

ΠΛΗ31

ΕΝΟΤΗΤΑ 3: ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Μάθημα 3.7:
WEKA

Δημήτρης Ψούνης



www.psounis.gr



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.

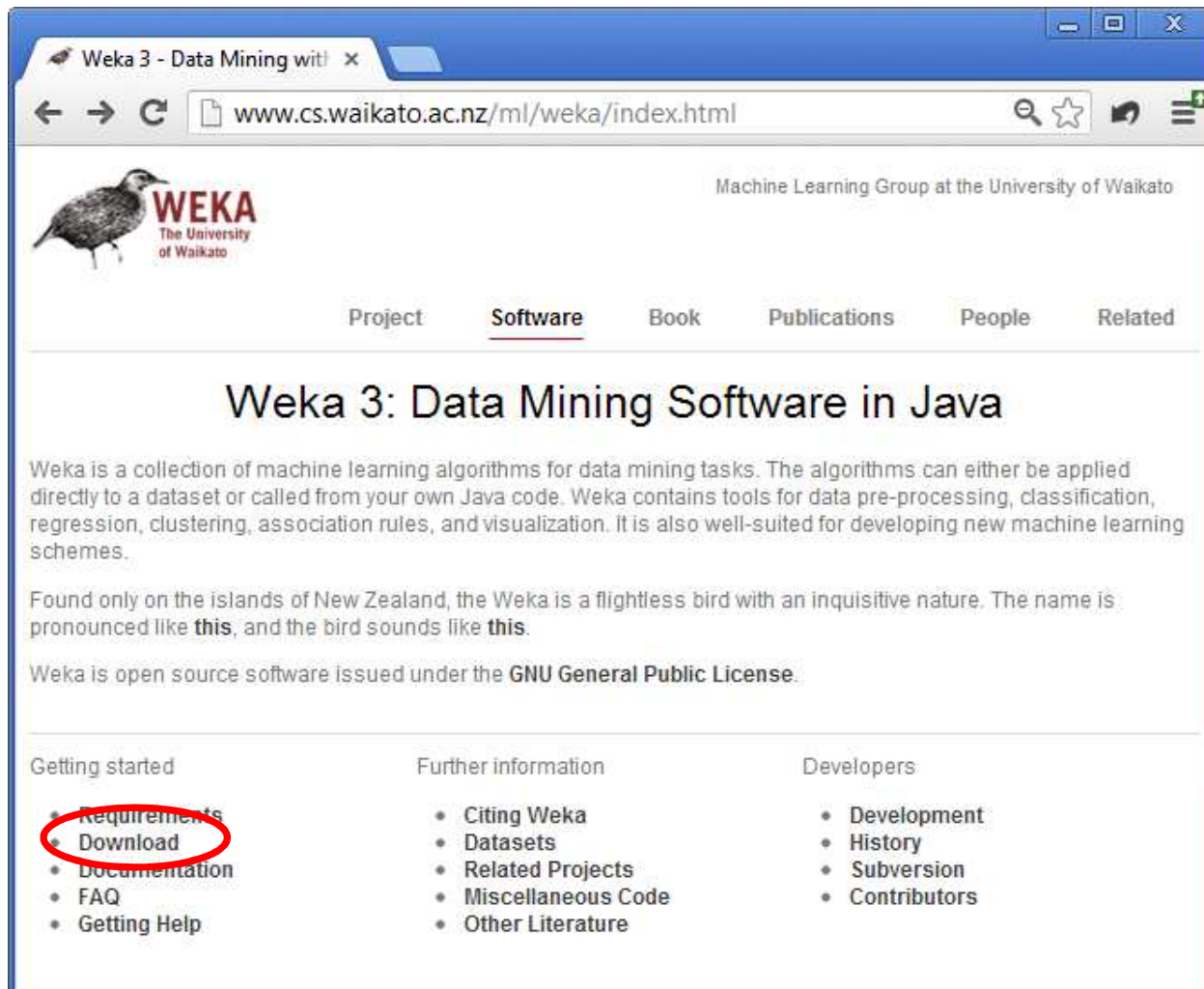
- Σημείωση: Το μάθημα αυτό είναι χρήσιμο για εκπόνηση ερωτημάτων εργασίας.

A. Θεωρία

1. Εισαγωγή

2. Εγκατάσταση

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



Επειτα 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 sepalength 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	} Πρότυπα: Πρώτα τα ονόματα των ιδιοχαρακτηριστικών και επείτα το όνομα της κλάσης που ταξινομείται.
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	
.....	

