



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

ΔΠΜΣ Επιστήμη Δεδομένων και Μηχανική Μάθηση

Ανάλυση και Επεξεργασία Γεωχωρικών Δεδομένων - Σειρά 5η

Σωτήρης Πελέκης

03400069

ΕΔΕΜΜ

10/05/2020

Περιεχόμενα

Άσκηση 1.....	3
Άσκηση 2.....	3
Ερωτήματα B1, B2 – Εκπαίδευση, επιλογή παραμέτρων και αξιολόγηση ταξινομητών και σχολιασμοί.....	3
Κλασσική μηχανική μάθηση	3
Βαθιά μάθηση.....	6
Ερωτήμα B3	9
B.3.1 Διαφορές ακρίβειας ανάμεσα στις κατηγορίες	9
B.3.1 Overfitting / Underfitting	10
B.3.3 Συντελεστές μάθησης	10
B.3.3 Αριθμός εποχών	11

Άσκηση 1

Τα ζητούμενα της άσκησης 1 απαντώνται αναλυτικά εντός του αντίστοιχου notebook για αυτό δεν παρατίθενται και εδώ.

Η γλώσσα στο notebook είναι Αγγλικά για λόγους συμπερίληψης σε portfolio. Εάν υπάρξει κάποιο ζήτημα με αυτό παρακαλώ ενημερώστε με να σας στείλω μια μεταφρασμένη έκδοση.

Όσον αφορά τον αλγόριθμο Perceptron που ζητήθηκε, υλοποιείται στο αρχείο `custom_perceptron.py`. Ο αλγόριθμος γράφτηκε καθαρά με χρήση `numpy` όπως ζητήθηκε, ωστόσο χρησιμοποιήθηκε το πρότυπο της `scikit-learn (base)` για λόγους τυποποίησης και ευκολίας στη χρήση

Άσκηση 2

Στο αντίστοιχο notebook υπάρχουν αναλυτικά σχόλια και επεξηγήσεις για όλα τμήματα της άσκησης. Ωστόσο παρατίθενται και εδώ οι άμεσες απαντήσεις στα επιμέρους ερωτήματα της άσκησης για να είναι πιο ευανάγνωστα. Όσον αφορά το pipeline εκπαίδευσης και αξιολόγησης των νευρωνικών δικτύων έχει υλοποιηθεί στο αρχείο `nn_training_evaluation.py`.

Ερωτήματα B1, B2 – Εκπαίδευση, επιλογή παραμέτρων και αξιολόγηση ταξινομητών και σχολιασμοί

Κλασσική μηχανική μάθηση

Έγινε χρήση Stratified 5-fold Cross Validation για πολλούς διαφορετικούς συνδυασμούς υπερπαραμέτρων για τους αλγορίθμους SVM και Random Forest (μέσω `GridSearchCV` πάνω στο training set). Τα scores είναι χαμηλά καθώς το Cross Validation αποτελεί αρκετά αυστηρή μέθοδο αξιολόγησης της επίδοσης ενός αλγορίθμου και πολλές φορές στο test set τα πράγματα είναι καλύτερα. Ωστόσο αποτελεί μια πολύ καλή μέθοδο συγκρίσης επίδοσης αλγορίθμων μάθησης.

SVM

Εκπαίδευση και επιλογή υπερπαραμέτρων

C	gamma	kernel	Score
10	scale	rbf	0.886254
10	scale	linear	0.848480
10	auto	rbf	0.886533
10	auto	linear	0.848480
100	scale	rbf	0.906746
100	scale	linear	0.836492
100	auto	rbf	0.906606
100	auto	linear	0.836492
1000	scale	rbf	0.903401
1000	scale	linear	0.836073
1000	auto	rbf	0.903401
1000	auto	linear	0.836073

Best parameters set found on training set:

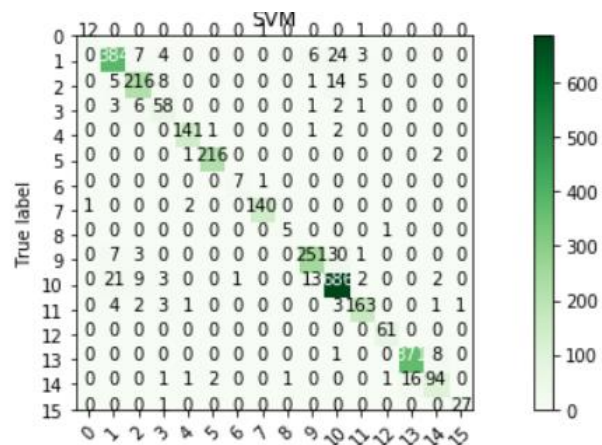
```
{'C': 100, 'gamma': 'scale', 'kernel': 'rbf'}
```

Η παράμετρος C του SVM δηλώνει την ανοχή της συνάρτησης κόστους σε σφάλματα. Η παράμετρος $kernel$ ορίζει τη μορφή της συνάρτησης πυρήνα που χρησιμοποιείται για προβολή των εσωτερικών γινομένων που υπολογίζονται κατά την ελαχιστοποίηση της συνάρτησης κόστους σε χώρο υψηλότερων διαστάσεων. Η παράμετρος $gamma$ ορίζει με ένα τρόπο τη διασπορά – περιοχή επιρροής των συναρτήσεων πυρήνα. Βλέπουμε ότι γενικά ο τρόπος ορισμού του $gamma$ δεν επιδρά στα αποτελέσματα μας καθόλου. Καίρια σημασία ωστόσο παίζει η υπερπαράμετρος ανοχής σφάλματος καθώς ρυθμίζει την ποιότητα του fitting αλλά και το σχήμα του πυρήνα. Στην προκειμένη ο Gaussian (rbf) πυρήνας φέρνει καλύτερα αποτελέσματα.

