

Υπολογιστική Νοημοσύνη
-Τρίτη Εργασία-

Αλέξανδρος Πετρίδης

Τελευταία ενημέρωση: 7 Μαρτίου 2022

Περιεχόμενα

1	Στόχος	3
2	Διερεύνηση απόδοσης μοντέλων	3
2.1	Αποτελέσματα και σχολιασμός	3
2.1.1	Ταχύτητα εκπαίδευσης	3
2.1.2	Καμπύλες ακρίβειας και κόστους	3
2.2	Σχολιασμός αποτελεσμάτων	4
3	Fine tuning δικτύου	4
4	Σχήματα	5

Κατάλογος Σχημάτων

1	Πίνακας σύγχυσης	4
2	Μετρικές	4
3	Καμπύλες εκμάθησης	5
4	Καμπύλες ακρίβειας και κόστους για τα πρώτα τρία μοντέλα	5
5	Καμπύλες ακρίβειας και κόστους για μοντέλα με RMSProp optimizer	5
6	Καμπύλες ακρίβειας και κόστους για μοντέλο με SGD optimizer	6
7	Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.1$ για διαφορετικά minibatches	6
8	Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.01$ για διαφορετικά minibatches	6
9	Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.001$ για διαφορετικά minibatches	7
10	Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.1$ και RMSProp optimizer	7
11	Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.01$ και RMSProp optimizer	8
12	Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.001$ και RMSProp optimizer	8
13	Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα και SGD optimizer	9
14	Καμπύλες ακρίβειας και κόστους για μοντέλο με κανονικοποίηση με L_1 -νόρμα	9

1 Στόχος

Στόχος είναι η σωστή ταξινόμηση κάθε εικόνας στην κλάση που αντιστοιχεί στο ψηφίο που απεικονίζει. Το Dataset που επιλέχθηκε είναι το MNIST Dataset, το οποίο περιλαμβάνει εικόνες χειρόγραφων ψηφίων από 28×28 pixels η κάθε μια. Τα δεδομένα είναι χωρισμένα σε training και testing υποσύνολα, το καθένα με 60.000 και 10.000 δείγματα αντίστοιχα.

2 Διερεύνηση απόδοσης μοντέλων

Το πρώτο κομμάτι της εργασίας έχει να κάνει με τη μελέτη της απόδοσης του δικτύου για διαφορετικές επιλογές όσον αφορά την επιλογή μεθόδου βελτιστοποίησης, κανονικοποίησης και αρχικοποίησης των παραμέτρων. Συγκεκριμένα, εξετάζεται η συμπεριφορά του δικτύου στους παρακάτω συνδυασμούς:

- Εκπαίδευση ενός default δικτύου με τα ακόλουθα minibatch μεγέθη: 1, 256, N_{train} - «online, minibatch, batch», όπου N_{train} το συνολικό πλήθος των δεδομένων εκπαίδευσης.
- RMSProp optimizer με $l_r = 0.001$, $r \in 0.01, 0.99$.
- SGD optimizer με $l_r = 0.01$ και αρχικοποίηση των συναπτικών βαρών W κάθε στρώματος με βάση μια κανονική κατανομή με μ.ο. 10.
- Ίδιες επιλογές με παραπάνω, με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα για τα συναπτικά βάρη κάθε στρώματος με παράμετρο κανονικοποίησης $a \in 0.1, 0.01, 0.001$.
- Κανονικοποίηση με L_1 -νόρμα για τα συναπτικά βάρη των στρωμάτων του δικτύου ($\alpha=0.01$) και ταυτόχρονη χρήση dropout με dropout probability 0.3.

Σε κάθε μία από τις παραπάνω περιπτώσεις, το δίκτυο περιλαμβάνει δύο κρυφά στρώματα με 128 και 256 νευρώνες αντίστοιχα, και συνάρτηση ενεργοποίησης την ReLU. Όπου δεν αναφέρεται ρητά, η ενεργοποίηση του στρώματος εξόδου επιλέγεται η softmax. Τέλος, ως μετρική αξιολόγησης θεωρείται η ακρίβεια accuracy, ενώ η αντικειμενική συνάρτηση προς βελτιστοποίηση επιλέγεται ως η categorical cross-entropy, η οποία για κάθε δείγμα i ορίζεται το άθροισμα του κόστους για κάθε κλάση ξεχωριστά:

$$J(i) = - \sum_{c=1}^{10} y_{i,c} \cdot \log_e \cdot \hat{y}_{i,c}$$

Η εκπαίδευση των δικτύων γίνεται για 100 εποχές, με ένα 20% του συνόλου των δεδομένων εκπαίδευσης να παρακρατείται για χρήση ως σύνολο επικύρωσης.

