Στατιστική Μοντελοποίηση και Αναγνώριση Προτύπων

Σειρά Ασκήσεων 4

Αναστάσιος Πατερίτσας ΑΜ: 2016030065

Θέμα 1.

Στο συγκεκριμένο θέμα υλοποιούμε την επίδοση του συστήματος με την τεχνική cross validation leave-one-out. Ο σκοπός της άσκησης είναι να ελέγξουμε την συσχέτιση των χαρακτηριστικών με τα labels , δηλαδή ποια χαρακτηριστικά έχουν μεγαλύτερη βαρύτητα και μεγαλύτερη συσχέτιση με το αν ένας άνθρωπος έχει αυτισμό. Χρησιμοποιούμε τον συντελεστή Pearson στην διαδικασία του feature selection. Η cross validation είναι η επαναληπτική κλήση του classifier μας. Η διαφοροποίηση που προκύπτει κατά την διαδικασία του cross validation είναι σε ποιο κομμάτι θα τεθεί προς training και ποιο προς testing. Όποτε θα αποφασίσουμε αν αξίζει το feature selection να γίνει με η χωρίς το sample που αφήνουμε εκτός κατά το leave-one-out.

Υπήρχαν τρία ενδεχόμενα:

Classify without feature selection

Σε αυτό το κομμάτι δε κάνουμε feature selection και χρησιμοποιούμε και τα 1000 χαρακτηριστικά για να κάνουμε training. Όμως αυτό δημιουργεί overfitting με αποτέλεσμα να χάνουμε την απόδοση της προβλέψεις εφόσον αυξάνουμε τις διαστάσεις. Το accuracy που παίρνουμε είναι 0.64 .

Classify with feature selection inside the cross validation

Σε αυτή την περίπτωση μειώσαμε τις διαστάσεις αλλά και features από 1000 σε 100. Θεωρητικά αυτή η μέθοδος είναι η πιο σωστή καθώς δε λαμβάνει υπόψη το testing data άρα το testing έχει νόημα να υπάρχει. Το accuracy που παίρνουμε είναι 0.52 .

Classify with feature selection outside the cross validation

Σε αυτή την περίπτωση κάνουμε το feature selection εκτός του cross validation οπότε λαμβάνει υπόψη όλα τα samples. Άρα το cross validation ως testing δε έχει νόημα καθώς δε ελέγχουμε κάποιο καινούργιο sample , με αποτέλεσμα να γίνετε overfitting. Το accuracy που παίρνουμε είναι 1 .

Επειδή τα δεδομένα μας είναι κανονικοποιημενα τυχαία δε μας βοηθά να προσεγγίσουμε με ακρίβεια το πρόβλημα μας καθώς όπως είναι προφανές καθώς δε γίνεται να είναι διαχωρισμένα στις δυο κλάσεις.

Το θέμα 1 υλοποιήθηκε σε Matlab 2016.

Θέμα 2.

Τα α,β,γ είναι υλοποιημένα στο τέλος του pdf

Ο σκοπός της άσκησης ήταν η δημιουργία ενός απλού νευρωνικού δικτύου. Το δίκτυο που δημιουργήθηκε είχε μόνο είσοδο και έξοδο χωρίς κάποιο κρυφό επίπεδο ενδιάμεσα. Η συνάρτηση ενεργοποίησης ήταν η sigmoid :

$$f(z) = \frac{1}{1 + e^{-z}}$$

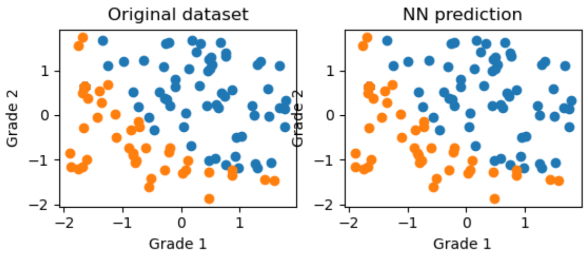
Ενώ για τον υπολογισμό του σφάλματος χρησιμοποιήθηκε η cross entropy όπως οριστικέ προηγουμένως.