Αξιολόγηση

Όσον αφορά την αξιολόγηση του μοντέλου επιβεβαιώνεται η αυστηρότητα του Cross Validation καθώς το επιλεγμένο μοντέλο `{'C': 100, 'gamma': 'scale', 'kernel': 'rbf'}` φέρνει τα ακόλουθα αποτελέσματα του Classification Report:

	precision	recall	f1-score	support
0	0.923	0.857	0.889	14
1	0.906	0.897	0.901	428
2	0.889	0.867	0.878	249
3	0.744	0.817	0.779	71
4	0.966	0.972	0.969	145
5	0.986	0.986	0.986	219
6	0.875	0.875	0.875	8
7	0.986	0.979	0.982	143
8	0.833	0.833	0.833	6
9	0.919	0.860	0.888	292
10	0.900	0.931	0.915	737
11	0.926	0.916	0.921	178
12	0.968	1.000	0.984	61
13	0.959	0.976	0.967	380
14	0.879	0.810	0.843	116
15	0.964	0.964	0.964	28
accuracy			0.921	3075
macro avg	0.914	0.909	0.911	3075
weighted avg	0.921	0.921	0.921	3075



Βλέπουμε ότι ο αλγόριθμος επιτυγχάνει 92.1 % με την κλάση 4 (καλαμπόκι) να είναι πολύ προβληματική τόσο ως προς precision και ως προς recall γεγονός που δείχνει ότι παράγονται και False Positives αλλά και False Negatives ως προς αυτή. Η συγκεκριμένη κλάση δηλαδή «κλέβει» προβλέψεις από άλλες κλάσεις ενώ ταυτόχρονα πολλά pixels της γίνονται assign σε άλλη κλάση.

Random Forest

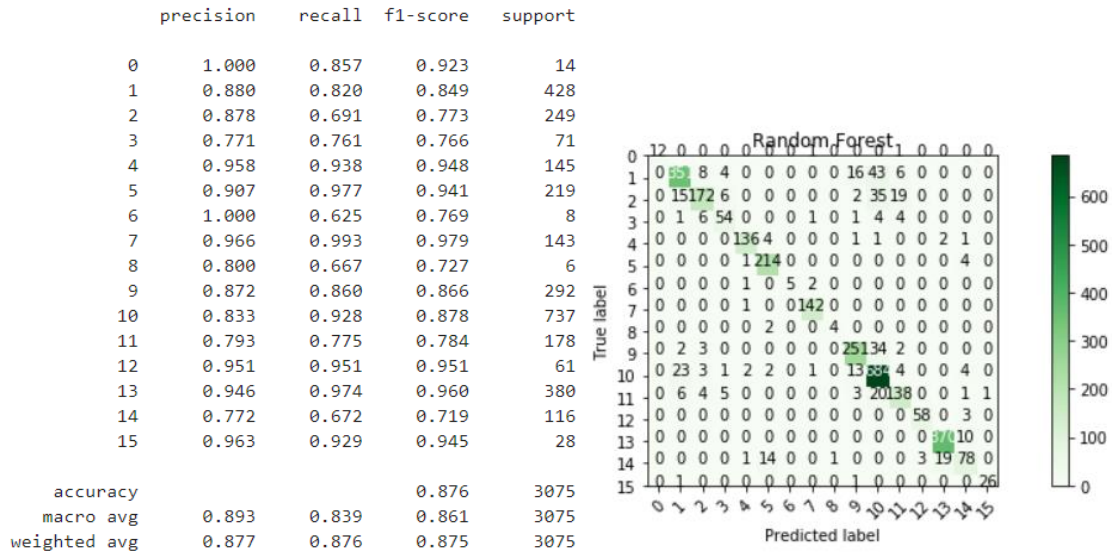
Εκπαίδευση και επιλογή υπερπαραμέτρων

max_depth	min_samples_leaf	n_estimators	Score
4.0	1	200	0.562308
4.0	1	600	0.562308
4.0	2	200	0.562029
4.0	2	600	0.561193
4.0	5	200	0.561890
4.0	5	600	0.562587
10.0	1	200	0.808752
10.0	1	600	0.810285
10.0	2	200	0.809728
10.0	2	600	0.810007
10.0	5	200	0.798158
10.0	5	600	0.801922
NaN	1	200	0.856563
NaN	1	600	0.857121
NaN	2	200	0.849873
NaN	2	600	0.852940
NaN	5	200	0.837048
NaN	5	600	0.836352

Όσον αφορά τον αλγόριθμο Random Forest, γίνεται προφανές ότι το κλάδεμα (pruning) δε λειτουργεί πολύ καλά, δηλαδή ούτε ο περιορισμός βάθους αλλά ούτε ο περιορισμός του ελάχιστου αριθμού δειγμάτων ανά φύλλο βοηθάνε στη βελτίωση του. Προτιμάται βέβαια η χρήση ενός μεγάλου αριθμού εκτιμητών (600) η οποία προσθέτει στιβαρότητα στο μοντέλο και συχνά λύνει από μόνο της τα προβλήματα overfitting που παρουσιάζει κατά κανόνα ο Random Forest. Το καλύτερο λοιπόν μοντέλο Random Forest έχει τις υπερπαραμέτρους **{'max_depth': None, 'min_samples_leaf': 1, 'n_estimators': 600}**

Αξιολόγηση

Όσον αφορά την αξιολόγηση του παραπάνω μοντέλου επιβεβαιώνεται η αυστηρότητα του Cross Validation καθώς το επιλεγμένο μοντέλο {'C': 100, 'gamma': 'scale', 'kernel': 'rbf'} φέρνει τα ακόλουθα αποτελέσματα του Classification Report:



Και πάλι ο αλγόριθμος τα πάει καλύτερα στο test set με 87.6%.

