

Τμήμα Μηχανικών Η/Υ & Πληροφορικής  
Πανεπιστήμιο Ιωαννίνων  
Μεταπτυχιακό μάθημα: Τ06 - Μηχανική Μάθηση

Σκορδά Ελένη Α.Μ.: 374  
Δημητριάδης Σωκράτης Α.Μ.: 359

**Εργασία 1**

Άσκηση 1

Α) Ο ταξινομητής των K-κοντινότερων γειτόνων, χρησιμοποιώντας την Ευκλείδεια απόσταση και έναν έως και δέκα γείτονες, κατηγοριοποίησε τα δεδομένα-χαρακτήρες δίνοντας τα εξής αποτελέσματα:

Πλήθος Γειτόνων : 1  
Απώλεια Μεθόδου = 0.0217  
Ποσοστό Επιτυχίας = 0.9638

Πλήθος Γειτόνων : 2  
Απώλεια Μεθόδου = 0.0256  
Ποσοστό Επιτυχίας = 0.9633

Πλήθος Γειτόνων : 3  
Απώλεια Μεθόδου = 0.0207  
Ποσοστό Επιτυχίας = 0.9649

Πλήθος Γειτόνων : 4  
Απώλεια Μεθόδου = 0.0243  
Ποσοστό Επιτυχίας = 0.9644

Πλήθος Γειτόνων : 5  
Απώλεια Μεθόδου = 0.0262  
Ποσοστό Επιτυχίας = 0.9649

Πλήθος Γειτόνων : 6  
Απώλεια Μεθόδου = 0.0269  
Ποσοστό Επιτυχίας = 0.9638

Πλήθος Γειτόνων : 7  
Απώλεια Μεθόδου = 0.0246  
Ποσοστό Επιτυχίας = 0.9638

Πλήθος Γειτόνων : 8  
Απώλεια Μεθόδου = 0.0235  
Ποσοστό Επιτυχίας = 0.9638

Πλήθος Γειτόνων : 9  
Απώλεια Μεθόδου = 0.0264  
Ποσοστό Επιτυχίας = 0.9622

Πλήθος Γειτόνων : 10  
Απώλεια Μεθόδου = 0.0269  
Ποσοστό Επιτυχίας = 0.9610

Την καλύτερη απόδοση, παρατηρούμε ότι την δίνει το μοντέλο των τριών γειτόνων, καθώς έχει το μέγιστο ποσοστό επιτυχίας (μαζί με αυτό των πέντε γειτόνων) και τη μικρότερη απώλεια.

Ο αντίστοιχος ταξινομητής των K-κοντινότερων γειτόνων που χρησιμοποιεί την απόσταση Hamming και έναν έως και δέκα γείτονες, κατηγοριοποίησε τα δεδομένα-χαρακτήρες, δίνοντας τα εξής αποτελέσματα:

Πλήθος Γειτόνων : 1  
Απώλεια Μεθόδου = 0.1138  
Ποσοστό Επιτυχίας = 0.8581

Πλήθος Γειτόνων : 2  
Απώλεια Μεθόδου = 0.1381  
Ποσοστό Επιτυχίας = 0.8525

Πλήθος Γειτόνων : 3  
Απώλεια Μεθόδου = 0.1067  
Ποσοστό Επιτυχίας = 0.8742

Πλήθος Γειτόνων : 4  
Απώλεια Μεθόδου = 0.1104  
Ποσοστό Επιτυχίας = 0.8826

Πλήθος Γειτόνων : 5  
Απώλεια Μεθόδου = 0.1054  
Ποσοστό Επιτυχίας = 0.8792

Πλήθος Γειτόνων : 6  
Απώλεια Μεθόδου = 0.1031  
Ποσοστό Επιτυχίας = 0.8798

Πλήθος Γειτόνων : 7  
Απώλεια Μεθόδου = 0.1020  
Ποσοστό Επιτυχίας = 0.8792

Πλήθος Γειτόνων : 8  
Απώλεια Μεθόδου = 0.1031  
Ποσοστό Επιτυχίας = 0.8770

Πλήθος Γειτόνων : 9  
Απώλεια Μεθόδου = 0.0999  
Ποσοστό Επιτυχίας = 0.8770

Πλήθος Γειτόνων : 10  
Απώλεια Μεθόδου = 0.1020  
Ποσοστό Επιτυχίας = 0.8776

Παρατηρείται ότι με τέσσερις γείτονες έχουμε μεγαλύτερο ποσοστό επιτυχίας αλλά μεγαλύτερη απώλεια από ότι άλλες με περισσότερους γείτονες.