Τα βήματα είναι τα εξής:

- Είναι ο υπολογισμός της εξόδου του δικτύου μας . Την είδος την πολλαπλασιάζουμε με τα βάρη ενώ προσθέτουμε το bias , ενώ στην συνεχεία το αποτέλεσμα αυτό το βάζουμε σαν είσοδο στην sigmoid.
- Επόμενο βήμα είναι η αξιολόγηση του αποτελέσματα με την cross entropy .

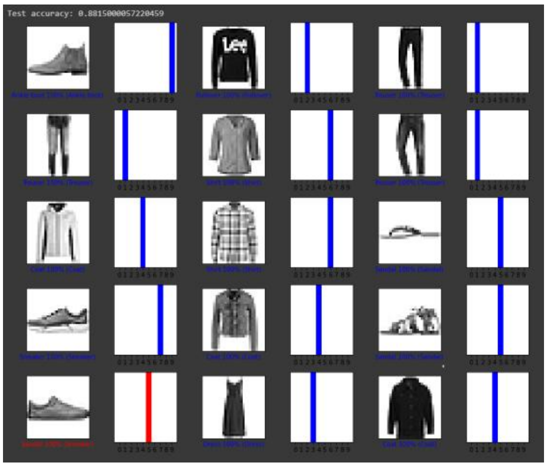
- Το τελευταίο βήμα είναι το backward που χρησιμοποιείται για να ελαχιστοποιηθούν οι απώλειες και να ανανεωθούν οι τιμές των βαρών και του bias.

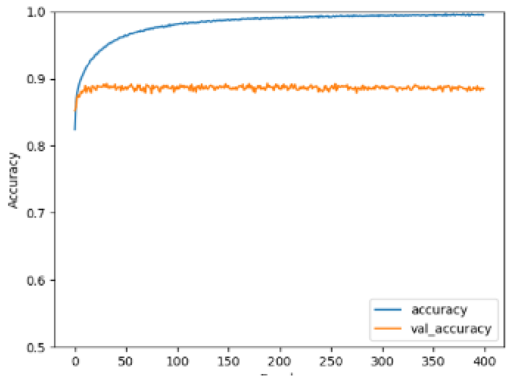
Τα συγκεκριμένα βήματα γίνονται επαναληπτικά τόσες φορές όσες και τα epoch. Για 55 epoch παίρνουμε γύρω στα 54% πιθανότητα προβλέψεις για τα δείγματα 45 έως 85 και 0.9 accuracy . Για 500 epochs βλέπουμε η πιθανότητα να βελτιώνεται στο 64% ενώ το accuracy μένει πάλι στο 0.9 .



Θέμα 3.

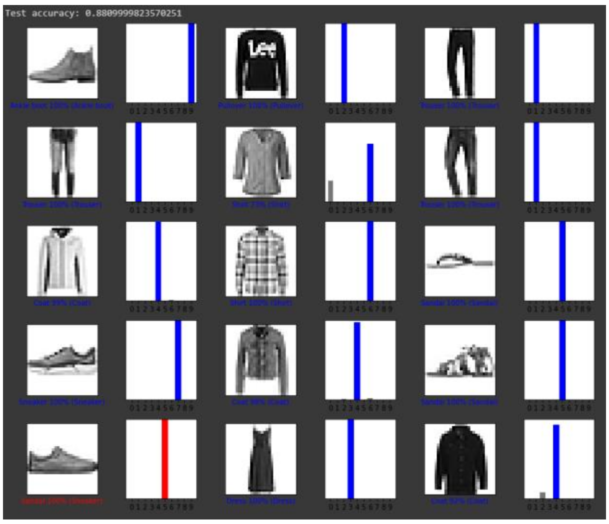
Στο συγκεκριμένο θέμα μας ζητήθηκε να δημιουργήσουμε ένα Deep CNN για την αναγνώριση των εικόνων από το γνωστό dataset fashion_mnist. Στην αρχή της άσκησης μας ζητήθηκε να πειραματιστούμε με έτοιμο κώδικα που μας δόθηκε . Αρχικά τρέξαμε τον έτοιμο κώδικα για να παρατηρήσουμε την εξέλιξη των accuracy και validation accuracy. Παρατηρησαμε ότι για τον Adam optimizer το accuracy τηνει στην μοναδα ενώ το validation accuracy φτάνει μέχρι το 88%.