Σημαντικές παρατηρήσεις:

1. Η εκπαίδευση του Random Forest είναι πολύ αργή πόσο μάλλον το κούρδισμα υπερπαραμέτρων με χρήση CV.
2. Η πολλαπλή επικύρωση και ο μεγάλος αριθμός εκτιμητών εμποδίζουν τον αλγόριθμο να κάνει overfit.
3. Και σε αυτή την περίπτωση η αυστηρότητα του Cross Validation προκαλεί αρκετά χαμηλά scores τα οποία αυξάνονται στο evaluation stage.
4. Η κλάση καλαμποκιού συνεχίζει να παρουσιάζει έντονα προβλήματα παρά την αλλαγή του αλγορίθμου. Ωστόσο εδώ βλέπουμε να προβλέπει άσχημα και την ανάμεικτη κλάση 15 (buildings-grass-trees-drives) που ενδείκνυται να οφείλεται στο overfitting του random forest όπου μια κλάση με αρκετά ετερόκλητα στοιχεία ενδεχομένως να τον μπερδεύει.

Βαθιά μάθηση

Δοκιμάστηκαν 3 MLPs διαφορετικού βάθους (αρχείο neural_nets.py) τα οποία έφεραν πολύ κοντινά αποτελέσματα. Έγινε χρήση early stopping χρησιμοποιώντας ένα 10 % του training set για σκοπούς επικύρωσης – επιλογής υπερπαραμέτρων (learning rate, number of epochs, regularization coefficients). Για επίτευξη του early stopping έγινε χρήση μιας μεταβλητής αντίστροφης μέτρησης για αποθήκευση του καλύτερου μοντέλου στο δίσκο μέχρι να βρεθεί καλύτερο.

Ρηχό MLP – 1 hidden layer

Εκπαίδευση και επιλογή υπερπαραμέτρων

Το ρηχό MLP (1 κρυφό στρώμα) επιτυγχάνει σταθερή απόδοση περίπου 91 % σε περίπου 60 εποχές χρησιμοποιώντας κανονικό ρυθμό εκμάθησης (0,001) Το early stopping εφαρμόζεται αποθηκεύοντας το μοντέλο της κόκκινης κουκκίδας στο δίσκο έτσι ώστε το καλύτερο μοντέλο να διατηρείται για περαιτέρω αξιολόγηση στο test set.

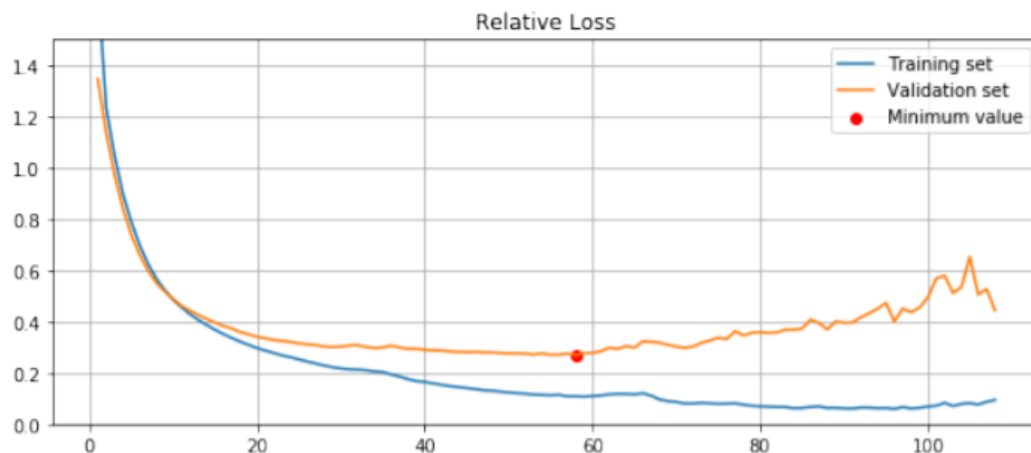


Figure 1: Καμπύλη μάθησης ρηχού MLP και early stopping point

Αξιολόγηση

Τα αποτελέσματα ήταν τα εξής της αξιολόγησης στο σύνολο ελέγχου είναι τα εξής:

	precision	recall	f1-score	support
0	0.867	0.929	0.897	14
1	0.903	0.911	0.907	428
2	0.864	0.843	0.854	249
3	0.786	0.775	0.780	71
4	0.952	0.959	0.955	145
5	0.990	0.950	0.970	219
6	1.000	1.000	1.000	8
7	0.965	0.972	0.969	143
8	0.857	1.000	0.923	6
9	0.900	0.866	0.883	292
10	0.901	0.898	0.899	737
11	0.853	0.916	0.883	178
12	0.984	0.984	0.984	61
13	0.971	0.889	0.929	380
14	0.658	0.879	0.753	116
15	0.862	0.893	0.877	28
accuracy			0.901	3075
macro avg	0.895	0.916	0.904	3075
weighted avg	0.905	0.901	0.902	3075

