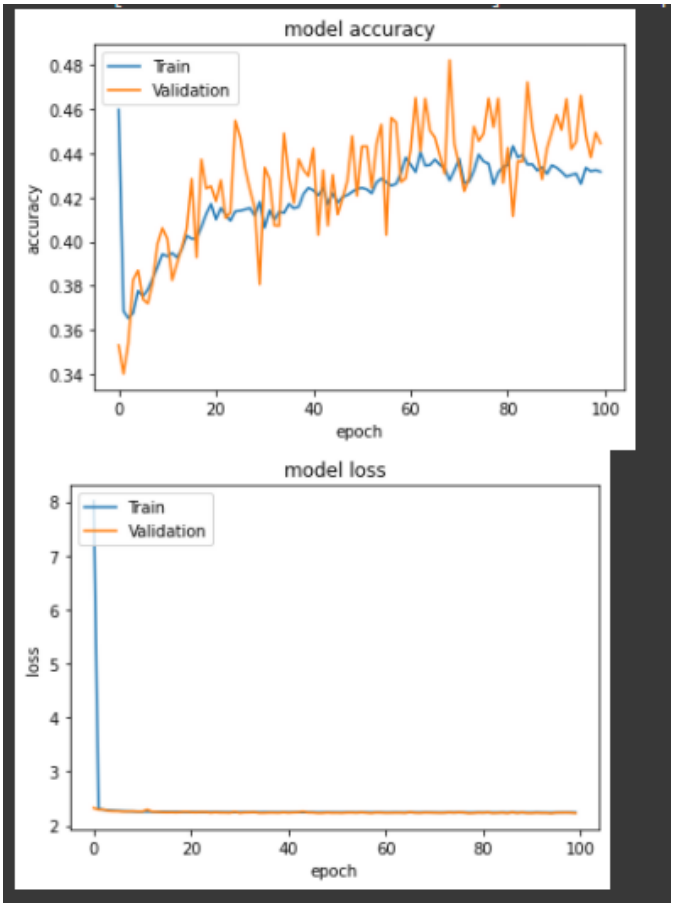
1.3.2 Σχολιασμός αποτελεσμάτων

Η εκπαίδευση του μοντέλου μας έχει αποτύχει καθώς έχουμε αρχικοποιήσει τα συναπτικά βάρη σε πολύ υψηλές τιμές (μέσος όρος=10). Ιδανικά, τα συναπτικά βάρη θα έπρεπε να έχουν αρχικοποιηθεί τυχαία σε κάποια περιοχή κοντά στο 0 ώστε να επιτρέπουν στον βελτιστοποιητή να κάνει τις κατάλληλες αλλαγές σε κάθε επανάληψη. Στην περίπτωση μας οι μεγάλες τιμές των βαρών οδηγούν σε τιμές της softmax κοντά στην μονάδα, με αποτέλεσμα η κλίση της παραγώγου να μεταβάλλεται πολύ αργά και έτσι να εμποδίζεται η διαδικασία εκμάθησης(**underfitting**).