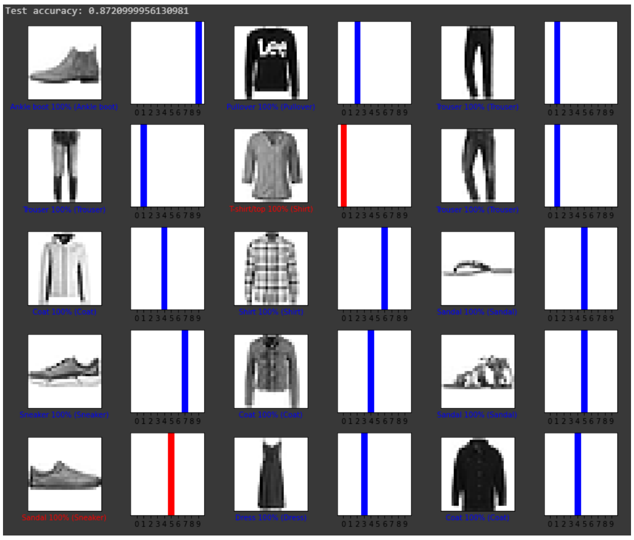


Adam

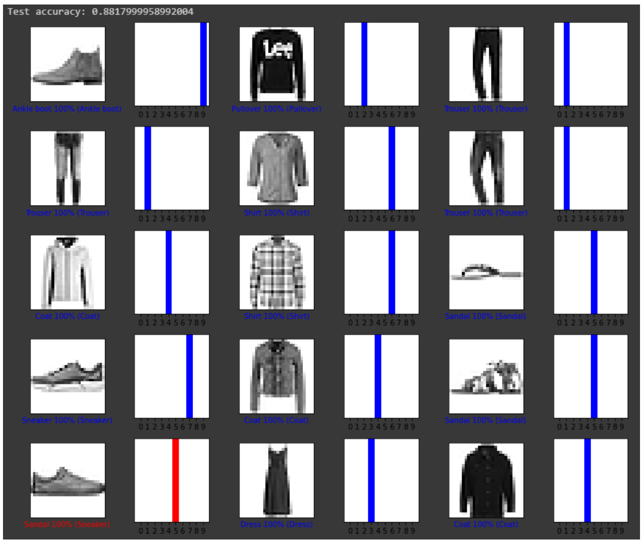
Έπειτα για την συνεχεία της άσκησης χρειάστηκε να ξανά τρέξουμε τον προηγούμενο κώδικα για διαφόρους optimizers (sgd, rmsprop, nadam, adamax, και ftrl).