Εν κατακλείδι, η Ευκλείδεια απόσταση δίνει εμφανώς καλύτερα αποτελέσματα στην ταξινόμηση από την Hamming.

Γ) Στην ταξινόμηση με πολυεπίπεδο νευρωνικό δίκτυο, έγινε χρήση του εργαλείου της Matlab 2016b nprtool όπου εκτελέσαμε δύο πειράματα με τις εξής ρυθμίσεις:

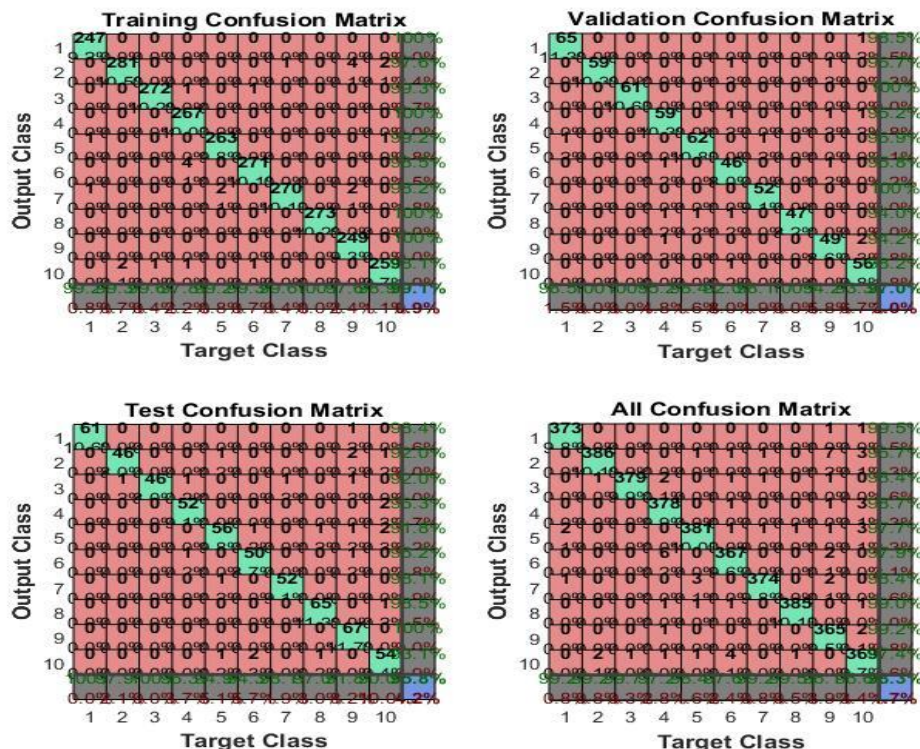
1<sup>η</sup> δοκιμή:

Επιλέχθηκε το 70% των δειγμάτων για εκπαίδευση (2677 δείγματα), 15% για εξακρίβωση (573 δείγματα) και 15% για έλεγχο (573 δείγματα). Επίσης τέθηκαν σε λειτουργία 20 νευρώνες με ένα κρυμμένο επίπεδο για τον σκοπό της ταξινόμησης.