2.1 Αποτελέσματα και σχολιασμός

2.1.1 Ταχύτητα εκπαίδευσης

Για τα τρία μοντέλα με τα ακόλουθα minibatch μεγέθη: 1, 256, N_{train} , έχουμε:

- Training time with minibatch = 1: 12316.667080402374 seconds
- Training time with minibatch = 256: 131.43236541748047 seconds
- Training time with minibatch = 60000: 49.831623554229736 seconds

Παρατηρούμε πως η ταχύτητα εκπαίδευσης βελτιώνεται ανάλογα με το πόσο μεγάλο σε μέγεθος είναι το minibatch κάτι που ήταν αναμενόμενο, καθώς το minibatch είναι το μέγεθος των δεδομένων που θα χρησιμοποιηθούν για να υπολογιστεί το error του μοντέλου και να ενημερωθούν οι συντελεστές.

2.1.2 Καμπύλες ακρίβειας και κόστους

Στο σχήμα 4 φαίνονται οι καμπύλες ακρίβειας και κόστους για τα πρώτα τρία μοντέλα με τα διαφορετικά minibatches. Ακολούθως στο σχήμα 5 απεικονίζονται οι καμπύλες ακρίβειας και κόστους για τα δύο μοντέλα με RMSProp optimizer με learning_rate = 0.001 και $r \in 0.01, 0.99$. Ακόμα στο σχήμα 6 οι καμπύλες ακρίβειας και κόστους για μοντέλο με SGD optimizer με learning_rate = 0.01 και αρχικοποίηση των συναπτικών βαρών W κάθε στρώματος με βάση μια κανονική κατανομή με μ.ο. 10. Επίσης στα σχήματα 7, 8, 9, 10, 11, 12 και 13 δίνονται οι καμπύλες ακρίβειας και κόστους για μοντέλα με ίδιες επιλογές με τα παραπάνω μοντέλα και επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα για τα συναπτικά βάρη κάθε στρώματος με παράμετρο κανονικοποίησης $a \in 0.1, 0.01, 0.001$. Τέλος στο σχήμα 14 δίνονται οι καμπύλες ακρίβειας και κόστους για μοντέλο με κανονικοποίηση με L_1 -νόρμα για τα συναπτικά βάρη των στρωμάτων του δικτύου ($\alpha = 0.01$) και ταυτόχρονη χρήση dropout με dropout probability = 0.3.

2.2 Σχολιασμός αποτελεσμάτων

Παρατηρώντας τα αποτελέσματα μέσω των σχημάτων, παρατηρείται πως τα μοντέλα με RMSProp optimizer είναι καλύτερα σε σχέση με τον SGD optimizer βασιζόμενοι είτε στην ακρίβεια είτε στην απώλεια. Από τα μοντέλα που κάνουν χρήση του L_2 regularizer ξεχωρίζουν και πάλι αυτά που κάνουν χρήση του RMSProp optimizer με $\alpha=0.01$ και $\rho=0.99$.

3 Fine tuning δικτύου

Στο συγκεκριμένο κομμάτι της εργασίας, σκοπός είναι η εύρεση των βέλτιστων τιμών για μερικές υπερπαραμέτρους του δικτύου, και η τελική εκπαίδευση και αξιολόγηση ενός μοντέλου με βάση της επιλεγμένες τιμές.

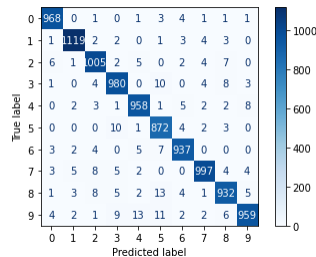
Το δίκτυο εκπαιδεύτηκε με τον αλγόριθμο βελτιστοποίησης RMSProp . Σε κάθε στρώμα του δικτύου εφαρμόστηκε κανονικοποίηση με βάση την L_2 -νόρμα και η αρχικοποίηση των συναπτικών βαρών κάθε στρώματος έγινε με βάση το He-normalization . Το δίκτυο αποτελείται από δύο κρυφά στρώματα. Οι υπερπαραμέτροι του δικτύου προς εξέταση, καθώς και το εύρος αναζήτησης για κάθε παράμετρο, είναι:

1. αριθμός νευρώνων πρώτου κρυφού στρώματος $n_{h1} \in 64, 128$
2. αριθμός νευρώνων δεύτερου κρυφού στρώματος $n_{h2} \in 256, 512$
3. παράμετρος κανονικοποίησης $a \in 0.1, 0.001, 0.000001$
4. ρυθμός εκμάθησης $l_r \in 0.1, 0.001, 0.001$

Η διαδικασία εύρεσης των βέλτιστων τιμών των παραμέτρων έγινε με τη χρήση του keras-tuner . Για την επιτάχυνση της διαδικασίας, χρησιμοποιήθηκε η μέθοδος του Early Stopping , χωρίς το βήμα της μετέπειτα επανεκπαίδευσης του δικτύου, με $\text{patience} = 200$ για κάθε ένα από τα εξεταζόμενα μοντέλα. Για τα κρυφά στρώματα και το στρώμα εξόδου, επιλέχθηκαν η ReLU και η softmax συναρτήσεις ενεργοποίησης αντίστοιχα. Ως μετρική αξιολόγησης χρησιμοποιήθηκε το F-measure . Ο αριθμός εποχών ορίστηκε στο 1000 και το validation split στο 20%.

Αφού βρέθηκαν οι βέλτιστες τιμές των υπερπαραμέτρων, εκπαιδεύτηκε το τελικό μοντέλο με βάση αυτές τις τιμές και τις υπόλοιπες επιλογές που θεωρήθηκαν παραπάνω. Η τελική αξιολόγηση έγινε στο testing μοντέλο. Υπολογίστηκαν:

1. Ο πίνακας σύγχυσης. Παρατηρούμε πως ο πίνακας έχει πολύ υψηλές τιμές στην διαγώνιο, εκεί δηλαδή όπου το προβλεπόμενο και το πραγματικό ψηφίο είναι ίδια. Αυτό σημαίνει πως το μοντέλο καταφέρνει να προβλέψει τα ψηφία που δίνονται σε αρκετά ικανοποιητικό βαθμό.



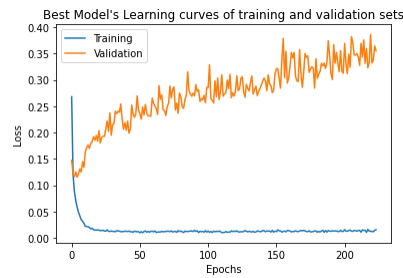
Σχήμα 1: Πίνακας σύγχυσης

2. Οι μετρικές accuracy, recall, precision, f-measure .

	precision	recall	f1-score	support
0	0.98	0.99	0.98	980
1	0.99	0.99	0.99	1135
2	0.97	0.97	0.97	1032
3	0.97	0.97	0.97	1010
4	0.97	0.98	0.97	982
5	0.95	0.98	0.96	892
6	0.98	0.98	0.98	958
7	0.98	0.97	0.98	1028
8	0.96	0.96	0.96	974
9	0.98	0.95	0.96	1009
accuracy			0.97	10000
macro avg	0.97	0.97	0.97	10000
weighted avg	0.97	0.97	0.97	10000

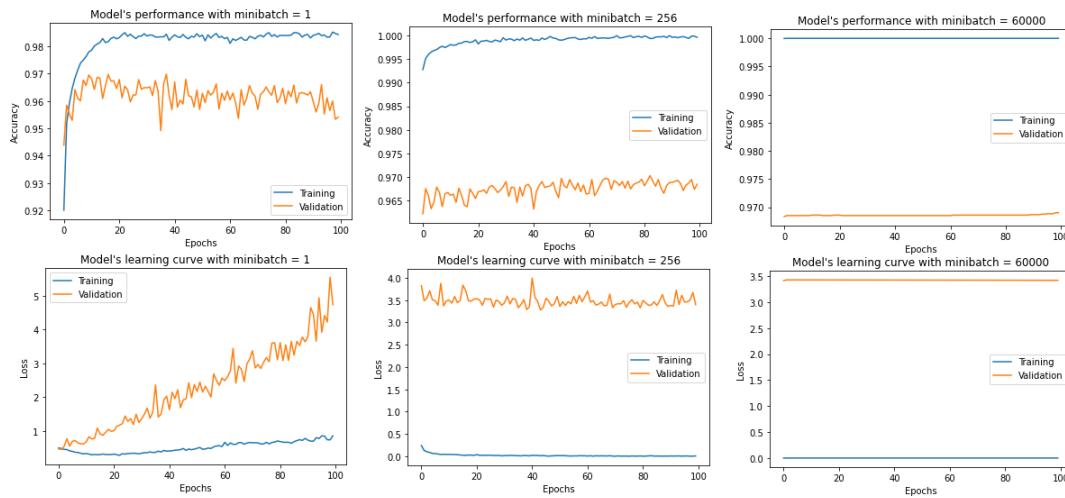
Σχήμα 2: Μετρικές

3. Καμπύλες εκμάθησης για training και validation δεδομένα.