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	390	6	3	0	0	0	0	0	0	5	20	4	0	0	0	0
1	3	210	9	0	0	0	0	0	1	21	5	0	0	0	0	0
2	0	1	8	55	0	0	0	3	0	0	4	0	0	0	0	0
3	0	0	0	0	139	0	0	1	0	0	0	2	0	2	1	0
4	0	0	0	0	0	208	0	0	1	1	1	0	0	0	8	0
5	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	2	0	0	0	0	0	0	139	0	0	2	0	0	0	0	0
8	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0
9	0	9	2	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	23	14	2	0	0	0	1	0	19	662	8	0	0	5	3
11	0	5	3	0	0	0	0	0	0	2	4	163	0	0	0	1
12	0	0	0	0	0	0	0	0	0	0	0	0	60	0	1	0
13	0	0	0	0	5	0	0	0	0	0	0	0	0	338	37	0
14	0	0	0	0	2	2	0	0	0	0	0	1	1	8	102	0
15	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	25

Παρατηρούμε έντονα προβλήματα (άλλης) κλίμακας στις 2 «δυσκολες» κλάσεις που προαναφέρθηκαν.

MLP μικρού βάθους – 2 hidden layers

Εκπαίδευση και επιλογή υπερπαραμέτρων

Το MLP μικρού βάθους (2 κρυμμένα στρώματα) επιτυγχάνει μια εξίσου καλή απόδοση (~ 91%) σε περίπου 50-60 εποχές με λίγο χαμηλότερο μάλιστα learning rate (0.0005).

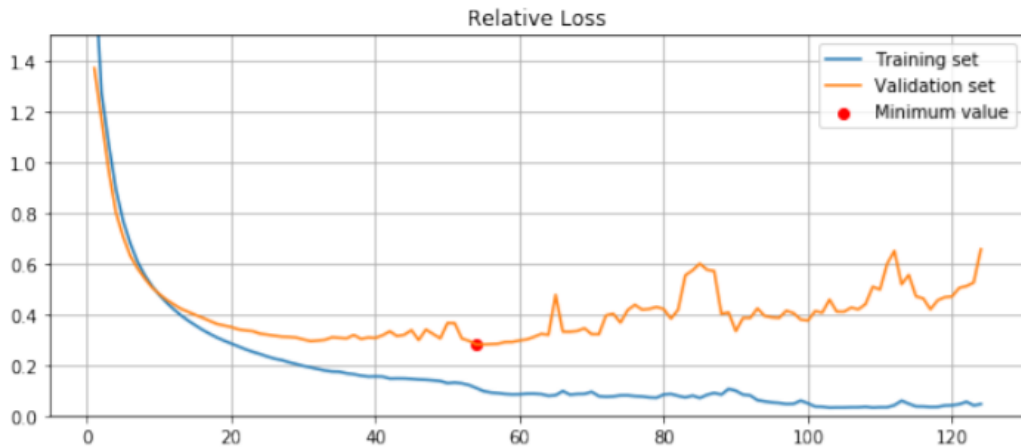
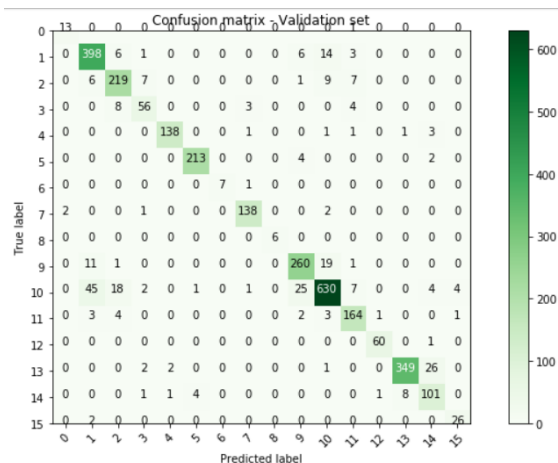


Figure 2: Καμπύλη μάθησης MLP μετρίου βάθους και early stopping point

Αξιολόγηση

Τα αποτελέσματα ήταν τα εξής της αξιολόγησης στο σύνολο ελέγχου είναι τα εξής:

	precision	recall	f1-score	support
0	0.867	0.929	0.897	14
1	0.856	0.930	0.891	428
2	0.855	0.880	0.867	249
3	0.800	0.789	0.794	71
4	0.979	0.952	0.965	145
5	0.977	0.973	0.975	219
6	1.000	0.875	0.933	8
7	0.958	0.965	0.962	143
8	1.000	1.000	1.000	6
9	0.872	0.890	0.881	292
10	0.928	0.855	0.890	737
11	0.872	0.921	0.896	178
12	0.968	0.984	0.976	61
13	0.975	0.918	0.946	380
14	0.737	0.871	0.798	116
15	0.839	0.929	0.881	28
accuracy			0.903	3075
macro avg	0.905	0.916	0.910	3075
weighted avg	0.907	0.903	0.904	3075



Βλέπουμε ότι εν γένει δεν έχουν αλλάξει και πάρα πολλά πράγματα.

Βαθύ MLP – 4 hidden layers

Εκπαίδευση και επιλογή υπερπαραμέτρων

Το βαθύ MLP (4 κρυμμένα στρώματα) επιβεβαίωσε το γεγονός ότι η εμβάθυνση της αρχιτεκτονικής και η αύξηση του αριθμού των παραμέτρων δεν βελτιώνουν περεταίρω τα αποτελέσματα. Ένας μικρός συντελεστής κανονικοποίησης (regularization) προστέθηκε επίσης στη συνάρτηση κόστους για την αποφυγή overfitting και την πέτυχε ωστόσο δε βελτίωσε τα αποτελέσματα.

