

ΠΛΗ31

ΕΝΟΤΗΤΑ 3: ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Μάθημα 3.7: WEKA

Δημήτρης Ψούνης



ΠΕΡΙΕΧΟΜΕΝΑ

A.Θεωρία

1. Εισαγωγή

1. Σκοπός
2. Εγκατάσταση

2. Το WEKA στην πράξη

1. Προετοιμασία Αρχείου Δεδομένων
2. Εκπαίδευση με Έλεγχο
3. Αρχική Οθόνη
4. Επιλογή Αρχείου Δεδομένων
5. Παραμετροποίηση Αλγόριθμου Οπισθοδιάδοσης
 1. Αξιολόγηση με Σύνολο Ελέγχου
 2. Αξιολόγηση με Επικύρωση K-δειγμάτων

B.Ασκήσεις

1. Εφαρμογές

A. Θεωρία

1. Εισαγωγή

1. Σκοπός

Το WEKA είναι ένα πρόγραμμα που υλοποιεί ένα σύνολο από αλγόριθμους εξόρυξης γνώσης,



Μεταξύ άλλων υλοποιεί αλγόριθμους εκπαίδευσης νευρωνικών δικτύων όπως ο αλγόριθμος οπισθοδιάδοσης του λάθους.

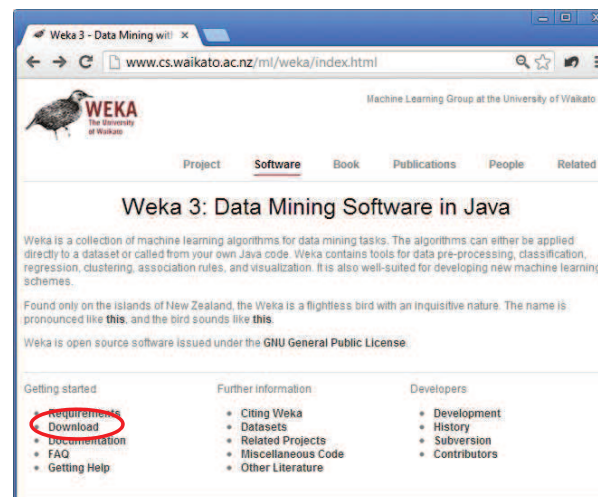
- Σημείωση: Το μάθημα αυτό είναι χρήσιμο για εκπόνηση ερωτημάτων εργασίας.

A. Θεωρία

1. Εισαγωγή

2. Εγκατάσταση

URL εγκατάστασης: <http://www.cs.waikato.ac.nz/ml/weka/>
Επιλογή "Download"



Επεται επιλογή κατάλληλης έκδοσης
"Stable Book 3rd Edition"

x86 – Για 32bit υπολογιστή
x64 – Για 64bit υπολογιστή

A. Θεωρία

2. Το WEKA στην πράξη

1. Προετοιμασία αρχείου δεδομένων

Προετοιμάζουμε το αρχείο δεδομένων μας σε απλό κειμενογράφο (notepad)

@RELATION iris	Ονομασία Αρχείου
@ATTRIBUTE sepallength REAL	Ιδιοχαρακτηριστικά Προτύπων (εδώ 4) και τύπος δεδομένων του κάθε ενός (REAL, INTEGER, BOOLEAN κ.λπ.)
@ATTRIBUTE sepalwidth REAL	
@ATTRIBUTE petallength REAL	
@ATTRIBUTE petalwidth REAL	
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}	Ονομασία κλάσεων στις οποίες χωρίζονται τα πρότυπα.
@DATA 5.1,3.5,1.4,0.2,Iris-setosa 4.9,3.0,1.4,0.2,Iris-setosa 4.7,3.2,1.3,0.2,Iris-setosa	Πρότυπα: Πρώτα τα ονόματα των ιδιοχαρακτηριστικών και επείτα το όνομα της κλάσης που ταξινομείται.

Η επέκταση του αρχείου για να αναγνωρίζεται από το πρόγραμμα είναι .arff

A. Θεωρία

2. Το WEKA στην πράξη

2. Εκπαίδευση με Έλεγχο

Η ακόλουθη διαδικασία συνηθίζεται όταν κατασκευάζουμε ένα νευρωνικό δίκτυο.

Έχουμε ένα σύνολο από δεδομένα του προβλήματος που θα εκπαιδεύσουμε. Τα χωρίζουμε σε δύο σύνολα.