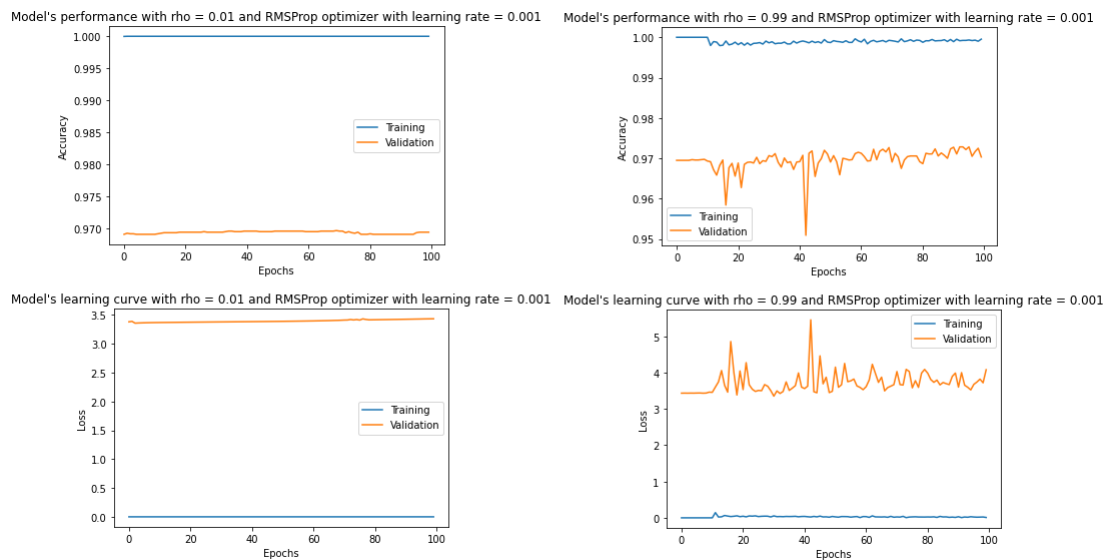


Σχήμα 3: Καμπύλες εκμάθησης

4 Σχήματα

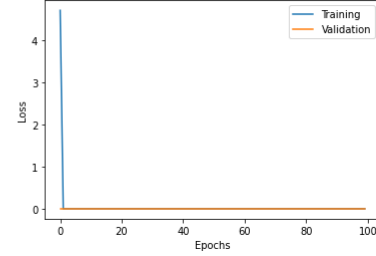
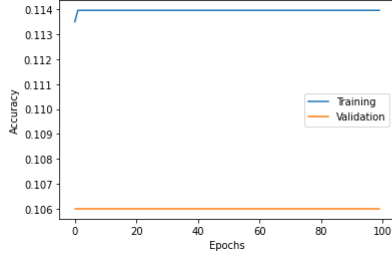


Σχήμα 4: Καμπύλες ακρίβειας και κόστους για τα πρώτα τρία μοντέλα



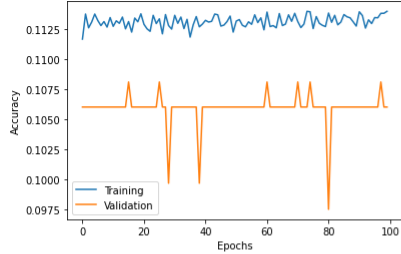
Σχήμα 5: Καμπύλες ακρίβειας και κόστους για μοντέλα με RMSProp optimizer

Model's performance with kernel initializer = (Random Normal with mean 10) and SGD optimizer with learning rate = 0.01 Model's learning curve with kernel initializer = (Random Normal with mean 10) and SGD optimizer with learning rate = 0.01

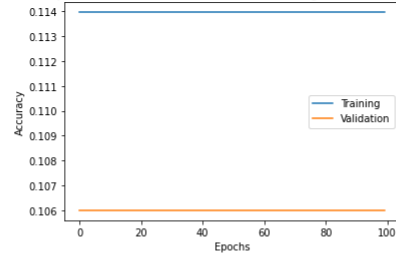


Σχήμα 6: Καμπύλες ακρίβειας και κόστους για μοντέλο με SGD optimizer

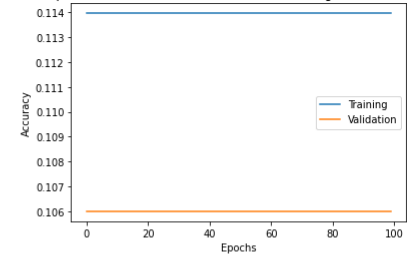
Model's performance with minibatch = 1 and regularizer L2 with $\alpha = 0.1$