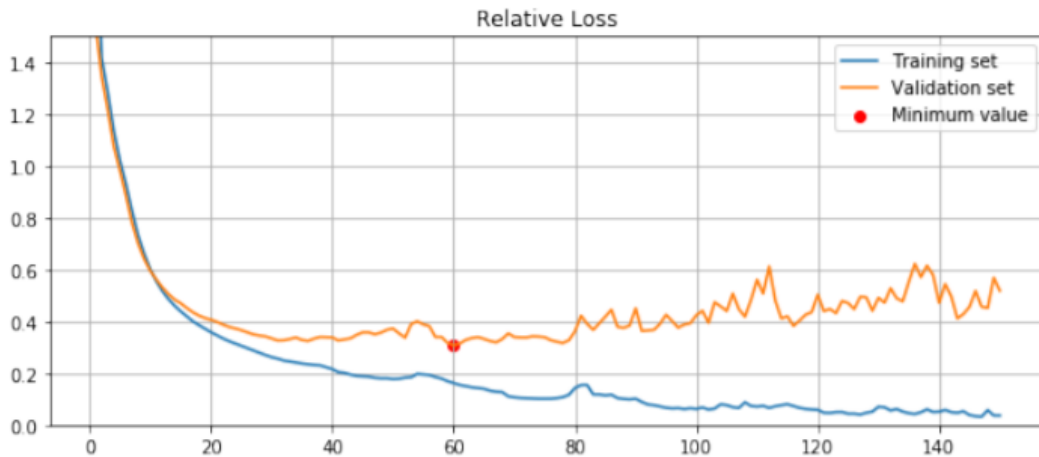


Figure 3: Καμπύλη μάθησης βαθύς MLP και early stopping point

Αξιολόγηση

Τα αποτελέσματα ήταν τα εξής της αξιολόγησης στο σύνολο ελέγχου είναι τα εξής:

	precision	recall	f1-score	support
0	0.929	0.929	0.929	14
1	0.882	0.904	0.893	428
2	0.900	0.795	0.844	249
3	0.782	0.859	0.819	71
4	0.922	0.972	0.946	145
5	0.973	0.982	0.977	219
6	0.778	0.875	0.824	8
7	0.972	0.972	0.972	143
8	1.000	0.667	0.800	6
9	0.863	0.884	0.873	292
10	0.900	0.881	0.890	737
11	0.835	0.938	0.884	178
12	0.984	1.000	0.992	61
13	0.969	0.916	0.942	380
14	0.762	0.828	0.793	116
15	0.926	0.893	0.909	28
accuracy			0.900	3075
macro avg	0.898	0.893	0.893	3075
weighted avg	0.902	0.900	0.901	3075

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	387	4	0	0	0	0	0	0	10	23	4	0	0	0	0
2	0	2	198	13	0	0	0	0	0	0	22	14	0	0	0	0
3	0	0	3	61	0	0	0	2	0	2	0	3	0	0	0	0
4	0	0	0	0	141	0	0	1	0	0	1	1	0	1	0	0
5	0	0	0	0	1	215	0	0	0	2	0	0	0	0	1	0
6	0	0	0	0	0	7	1	0	0	0	0	0	0	0	0	0
7	1	0	0	0	0	0	1	139	0	0	2	0	0	0	0	0
8	0	0	0	1	0	0	0	0	4	0	0	0	1	0	0	0
9	0	10	0	0	0	0	0	0	0	258	23	1	0	0	0	0
10	0	35	12	1	1	1	1	0	0	24	649	9	0	0	3	1
11	0	3	3	0	0	0	0	0	0	3	1	167	0	0	0	1
12	0	0	0	0	0	0	0	0	0	0	0	0	61	0	0	0
13	0	0	0	0	6	0	0	0	0	0	0	0	0	348	26	0
14	0	0	0	1	4	5	0	0	0	0	0	0	0	10	96	0
15	0	7	0	1	0	0	0	0	0	0	0	0	0	0	0	25

Παρατηρείται μια κάπως πιο ισορροπημένη συμπεριφορά ως προς τις 2 «ύποπτες» κλάσεις. Η βελτίωση αυτή όμως δε βελτιώνει τη συνολική ευστοχία καθώς αυτές οι 2 έχουν μικρό πλήθος παρατηρήσεων στο δείγμα.

Τα παραπάνω συμπεράσματα ισχύουν κατά προσέγγιση και επιβεβαιώθηκαν για πολλές επαναλήψεις των διαδικασιών εκπαίδευσης και αξιολόγησης.

Ερωτήμα Β3

Β.3.1 Διαφορές ακρίβειας ανάμεσα στις κατηγορίες

Όπως φαίνεται και στα σχετικά ιστογράμματα του notebook οι κλάσεις είναι εντελώς ανισοκατανομημένες ως προς το πλήθος των παρατηρήσεων τους. Όπως αναφέρθηκε και στην προηγούμενη άσκηση κάθε αλγόριθμος μάθησης αναπτύσσει δικά του χαρακτηριστικά τα οποία στο τέλος ενδεχομένως να ευνοούν την πρόβλεψη κάποια κλάσης υπέρ μιας άλλης. Αυτό