- Ένα σύνολο δεδομένων εκπαίδευσης που αποτελείται από αρκετά πρότυπα. Αυτά θα χρησιμοποιήσουμε για να εκπαιδεύσουμε το δίκτυο μας.
 - Ένα τέτοιο αρχείο είναι το plh31_lesson_3_7_train.arff
- Η αξιολόγηση της επιτυχίας της εκπαίδευσης γίνεται μέσα από το WEKA:
 - Από ένα σύνολο δεδομένων ελέγχου που θα το χρησιμοποιήσουμε για να αξιολογήσουμε το δίκτυο που θα κατασκευαστεί από τα αρχικά δεδομένα.
 - Ένα τέτοιο αρχείο είναι το plh31_lesson_3_7_test.arff
 - Με έτοιμους αλγόριθμους που παρέχει το WEKA

A. Θεωρία

2. Το WEKA στην πράξη

3. Αρχική Οθόνη

Εκπαίδευση Νευρωνικού Δικτύου:



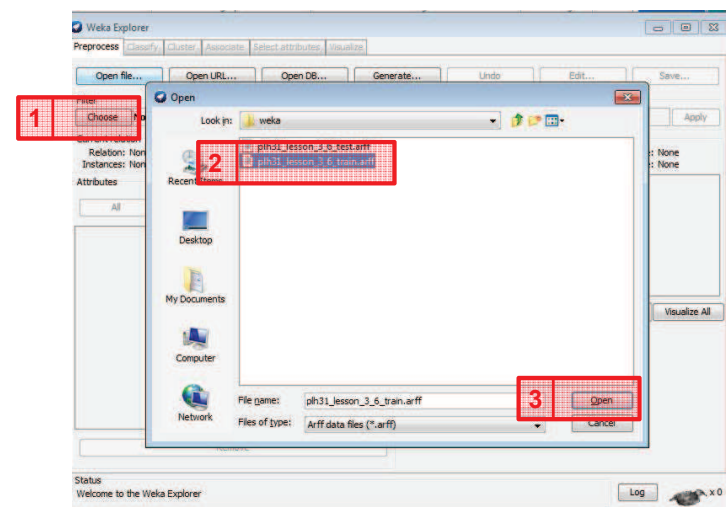
Επιλογή «Explorer»