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»

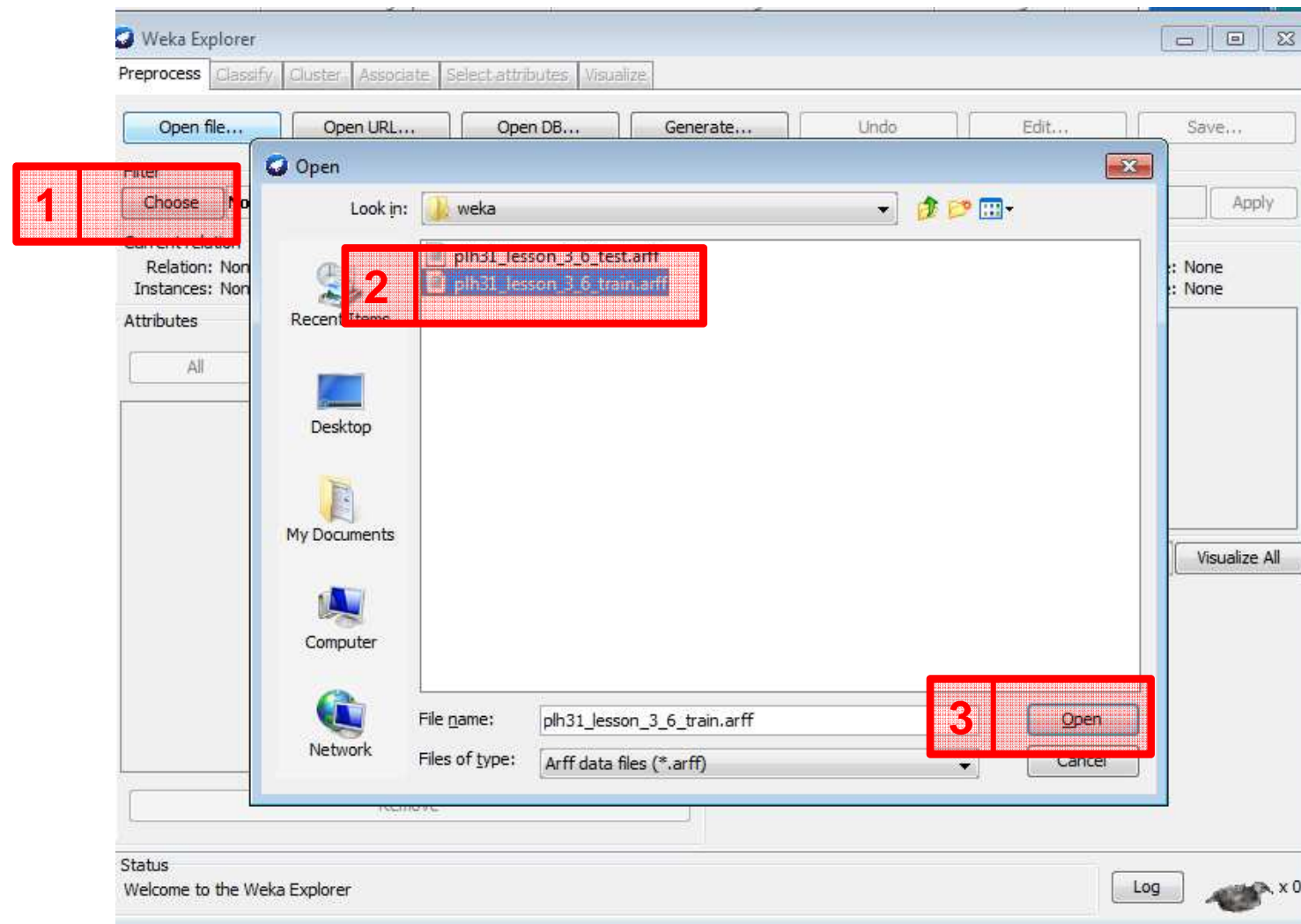


A. Θεωρία

2. Το WEKA στην πράξη

4. Επιλογή Αρχείου Δεδομένων