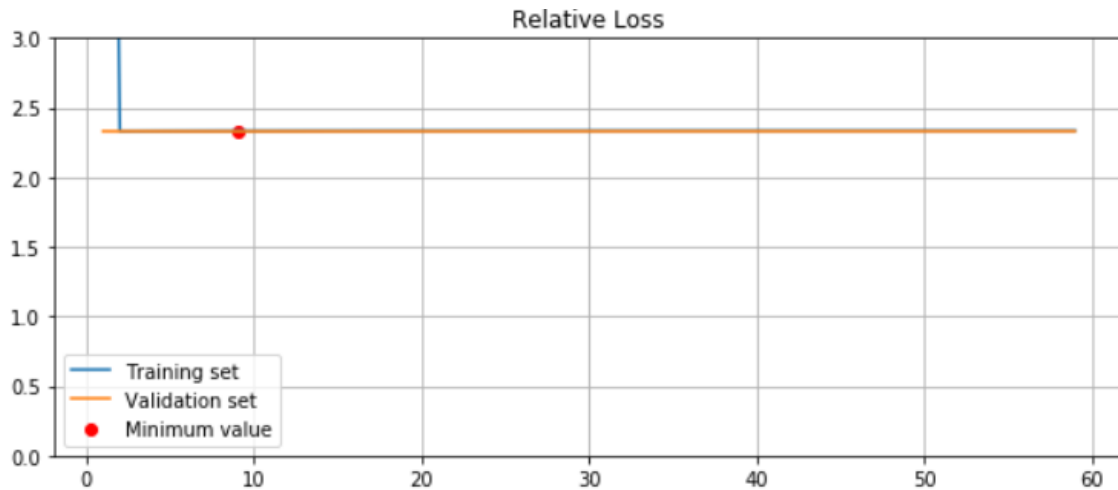
συμβαίνει, σε μεγάλο βαθμό καθώς η διαδικασία του hyperparameter tuning είναι εστιασμένη στο να μεγιστοποιεί την ευστοχία του αλγορίθμου. Συνεπώς κλάσεις με πολύ λίγα δείγμα, αν μπερδεύουν το μοντέλο, είναι εύκολο να αμεληθούν κατά τη διαδικασία μάθησης προκειμένου να δοθεί βαρύτητα σε πολυπληθείς κλάσεις όπου η σωστή πρόβλεψη τους ανεβάζει πολύ το συνολικό accuracy. Για παράδειγμα το μεσαίο MLP βλέπουμε ότι προβλέπει πολύ αποτελεσματικά την 9^η κλάση ενώ το ρηχό την 7^η. Ωστόσο, αυτές οι αποκλίσεις επιδρούν ελάχιστα στη συνολική ευστοχία του μοντέλου και είναι σε ένα βαθμό θέμα τύχης το πως θα καταλήξουν. Είναι βασικό εάν κάποιος επιθυμεί καλή πρόβλεψη κάποιας συγκεκριμένης κλάσης να εστιάσει σε βελτιστοποίηση βάσει precision, recall, ή F1 score της συγκεκριμένης κλάσης.

B.3.1 Overfitting / Underfitting

Τα νευρωνικά δίκτυα τείνουν να κάνουν overfit καθώς συνήθως διαθέτουν μεγάλο αριθμό παραμέτρων σε σύγκριση με την πολυπλοκότητα του προβλήματος. Αυτό σημαίνει ότι το training loss μειώνεται πάντα διότι οι πολλές παράμετροι του νευρωνικού με ένα τρόπο προσπαθούν σταδιακά να «αποστηθίσουν» το training set. Ωστόσο το validation loss θα αρχίσει να αυξάνεται μετά από ένα συγκεκριμένο σημείο αφού με την αποστήθιση το δίκτυο χάνει τη δυνατότητα γενίκευσης του σε άγνωστα δεδομένα (overfitting). Αυτό είναι πρόβλημα που μπορεί να λυθεί διατηρώντας ένα μικρό μέρος του training set ως validation set προκειμένου να είναι σε θέση να καθορίσει τις υπερπαραμέτρους εκπαίδευσης (αριθμός εποχών, ρυθμός εκμάθησης, αρχιτεκτονική NN) για να αποφευχθεί η είσοδος στην περιοχή του overfitting (early stopping). Το μειονέκτημα αυτής της μεθόδου είναι ότι ένα μέρος των δεδομένων εκπαίδευσης χάνεται (10% στην περίπτωση μας). Επομένως, υπάρχει ένα tradeoff της απώλειας ορισμένων δεδομένων για την εφαρμογή της σωστής επιλογής αριθμού υπερπαραμέτρων και εποχής. Αντίστοιχα μπορούν να χρησιμοποιηθούν μέθοδοι όπως regularization, dropout κ.α