A. Θεωρία

2. Το WEKA στην πράξη

4. Επιλογή Αρχείου Δεδομένων

Επιλογή «Open File» και άνοιγμα αρχείου εκπαίδευσης

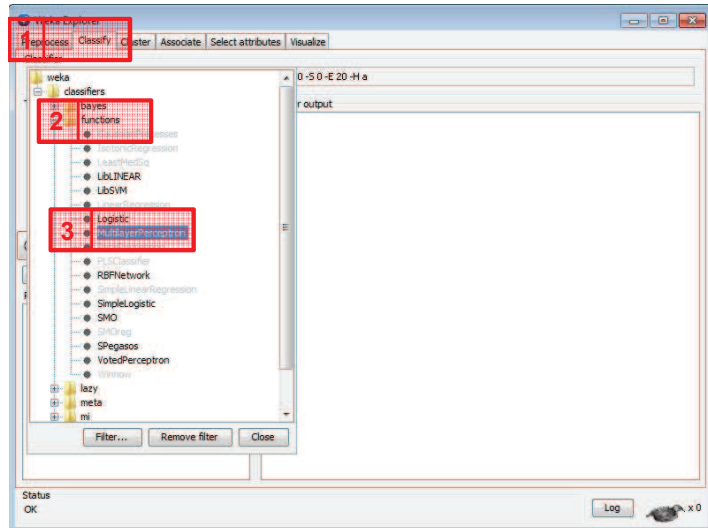


A. Θεωρία

2. Το WEKA στην πράξη

5. Παραμετροποίηση Αλγόριθμου Οπισθοδιάδοσης

Επιλογή του αλγορίθμου οπισθοδιάδοσης του Λάθους



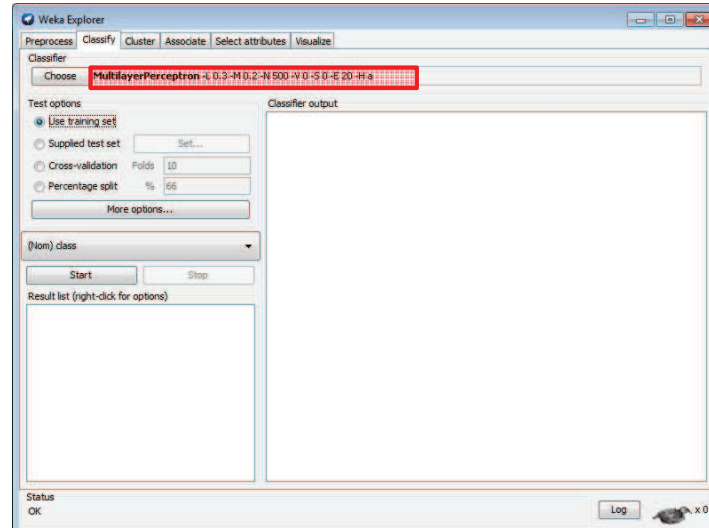
1. Επιλογή του tab "classify"
2. Επιλογή του "functions"
3. Επιλογή του «Multilayer Perceptron»

A. Θεωρία

2. Το WEKA στην πράξη

5. Παραμετροποίηση Αλγόριθμου Οπισθοδιάδοσης

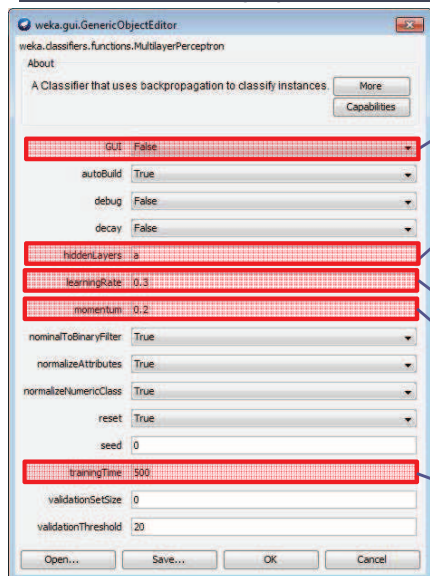
Κλικ στην περιοχή των παραμέτρων



A. Θεωρία

2. Το WEKA στην πράξη

5. Παραμετροποίηση Αλγόριθμου Οπισθοδιάδοσης



Γραφική Απεικόνιση του Νευρωνικού Δικτύου (Συνήθως False)

Κρυφά Επίπεδα
Π.χ.: 2,3,2 έχω 3 κρυφά επίπεδα με 2-3-2 νευρώνες αντίστοιχα
Default τιμή:
 $a = (\text{Νευρώνες_Εισόδου} + \text{Νευρώνες_Εξόδου}) / 2$

Οι εισοδοι και οι εξοδοι καθορίζονται από το αρχείο εκπαίδευσης

- Τα ιδιοχαρακτηριστικά είναι οι εισοδοι
- Έχουμε εξόδους όσες και οι κλάσεις

Τιμή Ρυθμού Εκπαίδευσης

Τιμή Παράγοντα Ορμής

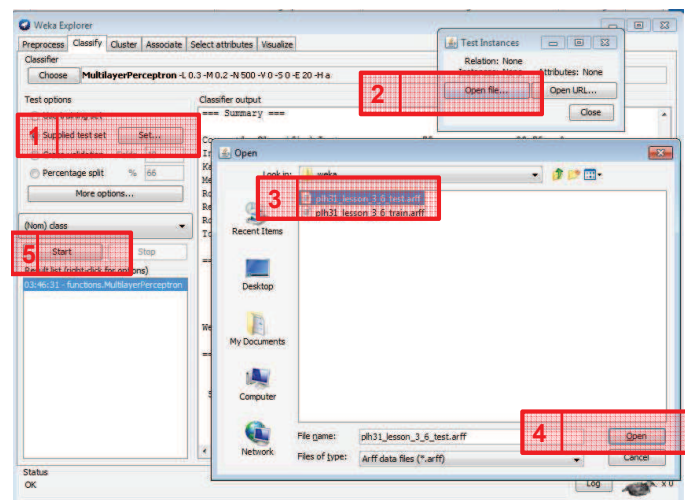
Αριθμός Κύκλων Εκπαίδευσης (Εποχών Εκπαίδευσης)

A. Θεωρία

2. Το WEKA στην πράξη

5α. Αξιολόγηση: Παροχή Συνόλου Ελέγχου

Επιλέγουμε το σύνολο των δεδομένων ελέγχου που έχουμε κατασκευάσει σε ξεχωριστό arff αρχείο.



1. Επιλέγουμε «supplied test set»

2. Επιλέγουμε «Open File»

3. Επιλέγουμε το αρχείο δεδομένων μας.