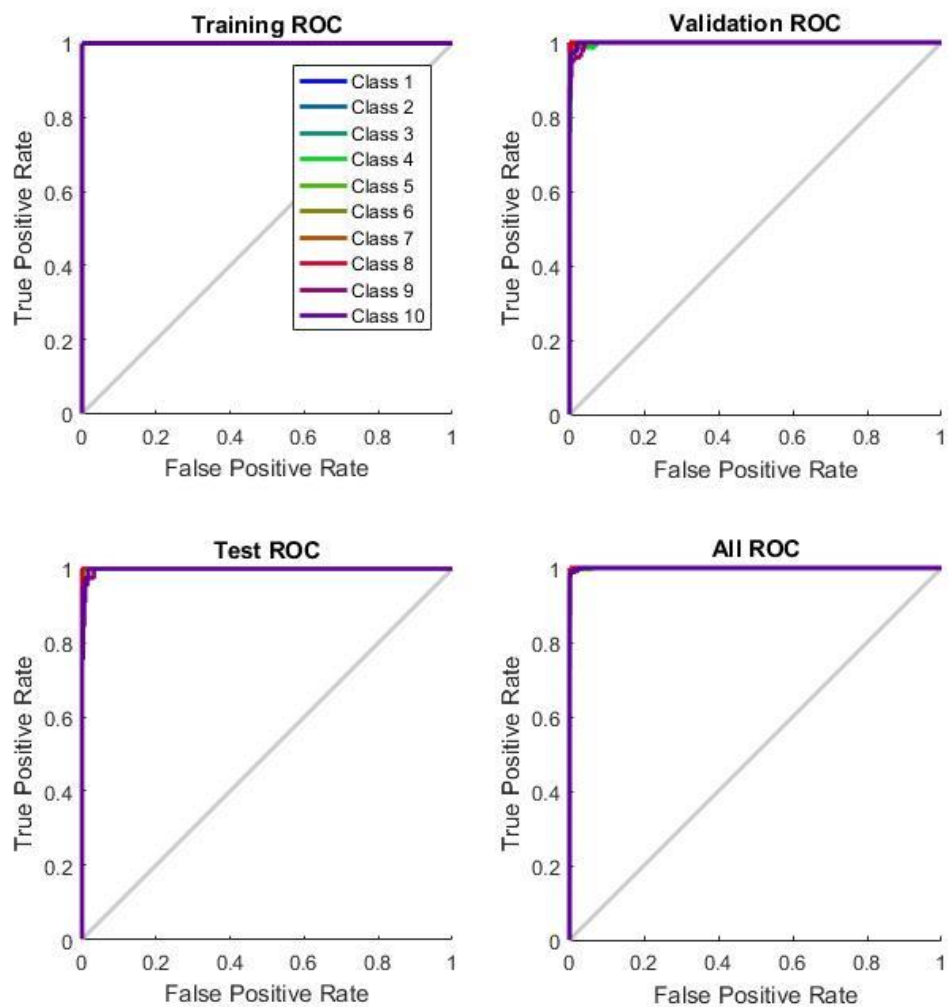
2<sup>η</sup> δοκιμή:

Επιλέχθηκε το 70% των δειγμάτων για εκπαίδευση (2676 δείγματα), 20% για εξακρίβωση (765 δείγματα) και 10% για έλεγχο (382 δείγματα). Επίσης τέθηκαν σε λειτουργία 30 νευρώνες με ένα κρυμμένο επίπεδο για τον σκοπό της ταξινόμησης.

Οι γραφικές παραστάσεις που προέκυψαν είναι αντίστοιχα οι εξής:







Οι μετρήσεις ήταν αντίστοιχα:

	Samples	CE	%E
Training:	2677	2.68139e-0	1.71834e-0
Validation:	573	7.71039e-0	4.18848e-0
Testing:	573	7.68145e-0	2.61780e-0

και

	Samples	CE	%E
Training:	2676	3.12238e-0	5.60538e-1
Validation:	765	8.36850e-0	2.61437e-0
Testing:	382	9.39905e-0	3.40314e-0



Παρατηρείται ότι η δεύτερη δοκιμή έχει μικρότερα ποσοστά σφαλμάτων στην εκπαίδευση και την εξακρίβωση, αλλά μεγαλύτερο στον έλεγχο από την πρώτη.

## Άσκηση 2:

Α) Τα αποτελέσματα της Γραμμικής Παλινδρόμησης είναι:

```
>> linear_regression
```

```
mdl =
```

Linear regression model:

NoShellRings ~ 1 + Sex + Length + Diameter + Height + WWeight + SWeight + VWeight + ShWeight

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	3.95235709880742	0.284838335246169	13.8757906143203	7.85919396400517e-43
Sex_F	-0.0577156749397003	0.0833464543657037	-0.692479066793389	0.488675114179323
Sex_I	-0.882591939744911	0.0957344127698478	-9.2191711862977	4.63789254561716e-20
Length	-0.458335416243381	1.80912456451878	-0.253346521976664	0.800012946959425
Diameter	11.0751025429074	2.22728021581143	4.97247830079276	6.87620027265983e-07
Height	10.7615366991231	1.5362029265219	7.00528329514918	2.86136617862729e-12
WWeight	8.97544462048555	0.725403939172189	12.3730298883241	1.46932031811157e-34
SWeight	-19.7868668634493	0.817350042378369	-24.208559169915	2.98716969766467e-121
VWeight	-10.5818270341597	1.29374885433731	-8.17919722107117	3.75652924505052e-16
ShWeight	8.74180579669376	1.12473147334201	7.77234922636108	9.63862158310486e-15

Number of observations: 4177, Error degrees of freedom: 4167

Root Mean Squared Error: 2.19

R-squared: 0.538, Adjusted R-Squared 0.537

F-statistic vs. constant model: 539, p-value = 0

Δ) Στην MLP με ένα κρυμμένο επίπεδο, χρησιμοποιήθηκαν 30 νευρώνες. Τα αποτελέσματα ήταν τα εξής:

