



**ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΠΣ «ΨΗΦΙΑΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΣΤΗΝ  
ΔΙΟΙΚΗΣΗ ΤΗΣ ΦΙΛΟΞΕΝΙΑΣ ΚΑΙ ΤΟΝ ΤΟΥΡΙΣΜΟ»**

**ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΝΟΗΜΟΣΥΝΗ ΣΤΟΝ ΤΟΥΡΙΣΜΟ**

**ΕΡΓΑΣΤΗΡΙΟ 2: INSTANCE-BASED LEARNING ΣΤΟ WEKA**

[\(\[HTTP://WWW.CS.WAIKATO.AC.NZ/ML/WEKA\]\(http://www.cs.waikato.ac.nz/ml/weka\)\)](http://www.cs.waikato.ac.nz/ml/weka)

Στόχος αυτού του εργαστηρίου είναι

- η εξοικείωση με πειράματα ταξινόμησης στο Weka
- η εξοικείωση με ταξινόμηση βάσει παραδειγμάτων

1. Φορτώστε στο Weka το αρχείο Tanzania.csv.

2. **Επιλογή αλγορίθμου ταξινόμησης:** Επιλέξτε το μενού *Classify*. Στο *Choose* επιλέξτε *classifiers* -> *lazy* -> *IBk*.

Κάντε κλικ πάνω στο όνομα του ταξινομητή, και θα σας ανοίξει ένα καινούριο παράθυρο με τις παραμέτρους του αλγορίθμου. Στην πρώτη παράμετρο (KNN) ορίζετε την τιμή του k (το πλήθος των πλησιέστερων γειτόνων που θα λάβετε υπόψη σας για την απόδοση τιμής εξόδου στο άγνωστο παράδειγμα. Επιβεβαιώστε ότι η τιμή σε αυτή την παράμετρο είναι 1.

3. **Επιλογή τρόπου διεξαγωγής του ελέγχου του πειράματος (test options):**

Στο *Test Options* επιλέγω τον τρόπο με τον οποίο θέλω να ελέγξω την απόδοση του αλγορίθμου.

*Use training set:* Ελέγχεται η απόδοση του αλγορίθμου σε όλα τα δεδομένα εκπαίδευσης (δεν χρησιμοποιούνται καθόλου καινούρια, άγνωστα παραδείγματα) – δεν ενδείκνυται!

*Supplied test set:* Παρέχω ένα αρχείο ελέγχου με καινούρια, άγνωστα παραδείγματα. Το αρχείο ελέγχου έχει ακριβώς την ίδια μορφή με το αρχείο εκπαίδευσης.

*Cross validation:* Χωρίζονται τα δεδομένα του αρχείου δεδομένων σε X ίσα τμήματα. Πραγματοποιούνται X πειράματα (folds), όπου σε κάθε ένα διατηρώ το ένα τμήμα του αρχείου δεδομένων για έλεγχο και τα υπόλοιπα για εκπαίδευση.

*Percentage Split:* Διαιρώ το αρχείο δεδομένων σε δύο τμήματα (δίνοντας ποσοστό): το ένα χρησιμοποιείται για εκπαίδευση και το άλλο για έλεγχο.

Επιλέξτε *Cross Validation* με 10 folds. Πατήστε το *Start*. Στο δεξί τμήμα του παραθύρου θα εμφανιστούν τα αποτελέσματα της ταξινόμησης.

4. **Απόδοση ταξινόμησης:** Παρατηρήστε στο δεξί παράθυρο τις ενδείξεις *Correctly classified instances* και *Incorrectly classified instances*. Η πρώτη ονομάζεται **Accuracy (Ορθότητα)**.

**Η ορθότητα δεν αποτελεί αξιόπιστο μέτρο απόδοσης από μόνη της. Γιατί; \_\_\_\_\_**

---

---

---

5. **Αξιολόγηση των αποτελεσμάτων:** Κάνοντας δεξί κλικ πάνω στο *id3* στο πεδίο *result list* (κάτω αριστερά), εμφανίζονται μια σειρά από ενέργειες που μπορώ να επιλέξω για το μοντέλο ταξινομητή που έφτιαξα. Μπορώ να σώσω τα αποτελέσματά μου ώστε να αποθηκευτούν σε κάποιο αρχείο, μπορώ να σώσω το μοντέλο του ταξινομητή, ώστε να μπορώ να το ξαναχρησιμοποιήσω στο μέλλον για να ταξινομήσω καινούρια παραδείγματα, μπορώ να δω τα λάθη ταξινόμησης.

Επιλέξτε *Visualize classifier errors*. Εμφανίζεται ένα διάγραμμα σε δύο άξονες. Μπορείτε να επιλέξετε οποιοδήποτε από τα χαρακτηριστικά να απεικονίζεται στον ένα ή στον άλλο άξονα, από τα αντίστοιχα πεδία (X και Y). Κάθε τιμή της κλάσης ταξινόμησης έχει το δικό της χρώμα (φαίνεται στο υπόμνημα στο κάτω μέρος του διαγράμματος). Οι σωστές ταξινομήσεις φαίνονται ως σταυροί, οι λάθος φαίνονται ως τετραγωνάκια. Με κλικ πάνω σε ένα παράδειγμα μπορείτε να δείτε πληροφορίες σχετικές με το παράδειγμα (τον αριθμό του, τις τιμές των χαρ/κων του, την πραγματική του κλάση και την κλάση στην οποία ταξινομήθηκε).

6. **Πειράματα για διαφορετικές τιμές του k:** Από το παράθυρο των παραμέτρων του αλγορίθμου επιλέξτε διαφορετικές τιμές του *k*. Τι παρατηρείτε; Ποιά τιμή του *k* οδηγεί στην καλύτερη ορθότητα;

---

---

---

7. **Πειράματα με διαφορετικούς τρόπους αξιολόγησης.** Για την βέλτιστη τιμή του *k* τρέξτε πειράματα με διαφορετική μέθοδο αξιολόγησης (πάνω στο σώμα εκπαίδευσης, 10-fold cross validation, percentage split). Τι παρατηρείτε;

---

---

---

8. **Πειράματα για διαφορετικά σετ χαρακτηριστικών:** Από το μενού *Preprocess* επιλέξτε από το πινακάκι των χαρακτηριστικών κάποιο χαρακτηριστικό, κάντε *tick* στο κουτάκι του και πατήστε από κάτω *Remove*. Αυτό το χαρακτηριστικό δεν θα συμμετέχει από εδώ και πέρα σε όποιο πείραμα ταξινόμησης τρέξετε. Τρέξτε τον *IBk* με την βέλτιστη τιμή του *k* και παρατηρείστε αν η διαγραφή του χαρακτηριστικού επηρεάζει την απόδοση του πειράματός σας.

---

---

---