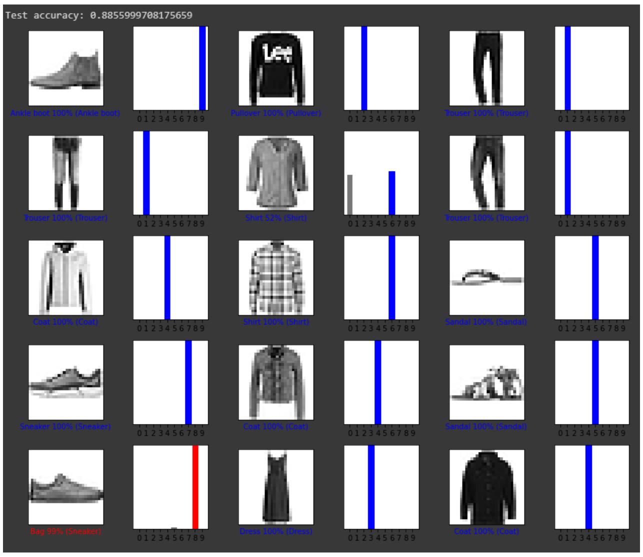
sgd



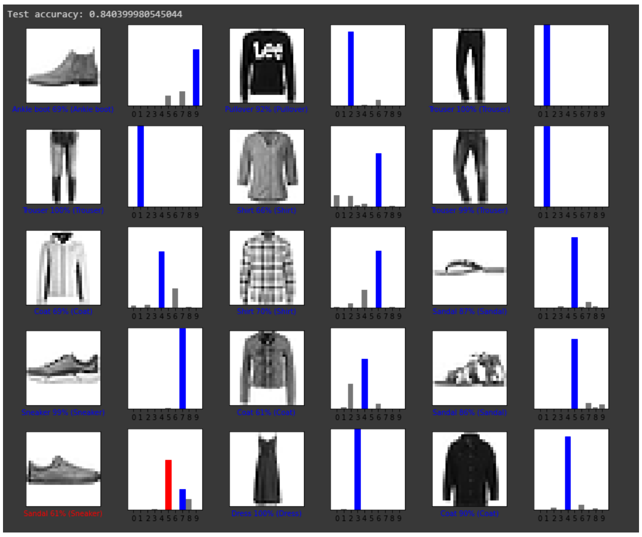
rmsprop



nadam

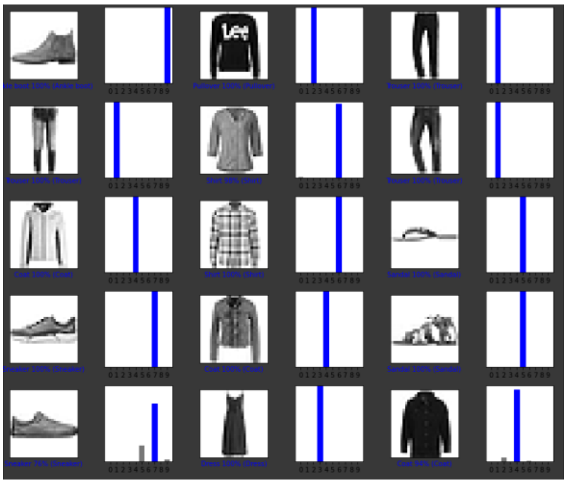


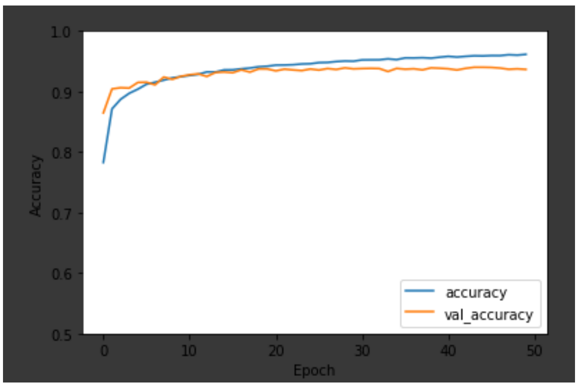
adamax



ftrl

Έπειτα φτιάξαμε ένα Deep CNN με βάση την αρχιτεκτονική που μας έδωσε η άσκηση. Αυτό που παρατηρούμε είναι ότι μετά την ολοκλήρωση του δικτύου και την εκπαίδευση του μοντέλου το accuracy και το validation accuracy σχεδόν ταυτίζονται (96% και 94% αντίστοιχα) . Το τελικό αποτέλεσμα είναι το εξής :





***

Η υλοποίηση των α,β,γ από το θέμα 2:

Άσκηση 2

$$J(Y,\hat{Y};w,b)=\frac{1}{B}\sum_i\left(-y^{(i)}\ln(\hat{y}^{(i)})-(1-y^{(i)})\ln(1-\hat{y}^{(i)})\right) \quad (1)$$

α) Θέτω $\hat{y}^{(i)}=f(z^{(i)})=\frac{1}{1+e^{-z^{(i)}}}$

$$(1)\Rightarrow \frac{1}{B}\sum_i\left(-y^{(i)}\ln\left(\frac{1}{1+e^{-z^{(i)}}}\right)-(1-y^{(i)})\ln\left(1-\frac{1}{1+e^{-z^{(i)}}}\right)\right)=$$

$$=\frac{1}{B}\sum_i\left(y^{(i)}\ln(e^{-z^{(i)}}+1)-\ln\left(1-\frac{1}{1+e^{-z^{(i)}}}\right)+y^{(i)}\ln\left(1-\frac{1}{1+e^{-z^{(i)}}}\right)\right)=$$

$$=\frac{1}{B}\sum_i\left(y^{(i)}\left(\ln\left((e^{-z^{(i)}}+1)\left(1-\frac{1}{1+e^{-z^{(i)}}}\right)\right)\right)-\ln\left(1-\frac{1}{1+e^{-z^{(i)}}}\right)\right)=$$