Επιλογή «Open File» και άνοιγμα αρχείου εκπαίδευσης



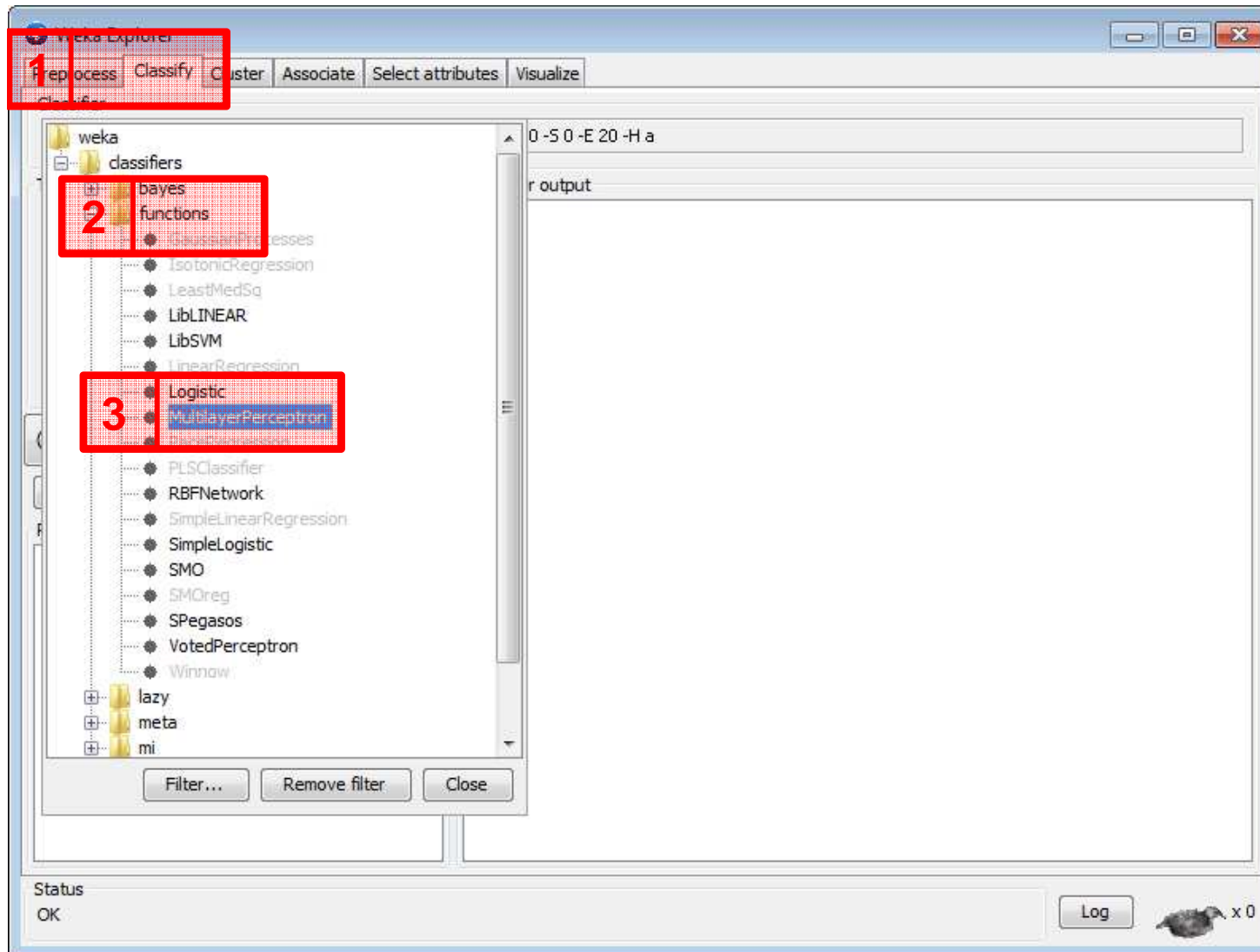


A. Θεωρία

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»