4. Επιλέγουμε «Open»

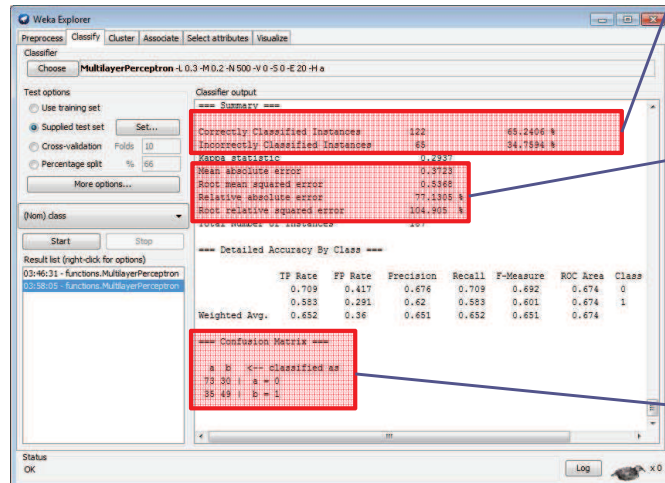
5. Επιλέγουμε «Start»

A. Θεωρία

2. Το WEKA στην πράξη

5α. Αξιολόγηση: Παροχή Συνόλου Ελέγχου

Αποκωδικοποίησης Αποτελεσμάτων



Αποτέλεσμα για το σύνολο ελέγχου:
Εδώ από 187 πρότυπα για έλεγχο:
122 ταξινομήθηκαν σωστά
65 ταξινομήθηκαν λάθος

Στατιστικά Στοιχεία:
Mean Absolute Error = Απόλυτο σφάλμα: Λάθος/(Σωστά+Λάθος)

Mean Absolute Error =
Μέσο Απόλυτο Σφάλμα=

$$\frac{1}{N} \sum_{i=1}^N error(i)$$

Root Mean Squared Error =
Σφάλμα Γενίκευσης=

$$RMSD = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}$$

Confusion Matrix:

Από τα 103 δείγματα A
Τα 73 ταξινομήθηκαν σε A
Τα 30 ταξινομήθηκαν σε B

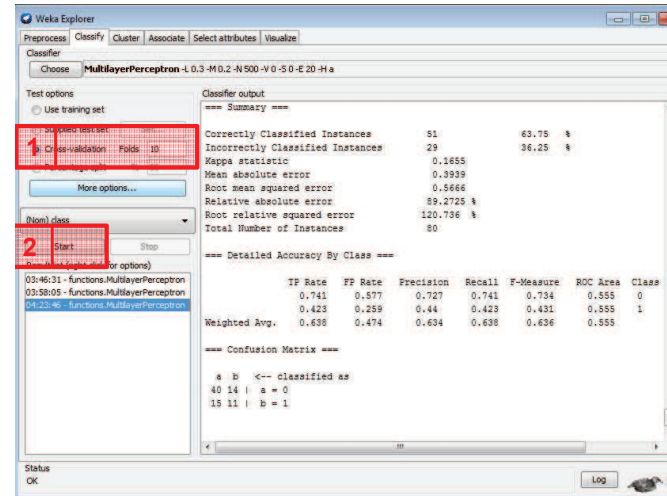
Από τα 80 δείγματα B
Τα 35 ταξινομήθηκαν σε A
Τα 49 ταξινομήθηκαν σε B

A. Θεωρία

2. Το WEKA στην πράξη

5β. Αξιολόγηση: Επικύρωση K-Δειγμάτων

Ενας Δεύτερος Τρόπος Αξιολόγησης του Δικτύου μας χωρίς χρήση δεδομένων ελέγχου είναι η «Επικύρωση K δειγμάτων»



1. Επιλέγουμε «cross-validation»

2. Εισάγουμε το πλήθος των δειγμάτων που θα χρησιμοποιήσουμε.

3. Επιλέγουμε «Start»

A. Θεωρία

2. Το WEKA στην πράξη

5β. Αξιολόγηση: Επικύρωση K-Δειγμάτων

Η Επικύρωση K δειγμάτων (K-cross validation) είναι μια διαδικασία ελέγχου όπου:

- Κάθε ένα από τα υποσύνολα $i=1, \dots, K$:
 - Το υποσύνολο i είναι τα δεδομένα ελέγχου
 - Τα υπόλοιπα υποσύνολα είναι τα δεδομένα εκπαίδευσης.