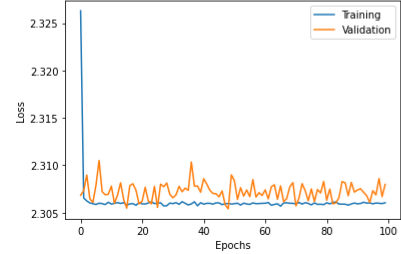
Model's performance with minibatch = 256 and regularizer L2 with $\alpha = 0.1$



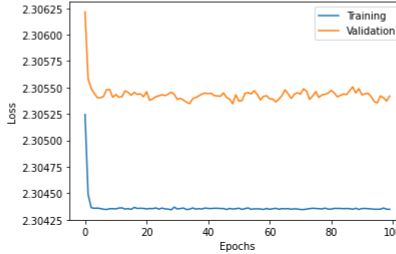
Model's performance with minibatch = 60000 and regularizer L2 with $\alpha = 0.1$



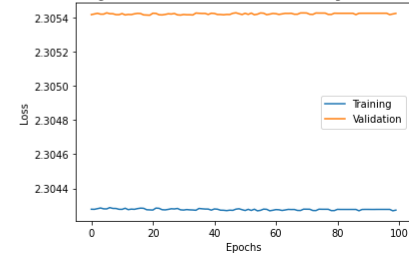
Model's learning curve with minibatch = 1 and regularizer L2 with $\alpha = 0.1$



Model's learning curve with minibatch = 256 and regularizer L2 with $\alpha = 0.1$

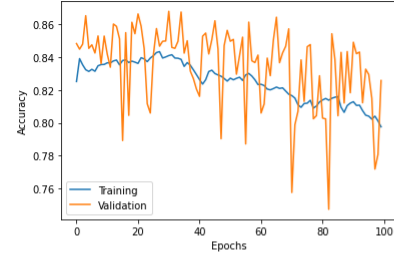


Model's learning curve with minibatch = 60000 and regularizer L2 with $\alpha = 0.1$

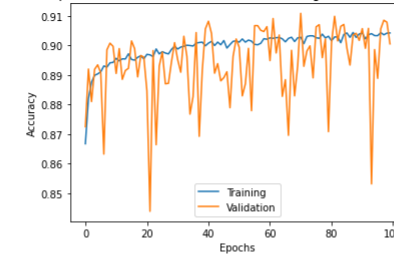


Σχήμα 7: Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.1$ για διαφορετικά minibatches

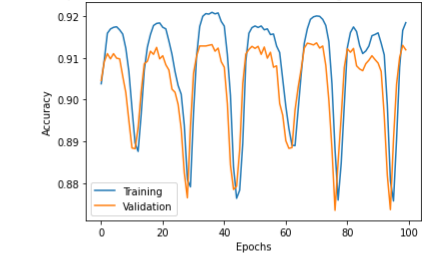
Model's performance with minibatch = 1 and regularizer L2 with $\alpha = 0.01$



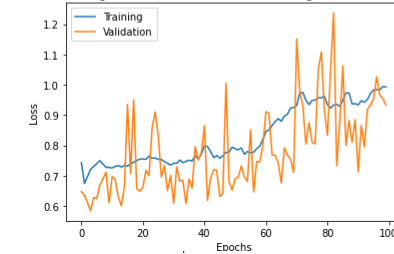
Model's performance with minibatch = 256 and regularizer L2 with $\alpha = 0.01$



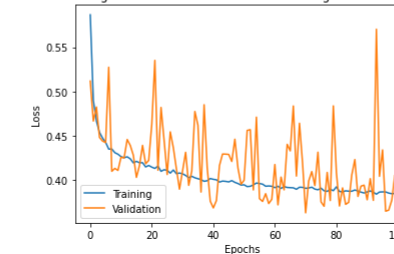
Model's performance with minibatch = 60000 and regularizer L2 with $\alpha = 0.01$



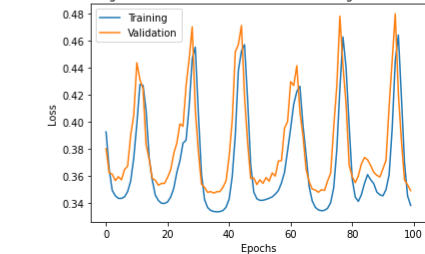
Model's learning curve with minibatch = 1 and regularizer L2 with $\alpha = 0.01$