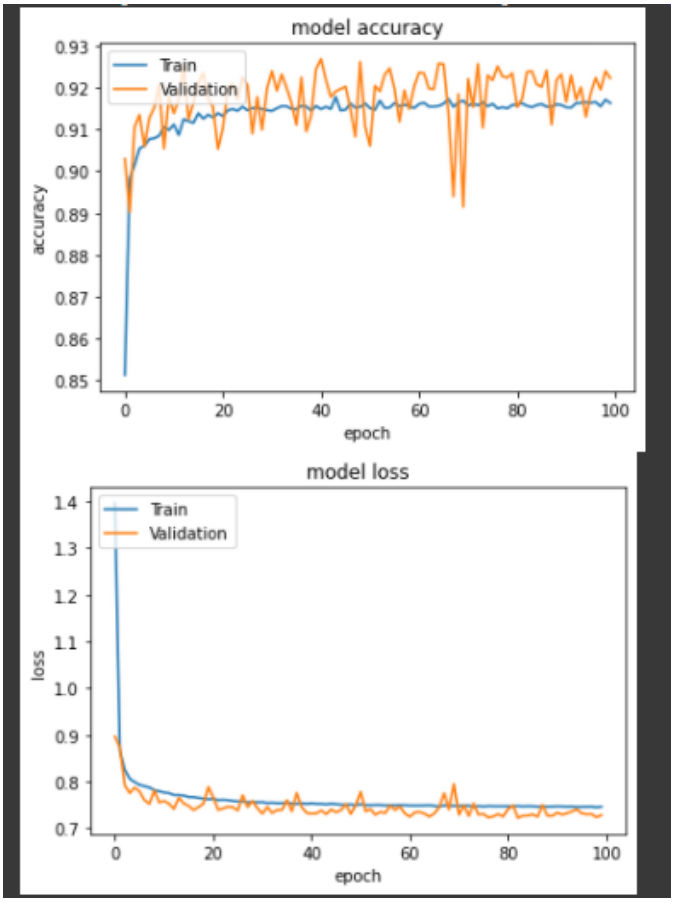
1.4 L_2 Normalization

Εν συνεχεία, κανονικοποιούμε τα βάρη με χρήση της L_2 νόρμας. Για διαφορετικές τιμές της παραμέτρου κανονικοποίησης α λαμβάνουμε τις καμπύλες 1.10, 1.11, 1.12, 1.13, 1.14, 1.15.

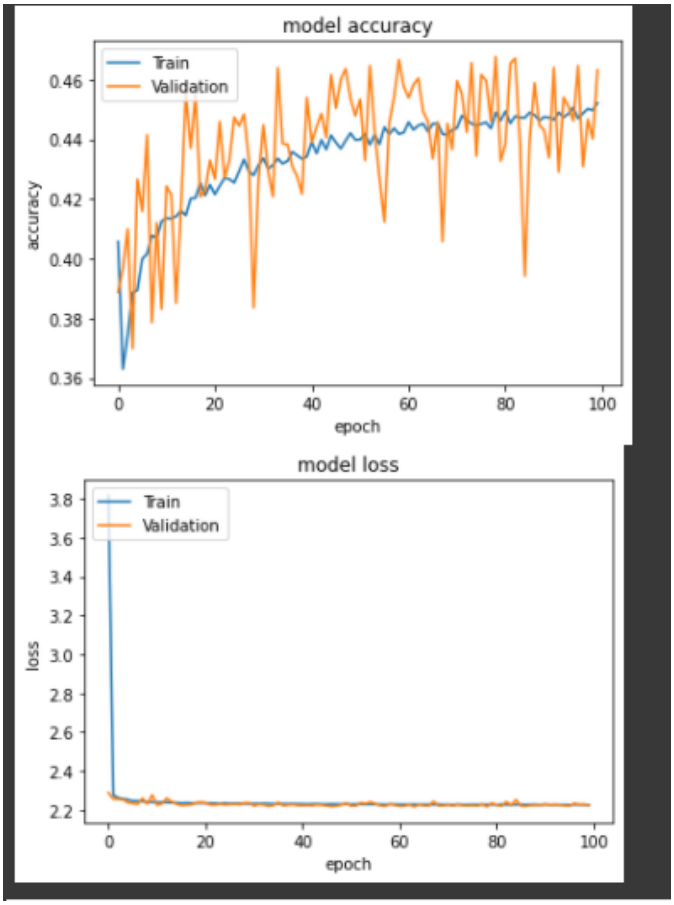
1.4.1 Καμπύλες ακριβείας και κόστους



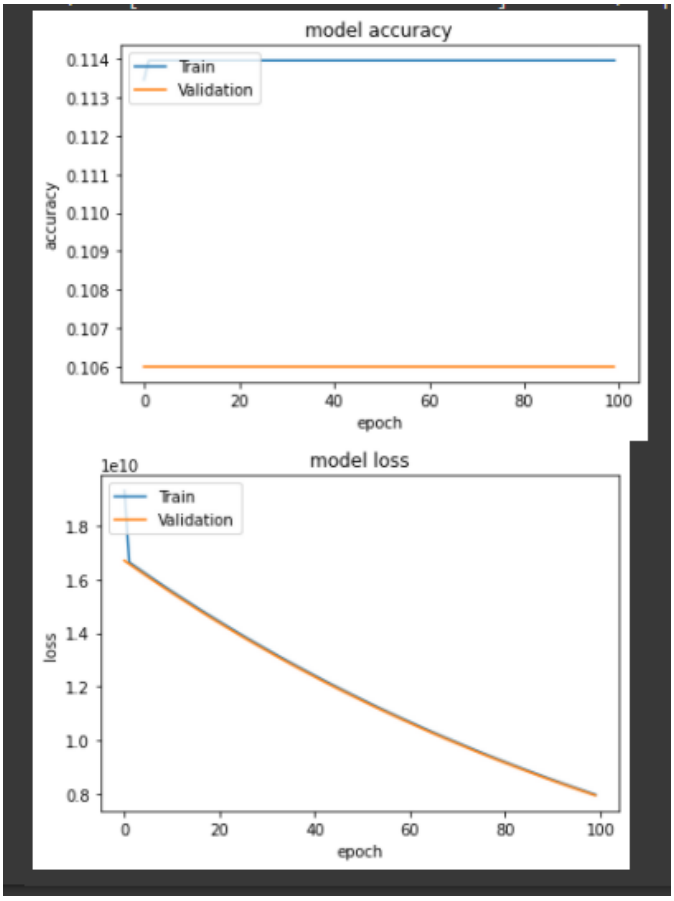
Σχήμα 1.10: Καμπύλες ακριβείας και κόστους για RMSProp optimizer(ρ=0.01) και α=0.1