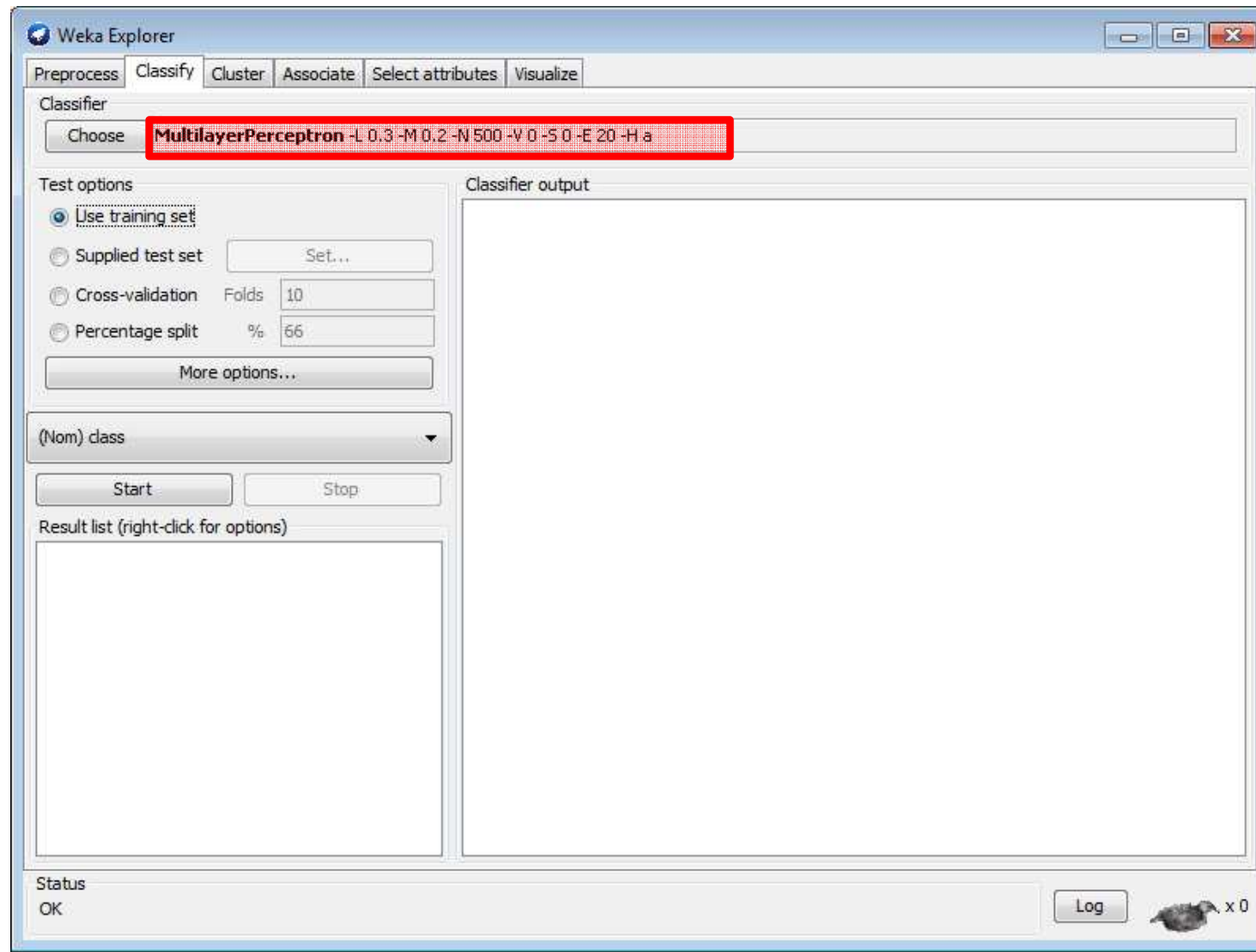


A. Θεωρία

2. Το WEKA στην πράξη

5. Παραμετροποίηση Αλγόριθμου Οπισθοδιάδοσης

Κλικ στην περιοχή των παραμέτρων

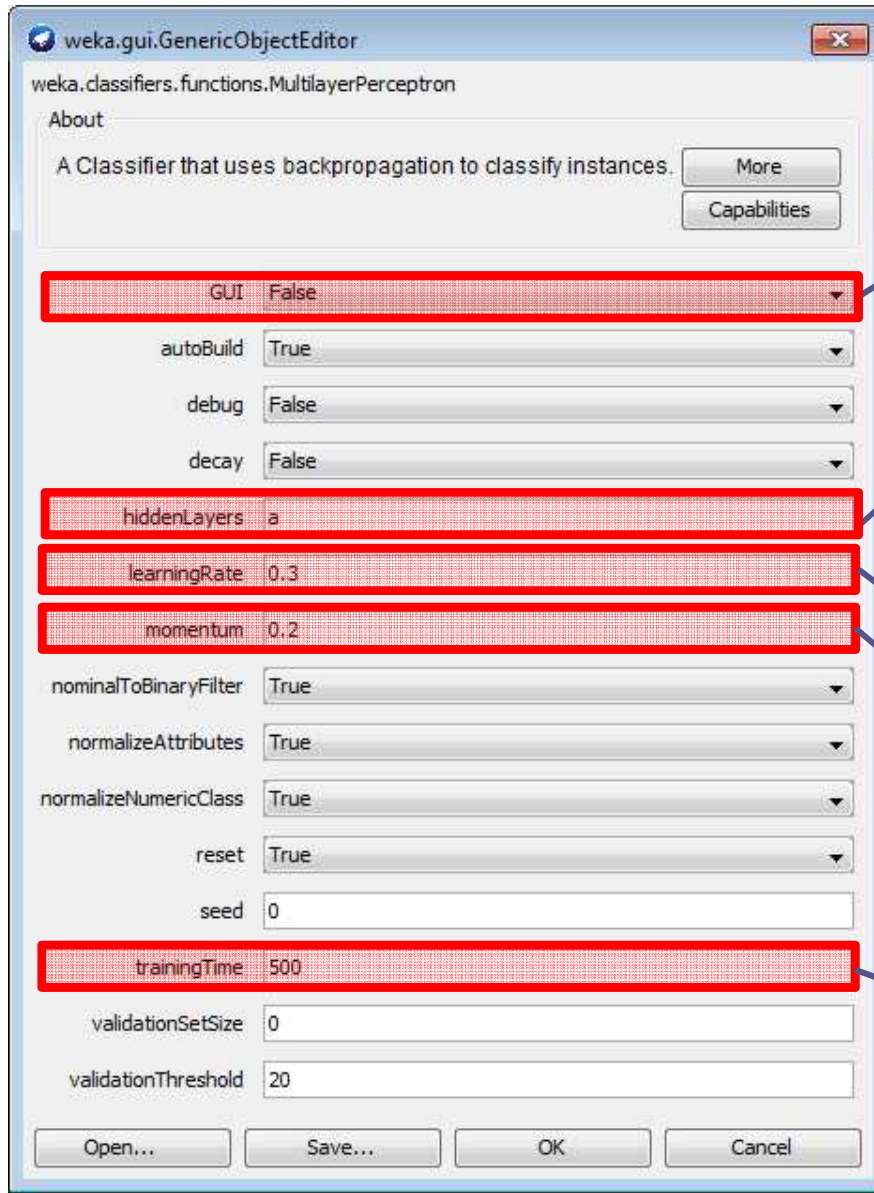




A. Θεωρία

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 τιμη:
$$a = (\text{Νευρώνες_Εισόδου} + \text{Νευρώνες_Εξόδου}) / 2$$

Οι είσοδοι και οι έξοδοι καθορίζονται από το αρχείο εκπαίδευσης

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

Τιμή Ρυθμού Εκπαίδευσης

Τιμή Παράγοντα Ορμής

Αριθμός Κύκλων Εκπαίδευσης (Εποχών Εκπαίδευσης)