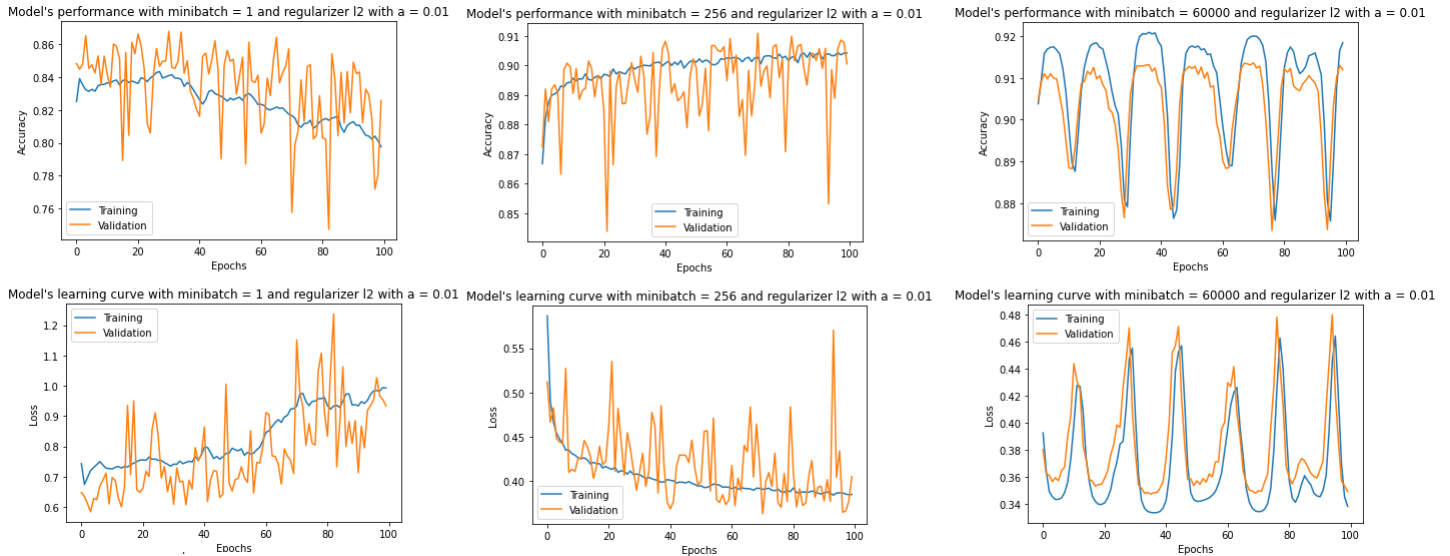
Model's learning curve with minibatch = 256 and regularizer L2 with $\alpha = 0.01$



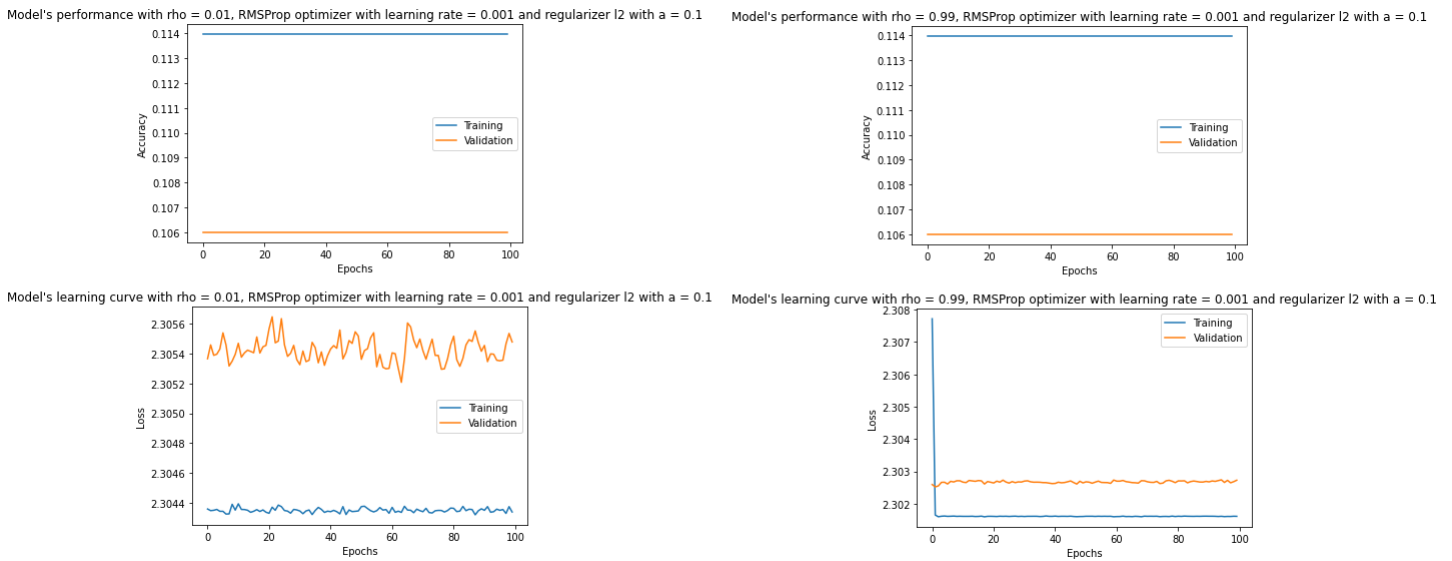
Model's learning curve with minibatch = 60000 and regularizer L2 with $\alpha = 0.01$