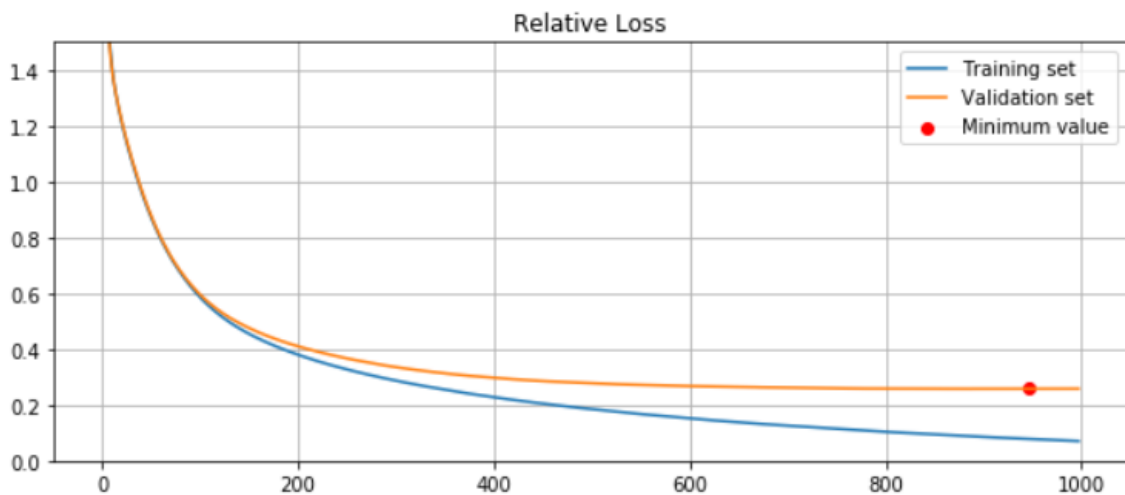
B.3.3 Συντελεστές μάθησης

Η επιλογή συντελεστή μάθησης αποδείχθηκε σημαντική κυρίως όσον αφορά τους χρόνους εκπαίδευσης των μοντέλων. Λόγω του παράγοντα early stopping, οι τιμές του δεν επηρεάζουν τόσο πολύ τις τελικές απώλειες ή την ακρίβεια των μοντέλων. Αυτά δεδομένου ότι κινούνται σε λογικό διάστημα [10^{-2} , 10^{-5}] καθώς υψηλότερα από αυτές τις τιμές learning rates προκαλούν αστάθεια ή αδυναμία μάθησης (Σχ. 4). Τα χαμηλά learning rates απαιτούσαν πολύ μεγάλο αριθμό εποχών για να φτάσουν στα κατάλληλα επίπεδα εκπαίδευσης με ελάχιστα καλύτερα η καθόλου καλύτερα αποτελέσματα (Σχ. 5). Σε ακραίες περιπτώσεις φάνηκε να κολλάνε σε τοπικά ελάχιστα της συνάρτησης κόστους επιτυγχάνοντας λίγο χαμηλότερες ακρίβειες επικύρωσης και υψηλότερα validation losses. Οι συντελεστές μάθησης που επέλεξα δεν εμπίπτουν σε καμία από τις 2 παραπάνω ακραίες κατηγορίες οπότε και ήταν αποτελεσματικοί. Επιπλέον δοκιμάστηκαν διάφοροι συντελεστές μάθησης οι οποίοι δεν έδωσαν καλύτερα tradeoffs μάθησης-χρόνου.



The accuracy achieved is: 0.238 at epoch 9

Figure 4: Καμπύλη μάθησης με μεγάλο learning rate = 0.1 στο ρηχό MLP



The accuracy achieved is: 0.919 at epoch 947

Figure 5: Καμπύλη μάθησης με πολύ μικρό learning rate = 0.00001 στο ρηχό MLP. Το αποτέλεσμα είναι λίγο καλύτερο ωστόσο χρειάστηκαν 950 εποχές για να το φτάσουμε.

B.3.3 Αριθμός εποχών

Early stopping βάσει του Validation set (10% επί του training set)

Εδώ παρατίθενται τα αποτελέσματα ως προς τις εποχές του pipeline μάθησης που ακολουθήθηκε (με κρυφό το test set). Οι κατάλληλοι αριθμοί εποχών για τα 3 MLPs εξάγονται άμεσα από τις κουκκίδες του early stopping. (Σχ. 1,2,3)

MLP1: ~ 60

MLP2: ~ 55

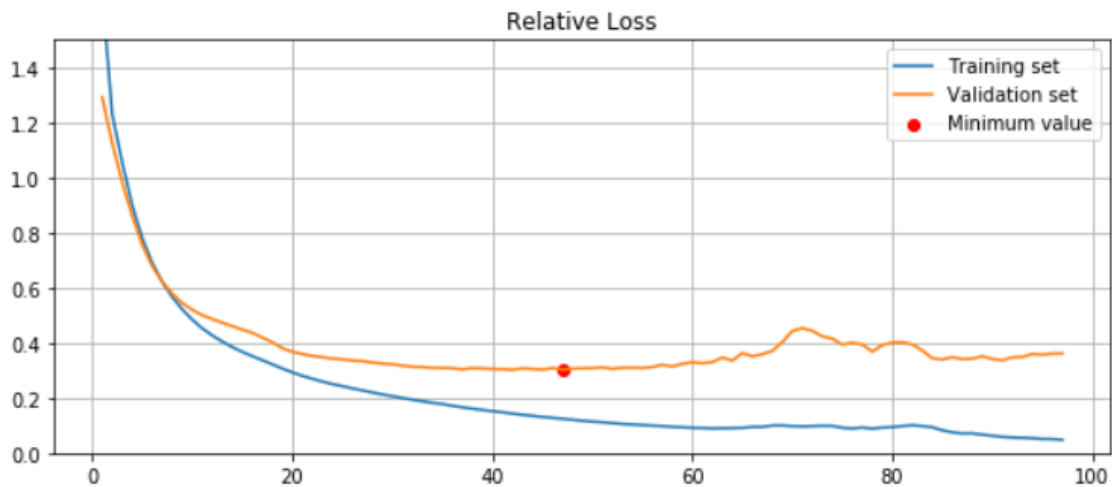
MLP3: ~ 60

Είναι σημαντικό να αναφερθεί πως οι αριθμοί αυτοί αλλάζουν από τρέξιμο σε τρέξιμο λόγω της στοχαστικότητας των αλγορίθμων μάθησης γι' αυτό και αναφέρονται προσεγγιστικά. Ταυτόχρονα, όπως προαναφέρθηκε είναι αρρηκτα συνδεδεμένοι με την επιλογή του learning rate (αντίστροφη σχέση).