$$=\frac{1}{B}\sum_i\left(y^{(i)}\ln(e^{-z^{(i)}}+1-1)-\ln\left(1-\frac{1}{1+e^{-z^{(i)}}}\right)\right)=\frac{1}{B}\sum_i\left(-y^{(i)}z^{(i)}-\ln\left(\frac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}}\right)\right)=$$

$$=\frac{1}{B}\sum_i\left(-y^{(i)}z^{(i)}-\ln(e^{-z^{(i)}})+\ln(1+e^{-z^{(i)}}\right)=\frac{1}{B}\sum_i\left(z^{(i)}-y^{(i)}z^{(i)}+\ln(1+e^{-z^{(i)}})\right)$$

β) $$\frac{d\hat{y}^{(i)}}{dz^{(i)}}=\frac{d\sigma(z^{(i)})}{dz^{(i)}}=\frac{d\frac{1}{1+e^{-z^{(i)}}}}{dz^{(i)}}=\frac{-1}{(1+e^{-z^{(i)}})^2}e^{-z^{(i)}}(-1)=\frac{1}{1+e^{-z^{(i)}}}\frac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}} \quad (2)$$

Επίσης $$1-\sigma(z^{(i)})=1-\frac{1}{1+e^{-z^{(i)}}}=\frac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}} \quad (3)$$

$$(2)(3)\Rightarrow \frac{d\hat{y}^{(i)}}{dz^{(i)}}=\frac{1}{1+e^{-z^{(i)}}}\frac{e^{-z^{(i)}}}{1+e^{-z^{(i)}}}=\sigma(z^{(i)})(1-\sigma(z^{(i)}))=\hat{y}^{(i)}(1-\hat{y}^{(i)})$$

Για να υπολογίσω $\frac{dJ}{dz^{(i)}}$, θα βρω πρώτα:

$$\frac{dJ}{d\hat{y}^{(i)}}=\frac{d\left(-y^{(i)}\ln(\hat{y}^{(i)})-(1-y^{(i)})\ln(1-\hat{y}^{(i)})\right)}{d\hat{y}^{(i)}}=\frac{d(-y^{(i)})\ln(\hat{y}^{(i)})}{d\hat{y}^{(i)}}+\frac{d(-(1-y^{(i)})\ln(1-\hat{y}^{(i)}))}{d\hat{y}^{(i)}}=$$

$$=-\frac{y^{(i)}}{\hat{y}^{(i)}}+\frac{1-y^{(i)}}{1-\hat{y}^{(i)}}=\frac{\hat{y}^{(i)}-y^{(i)}}{\hat{y}^{(i)}(1-\hat{y}^{(i)})}$$

Από κανόνα αλυσίδας

$$\frac{dJ}{dz^{(i)}} = \frac{dJ}{d\hat{y}^{(i)}} \cdot \frac{d\hat{y}^{(i)}}{dz^{(i)}} = \frac{\hat{y}^{(i)} - y^{(i)}}{\hat{y}^{(i)}(1-\hat{y}^{(i)})} \cdot \frac{\hat{y}^{(i)}(1-\hat{y}^{(i)})}{1} = \hat{y}^{(i)} - y^{(i)}$$

1) $$\frac{dJ}{dw} = \frac{dJ}{d\hat{y}^{(i)}} \frac{d\hat{y}^{(i)}}{dz^{(i)}} \cdot \frac{dz^{(i)}}{dw}$$

Είναι $$\frac{dz^{(i)}}{dw} = \frac{d(x^{(i)}w+b)}{dw} = x^{(i)T}$$

Άρα $$\frac{dJ}{dw} = (\hat{y}^{(i)} - y^{(i)})x^{(i)T}$$

$$\frac{dJ}{db} = \frac{dJ}{d\hat{y}^{(i)}} \cdot \frac{d\hat{y}^{(i)}}{dz^{(i)}} \cdot \frac{dz^{(i)}}{db}$$

Είναι $$\frac{dz^{(i)}}{db} = \frac{d(x^{(i)}w+b)}{db} = \cancelto{0}{\frac{d(x^{(i)}w)}{db}} + \frac{db}{db} = 1$$

Άρα $$\frac{dJ}{db} = (\hat{y}^{(i)} - y^{(i)})1 = \hat{y}^{(i)} - y^{(i)}$$