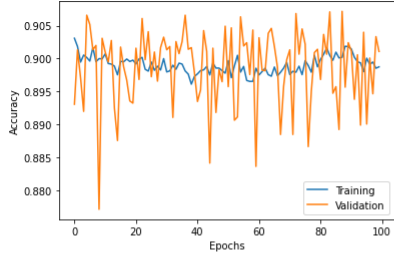
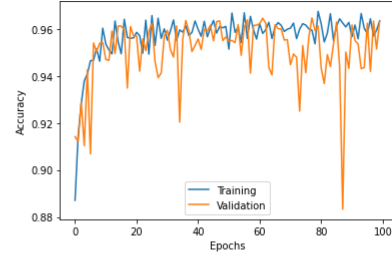
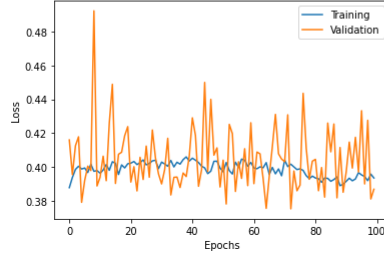
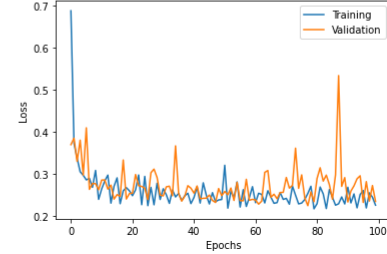
Σχήμα 8: Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.01$ για διαφορετικά minibatches



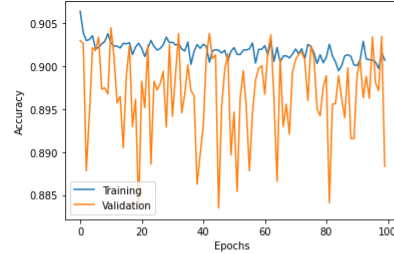
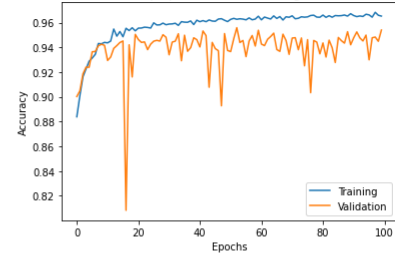
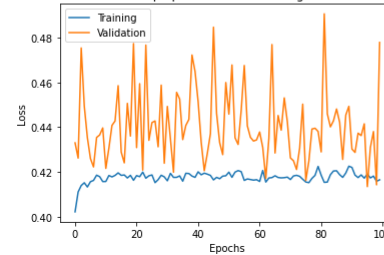
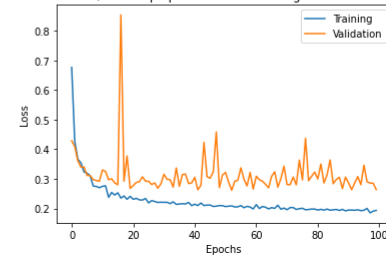
Σχήμα 9: Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.001$ για διαφορετικά minibatches