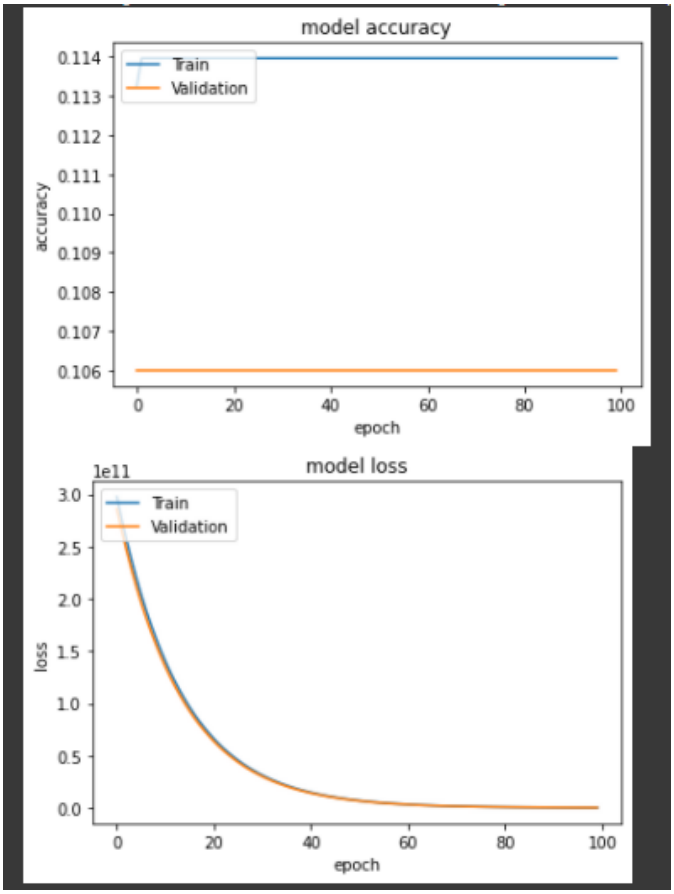
Σχήμα 1.11: Καμπύλες ακριβείας και κόστους για RMSProp optimizer(ρ=0.99) και α=0.01



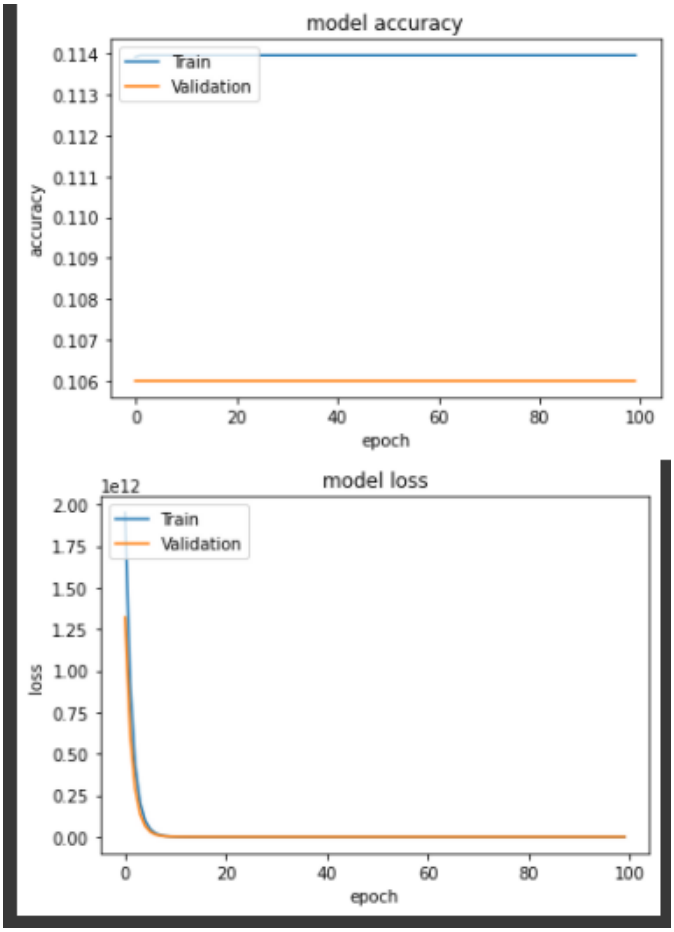
Σχήμα 1.12: Καμπύλες ακριβείας και κόστους για RMSProp optimizer(ρ=0.99) και α=0.1