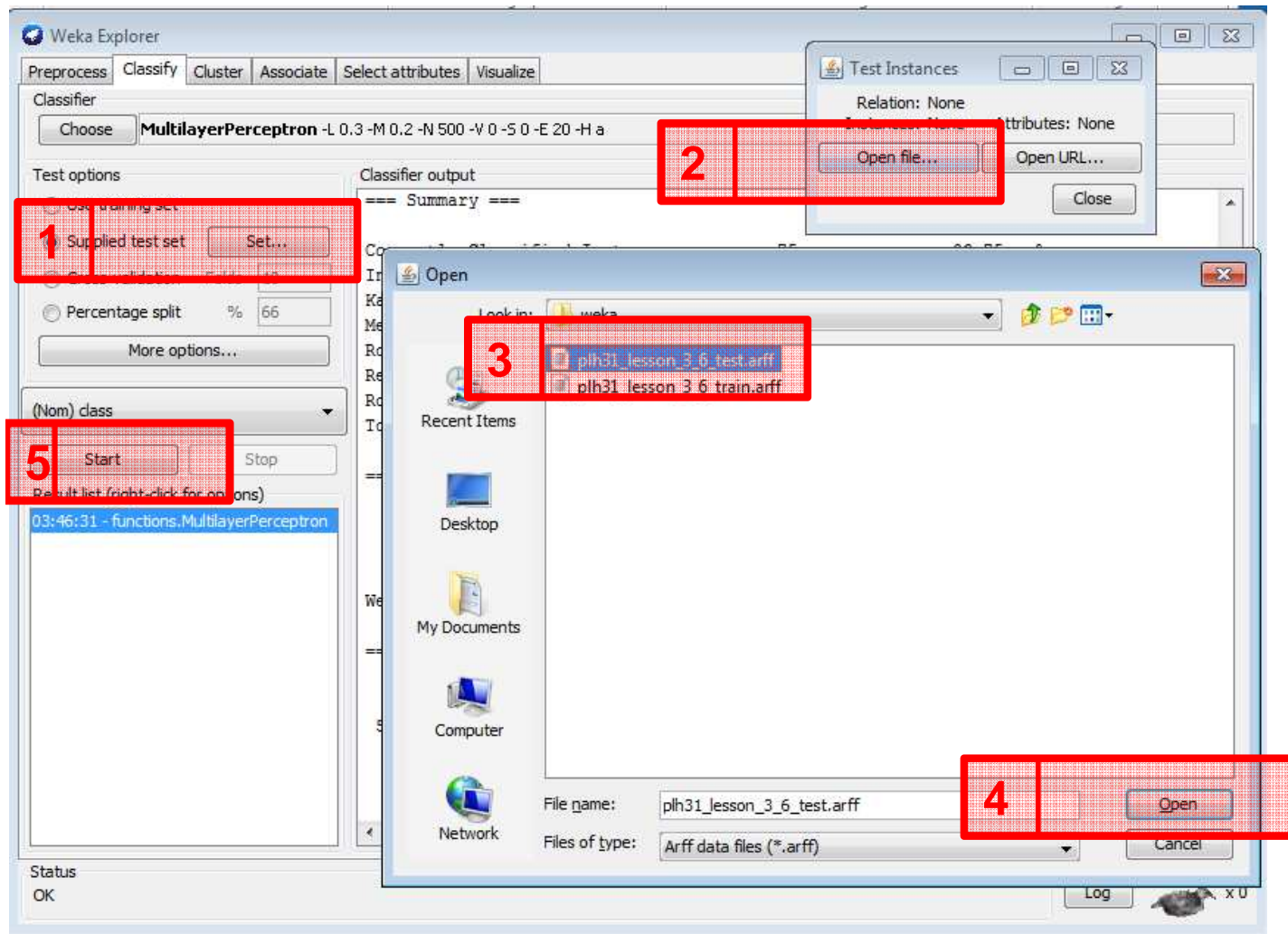


A. Θεωρία

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

Αποκωδικοποίηση Αποτελεσμάτων

Classifier output

=== Summary ===

Correctly Classified Instances	122	65.2406 %
Incorrectly Classified Instances	65	34.7594 %
Kappa statistic	0.2937	
Mean absolute error	0.3723	
Root mean squared error	0.5368	
Relative absolute error	77.1305 %	
Root relative squared error	104.905 %	
Total Number of Instances	187	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.709	0.417	0.676	0.709	0.692	0.674	0
	0.583	0.291	0.62	0.583	0.601	0.674	1
Weighted Avg.	0.652	0.36	0.651	0.652	0.651	0.674	

=== Confusion Matrix ===

a	b	<-- classified as
73	30	a = 0
35	49	b = 1

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»

The screenshot shows the Weka Explorer window with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a'. Under 'Test options', 'Cross-validation' is selected with 'Folds' set to 10. The 'Start' button is highlighted with a red box and the number 2. The 'Classifier output' pane displays the following summary and detailed accuracy by class:

```
=== Summary ===
Correctly Classified Instances      51           63.75 %
Incorrectly Classified Instances   29           36.25 %
Kappa statistic                    0.1655
Mean absolute error                0.3939
Root mean squared error            0.5666
Relative absolute error            89.2725 %
Root relative squared error        120.736 %
Total Number of Instances         80

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.741	0.577	0.727	0.741	0.734	0.555	0
	0.423	0.259	0.44	0.423	0.431	0.555	1
Weighted Avg.	0.638	0.474	0.634	0.638	0.636	0.555	

```

=== Confusion Matrix ===
 a b  <-- classified as
40 14 | a = 0
15 11 | b = 1

```

The 'Status' bar at the bottom shows 'OK' and a 'Log' button.

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.