Early stopping βάσει του Test set (30% επί του dataset)

Έχοντας αποφασισεί τον κατάλληλο αριθμό epochs με βάση το validation set και αφήνοντας σαν άγνωστο το test set (προσομοίωση πραγματικού προβλήματος) ενδεχομένως οι τιμές να αποκλίνουν από τους πραγματικούς που θεωρητικά και ιδανικά απαιτεί το test set αυτό. Για αυτό και δοκιμάστηκε να βρεθούν και οι πραγματικές τους τιμές χρησιμοποιώντας αυτή τη φορά το test set ως validation set και κρατώντας το training set ακεραίο (70%).

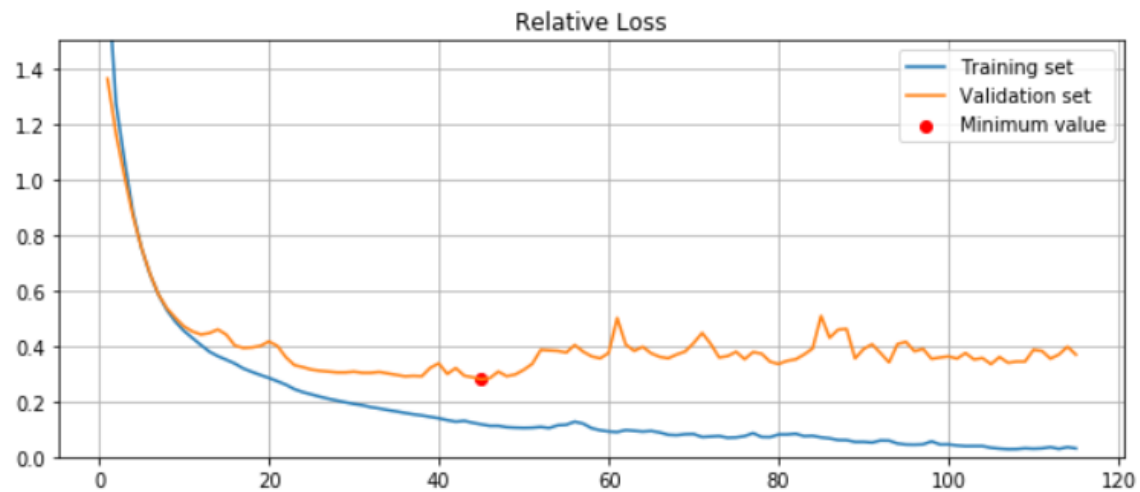
MLP1: Τα αποτελέσματα επιβεβαιώνονται σε ένα βαθμό καθώς τα 58 epochs εκπαίδευσης βρίσκονται μέσα στην «κρίσιμη περιοχή» (Σχ. 6).



The accuracy achieved is: 0.904 at epoch 47

Figure 6: Καμπύλες μάθησης χρησιμοποιώντας το πραγματικό test set (30%) σαν validation set για προσδιορισμό του κατάλληλου αριθμού epochs. Ανάμεσα στα 47 και 60 epochs εντοπίζει κανείς την κρίσιμη περιοχή.

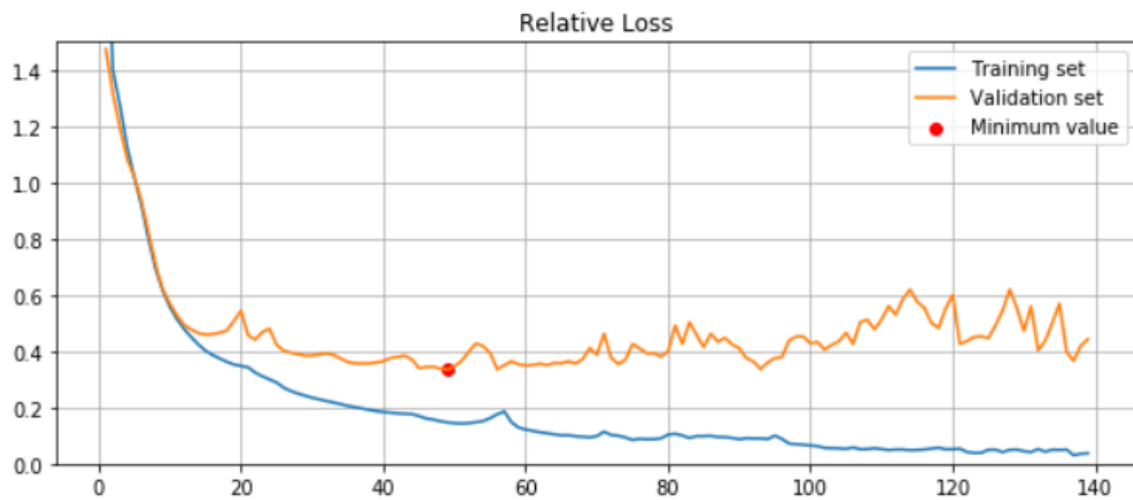
MLP2: Και πάλι βλέπουμε ότι το test set απαιτεί λίγο λιγότερα epochs σε σχέση με το validation set πριν το δίκτυο αρχίσει το overfitting χωρίς όμως να ξεφεύγουμε από τον υπολογισμένο αριθμό



The accuracy achieved is: 0.908 at epoch 45

Figure 7: Καμπύλες μάθησης χρησιμοποιώντας το πραγματικό test set (30%) σαν validation set για προσδιορισμό του κατάλληλου αριθμού epochs. Ανάμεσα στα 45 και 50 epochs εντοπίζει κανείς την κρίσιμη περιοχή.

MLP3: Όμοια



The accuracy achieved is: 0.899 at epoch 49

Figure 8: Καμπύλες μάθησης χρησιμοποιώντας το πραγματικό test set (30%) σαν validation set για προσδιορισμό του κατάλληλου αριθμού epochs.