Σχήμα 1.13: Καμπύλες ακριβείας και κόστους SGD optimizer και α=0.001



Σχήμα 1.14: Καμπύλες ακριβείας και κόστους SGD optimizer και α=0.01



Σχήμα 1.15: Καμπύλες ακριβείας και κόστους SGD optimizer και α=0.1

1.4.2 Σχολιασμός Αποτελεσμάτων

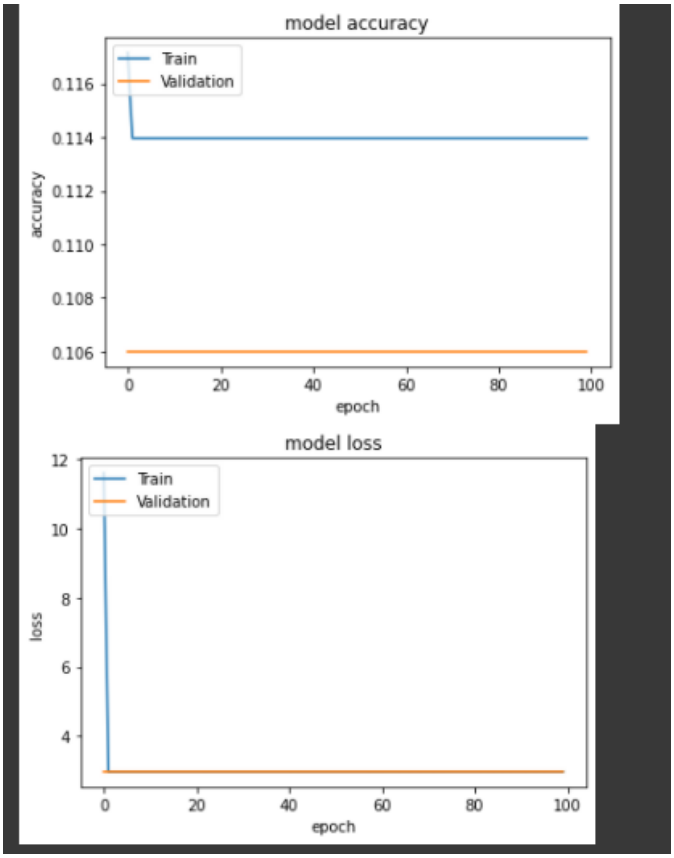
Στην περίπτωση ταυτόχρονης χρήσης L_2 νόρμας με βελτιστοποιητή RMSProp και $\alpha = 0.1$ δεν έχουμε την επιθυμητή εκπαίδευση καθώς οι τιμές της ακρίβειας δεν ξεπερνάνε το 0.5 για οποιαδήποτε τιμή του rho(**underfitting**). Με την μείωση του α οι τιμές της ακρίβειας και του κόστους βελτιώνονται αισθητά και για τα δύο σύνολα μας($acc > 0.9$ και $loss \leq 0.8$), ενώ μειώνεται και το μέγεθος των ταλαντώσεων.

Η αλλαγή βελτιστοποιητή (χρήση Stochastic Gradient Descent) έχει ακόμα πιο ανεπιθύμητα αποτελέσματα καθώς η ακρίβεια και των δύο συνόλων κυμαίνεται στο 0.1(**underfitting**).

1.5 L_1 Normalization

Χρησιμοποιώντας το default δίκτυο με batch_size=256 και εισάγωντας κανονικοποίηση με L_1 νόρμα για τα συναπτικά βάρη(με $\alpha = 0.01$) και ταυτόχρονη χρήση dropout με dropout propability 0.3 λαμβάνουμε τις καμπύλες του σχ. 1.16.

1.5.1 Καμπύλες ακριβείας και κόστους



Σχήμα 1.16: Καμπύλες ακριβείας και κόστους default δικτύου με L_1 norm dropout

1.5.2 Σχολιασμός αποτελεσμάτων

Η εκπαίδευση του μοντέλου δεν κρίνεται επαρκής καθώς η ακρίβεια κυμαίνεται περίπου στο 10%. Το συγκεκριμένο αποτέλεσμα μπορεί να οφείλεται είτε στην μεγάλη τιμή του α το οποίο υπερ απλούστευσε το μοντέλο, είτε σε λανθασμένη τιμή του rate στο dropout(**underfitting**).

