

Πασόη Σοφία, ΑΜ: 2798

Κωστάκης Ελευθέριος-Παναγιώτης, ΑΜ: 2741

Υλοποιήσαμε την εργασία χρησιμοποιώντας το juryter, διότι ήταν χρήσιμο κατά το φόρτωμα των αρχείων αλλά και στην ανάπτυξη του εκάστοτε αλγορίθμου καθώς μπορούσαμε εύκολα να “τεστάρουμε” κάποια πράγματα χωρίς να πρέπει να ξανα-αρχικοποιούμε. Το αρχείο με τα στοιχεία των credit cards, ήταν πολύ μεγάλο και γενικά δεν έβγαζε καλά σκορ σε σχέση με το αρχείο spambase. Ίσως θα έπρεπε να αφαιρέσουμε κάποιες ακόμα στήλες εκτός από την πρώτη που αφαιρέσαμε με τα ID σε αύξουσα σειρά. Έτσι ήταν εφικτό να γίνει τεστ για τη μέθοδο LVQ χρησιμοποιώντας gaussian similarity.

Για το spambase αρχείο, όπως ήταν αναμενόμενο τα καλύτερα αποτελέσματα ήρθαν από τις μεθόδους SVM και τα νευρωνικά δίκτυα(MLP), αφού το πρόβλημα μας ήταν binary classification. Μετά με πολύ μικρή διαφορά τις καλύτερες επιδόσεις έφερε το νευρωνικό δίκτυο ενός κρυφού επιπέδου. Τέλος ακολουθούν οι μέθοδοι k-Nearest Neighbor, και αρκετά πίσω Naive Bayes και LVQ.

Στο νευρωνικό δίκτυο χρησιμοποιήσαμε τον optimizer “adam” καθώς ήταν πολύ γρήγορος και είχε τα καλύτερα σκορ και είναι μέθοδος που χρησιμοποιεί stochastic gradient descent για την εκπαίδευση του δικτύου.

Μια σύντομη κατάταξη βάσει των αποτελεσμάτων θα ήταν(spambase αρχείο):

1. Neural network 2 hidden layers
2. SVM Gaussian kernel
3. Neural network 1 hidden layer
4. SVM Linear kernel
5. k-Nearest Neighbors
6. Naive Bayes
7. LVQ

*Το νευρωνικό δίκτυο υλοποιήθηκε με το keras.

Επίσης στο LVQ οι προβλέψεις για το test dataset, έγινε όπως και η εκπαίδευση, δηλαδή βρίσκοντας την μεγαλύτερη gaussian ομοιότητα.

Ολοι οι μέθοδοι -εκτός από το νευρωνικό δίκτυο και το LVQ- υλοποιήθηκαν με το εργαλείο sklearn και για το “φόρτωμα” των δεδομένων χρησιμοποιήθηκε το pandas ενώ για την δίκαιη κατανομή των δεδομένων σε training και test dataset, χρησιμοποιήθηκε το k-Fold Cross Validation.

SVM linear kernel		
File: spambase	Accuracy score	F1 score
Gamma = “scale”	0.928	0.906
Gamma = “auto”	0.929	0.908

SVM Gaussian kernel		
File: spambase	Accuracy score	F1 score
Gamma = “scale”	0.934	0.913
Gamma = “auto”	0.934	0.914

SVM Linear kernel		
File: credit cards	Accuracy score	F1 score
Gamma = “scale”	0.809	0.355
Gamma = “auto”	0.809	0.357

SVM Gaussian kernel		
File: credit cards	Accuracy score	F1 score
Gamma = “scale”	0.819	0.450
Gamma = “auto”	0.819	0.449

Naive Bayes		
File: spambase	Accuracy score	F1 score
	0.816	0.804

Naive Bayes		
File: credit cards	Accuracy score	F1 score
	0.711	0.499

k-Nearest Neighbors		
File: spambase	Accuracy score	F1 score
k=5	0.906	0.879
k=15	0.903	0.871
k=30	0.897	0.862

k-Nearest Neighbors		
File: credit cards	Accuracy score	F1 score
k=30	0.812	0.421
k=90	0.809	0.395
k=180	0.807	0.375

Neural network- 1 hidden layer	File: spambase	
Units	Accuracy score	F1 score
10	0.930	0.912
20	0.932	0.913
50	0.934	0.915

Neural network- 2 hidden layers	File: spambase	
Units	Accuracy score	F1 score
10-20	0.933	0.914
20-40	0.936	0.918
40-70	0.936	0.920

Neural network- 1 hidden layer	File: credit cards	
Units	Accuracy score	F1 score
60	0.819	0.472
120	0.819	0.474
240	0.819	0.4824
360	0.819	0.473

Neural network- 2 hidden layers	File: credit cards	
Units	Accuracy score	F1 score
60-120	0.820	0.474
120-140	0.819	0.480
240-360	0.820	0.477

LVQ	File: spambase	
	Accuracy score	F1 score
	0.584	0.655
	0.586	0.656

LVQ	File: credit cards	
	Accuracy score	F1 score
	0.391	0.474
	-	-