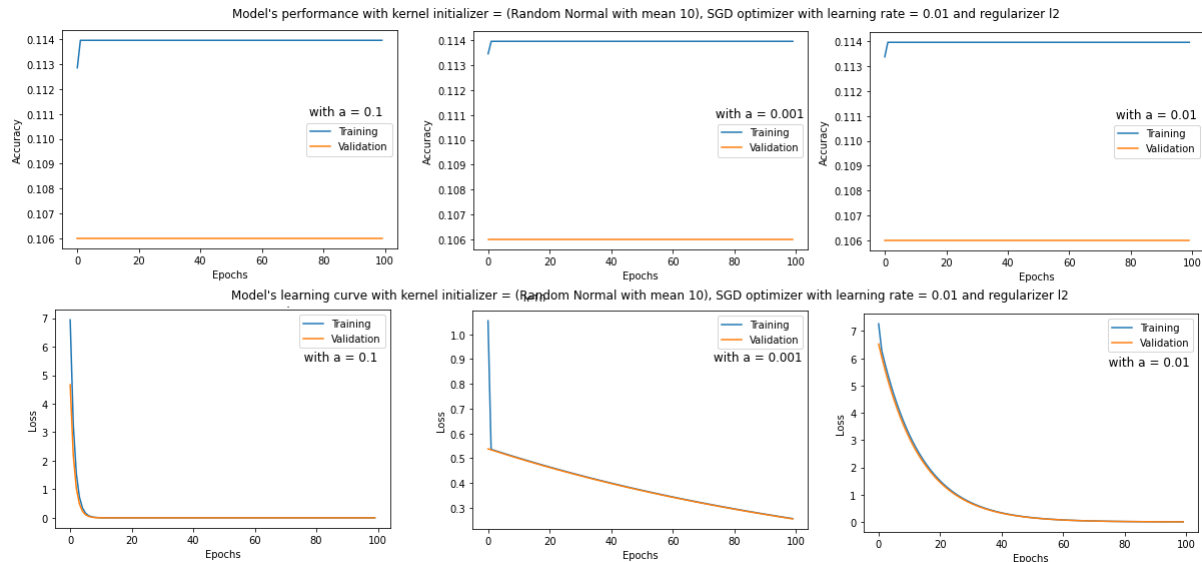
Σχήμα 10: Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.1$ και RMSProp optimizer

Model's performance with $\rho = 0.01$, RMSProp optimizer with learning rate = 0.001 and regularizer l_2 with $\alpha = 0.01$ Model's performance with $\rho = 0.99$, RMSProp optimizer with learning rate = 0.001 and regularizer l_2 with $\alpha = 0.01$ Model's learning curve with $\rho = 0.01$, RMSProp optimizer with learning rate = 0.001 and regularizer l_2 with $\alpha = 0.01$ Model's learning curve with $\rho = 0.99$, RMSProp optimizer with learning rate = 0.001 and regularizer l_2 with $\alpha = 0.01$ 

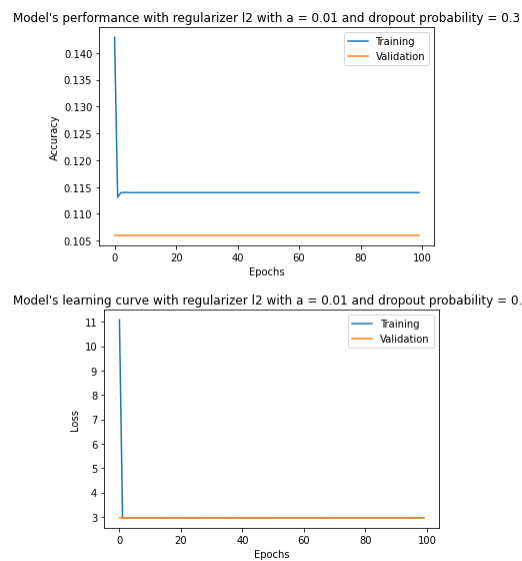
Σχήμα 11: Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.01$ και RMSProp optimizer

Model's performance with $\rho = 0.01$, RMSProp optimizer with learning rate = 0.001 and regularizer l_2 with $\alpha = 0.001$ Model's performance with $\rho = 0.99$, RMSProp optimizer with learning rate = 0.001 and regularizer l_2 with $\alpha = 0.001$ Model's learning curve with $\rho = 0.01$, RMSProp optimizer with learning rate = 0.001 and regularizer l_2 with $\alpha = 0.001$ Model's learning curve with $\rho = 0.99$, RMSProp optimizer with learning rate = 0.001 and regularizer l_2 with $\alpha = 0.001$ 

Σχήμα 12: Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα με $\alpha=0.001$ και RMSProp optimizer



Σχήμα 13: Καμπύλες ακρίβειας και κόστους με επιπλέον προσθήκη κανονικοποίησης με L_2 -νόρμα και SGD optimizer



Σχήμα 14: Καμπύλες ακρίβειας και κόστους για μοντέλο με κανονικοποίηση με L_1 -νόρμα