2 Fine tuning δικτύου

2.1 Παράμετροι

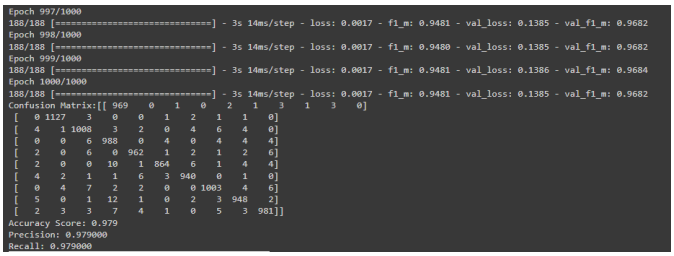
Όσον αφορά την διαδικασία εύρεσης των βέλτιστων τιμών των παραμέτρων του ζητούμενου δικτύου, χρησιμοποιήθηκε ο keras-tuner. Βάσει αυτού, προέκυψε ότι οι βέλτιστες τιμές των παραμέτρων είναι οι εξής:

- Αριθμός νευρώνων στο πρώτο κρυφό στρώμα: 128
- Αριθμός νευρώνων στο δεύτερο κρυφό στρώμα: 512

- Παράμετρος κανονικοποίησης α : 0.1
- Ρυθμός εκμάθησης: 10^{-6}

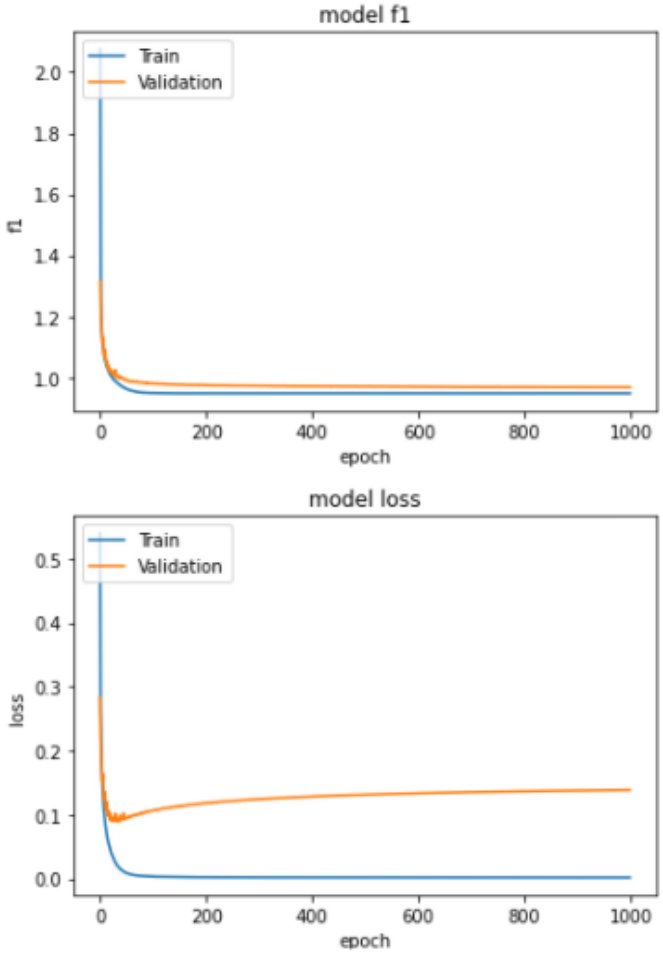
2.2 Εκπαίδευση μοντέλου

Με βάση τις τιμές αυτές και τα δεδομένα της εργασίας, γίνεται στη συνέχεια η εκπαίδευση του δικτύου. Οι καμπύλες εκμάθησης φαίνονται στο σχ. 2.2 ενώ ο πίνακας σύγχυσης και οι ζητούμενες μετρικές δίνονται στο σχ. 2.1.



Σχήμα 2.1: Confusion Matrix, Accuracy, Recall, Precision, f1

2.2.1 Καμπύλες ακριβείας και κόστους



Σχήμα 2.2: Καμπύλες εκμάθησης υπερμοντέλου

2.2.2 Σχολιασμός αποτελεσμάτων

Εφόσον η μετρική f1, όπως φαίνεται και από το πρώτο διάγραμμα του σχ. 2.2, βρίσκεται αρκετά κοντά στην μονάδα και το loss είναι αρκετά μικρό, η εκπαίδευση κρίνεται αποτελεσματική, πράγμα αναμενόμενο και μετά το fine tuning του δικτύου.