



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Τεχνητή Νοημοσύνη II

Εργαστηριακή Άσκηση 5

Μουστάκας Κωνσταντίνος

Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών

ΑΝΟΙΚΤΑ ακαδημαϊκά **ΠΠ**
μαθήματα

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΙΙ

Ακαδημαϊκό Έτος 2014-15

5η Εργαστηριακή Άσκηση¹

Καταληκτική Ημερομηνία Υποβολής: 10/6/2015

Μάθηση με Δένδρα Αποφάσεων (Decision Tree Learning)

ΜΕΡΟΣ Α'

Εξοικειωθείτε με το λογισμικό WEKA κάνοντας τα παρακάτω:

1. Εγκαταστήστε το λογισμικό WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>) και πειραματιστείτε με τη χρήση του.
2. Τρέξτε μερικά από τα έτοιμα παραδείγματα που περιγράφονται στο Κεφάλαιο 17 του βιβλίου [4] (Witten et al. 2011).
3. Για περισσότερες λεπτομέρειες ως προς τη χρήση του Explorer του WEKA μπορείτε να δείτε το Κεφάλαιο 11 του ίδιου βιβλίου [4].
4. Για πλήρεις οδηγίες μπορείτε να ανατρέξετε στο εγχειρίδιο του WEKA [3] (Bouckaert et al. 2013, εγκαθίσταται στον υπολογιστή σας μαζί με το λογισμικό). Δείτε ιδιαίτερα τα Κεφάλαια 2 και 4.

Δεν παραδίδετε κάτι για αυτό το μέρος της άσκησης.

ΜΕΡΟΣ Β'

Στο Πανελλήνιο Θερινό Σχολείο Τεχνητής Νοημοσύνης HAISS-2011² υπέβαλαν αιτήσεις συμμετοχής 77 φοιτητές. Τα στοιχεία τους περιέχονται (ανώνυμα) στο αρχείο HAISS2011.xls (φύλλο "Αιτήσεις") που συνοδεύει την εκφώνηση. Η οργανωτική επιτροπή του Θερινού Σχολείου απένειμε υποτροφίες σε 35 από τους συμμετέχοντες, με έναν αλγόριθμο που καλείστε να ανακαλύψετε δημιουργώντας το κατάλληλο δένδρο αποφάσεων από τα δεδομένα.

Οι στήλες του αρχείου HAISS2011.xls περιέχουν τα εξής:

A/A	Ο αύξων αριθμός της αίτησης του κάθε φοιτητή.
ΙΔΙΟΤΗΤΑ	Αν ο φοιτητής είναι προπτυχιακός ή μεταπτυχιακός.
ΜΑΘΗΜΑ TN	Αν ο φοιτητής έχει παρακολουθήσει μάθημα Τεχνητής Νοημοσύνης σε προπτυχιακό επίπεδο.

¹ Έκδοση άσκησης 2.02 (3/6/15). Κ.Σγάμπας.

² <http://technoisi.upatras.gr/haiss2011>

ΕΤΟΣ ΣΠΟΥΔΩΝ	Σε ποιο έτος σπουδών βρίσκεται ο φοιτητής.
ΤΜΗΜΑ	Σε ποιο Τμήμα σπουδάζει ο φοιτητής.
ΙΔΡΥΜΑ	Σε ποιο Πανεπιστήμιο σπουδάζει ο φοιτητής.
ΥΠΟΤΡΟΦΙΑ	ΝΑΙ/ΟΧΙ, ανάλογα αν ο φοιτητής έλαβε υποτροφία. Αυτή είναι η έξοδος (η απόφαση) του δένδρου.

Χρησιμοποιήστε το WEKA για να κατασκευάσετε το δένδρο αποφάσεων από τα παραπάνω δεδομένα. Σημειωτέον ότι το WEKA δεν διαβάζει απευθείας αρχεία Excel. Θα πρέπει να προεπεξεργαστείτε κατάλληλα τα δεδομένα σας πριν τα τροφοδοτήσετε στο WEKA (π.χ. μετατρέψτε τα σε αρχείο .csv ή .arff - δείτε το Κεφάλαιο 9 του εγχειριδίου [3]).

Στην έκθεσή σας:

- Γράψτε τί κάνατε στο στάδιο της προεπεξεργασίας για να περάσετε τα δεδομένα στο WEKA.
- Γράψτε ποιόν αλγόριθμο χρησιμοποιήσατε για την δημιουργία του δένδρου και με ποιές παραμέτρους (options) τον τρέξατε.
- Βάλτε ένα screen capture με το δένδρο αποφάσεων που βγάλατε.
- Προσθέστε περιγραφές και επεξηγήσεις κατά την κρίση σας.

ΜΕΡΟΣ Γ'

Στην πραγματικότητα, οι υποτροφίες στο HAISS-2011 δόθηκαν σύμφωνα με τον παρακάτω αλγόριθμο (τα βήματα εφαρμόστηκαν σειριακά):

1. Όσοι δεν είχαν παρακολουθήσει μάθημα TN ή ήταν άγνωστο αν είχαν παρακολουθήσει μάθημα TN δεν πήραν υποτροφία.
2. Προπτυχιακοί στο τελευταίο έτος σπουδών τους (δηλαδή 4ο-ετείς σε 4ετή τμήματα ή 5ο-ετείς σε 5ετή τμήματα) πήραν υποτροφία.
3. Προπτυχιακοί που είχαν ξεπεράσει το τελευταίο έτος σπουδών τους δεν πήραν υποτροφία.
4. Μεταπτυχιακοί και υποψήφιοι διδάκτορες πήραν υποτροφία.
5. Προπτυχιακοί μικρότερων ετών (1ο-3ο για 4ετή τμήματα, 1ο-4ο για 5ετή τμήματα) δεν πήραν υποτροφία.
6. Μεταδιδακτορικοί ερευνητές και φοιτητές με άγνωστη ιδιότητα δεν πήραν υποτροφία.

Προσθέστε κατάλληλα χαρακτηριστικά (attributes) στον πίνακα με τα δεδομένα σας και ξανατρέξτε τη διαδικασία κατασκευής δένδρου αποφάσεων στο WEKA (για διευκόλυνσή σας, στο φύλλο "Τμήματα" του αρχείου HAISS2011.xls δίνεται η διάρκεια σπουδών των τμημάτων).

Πειραματιστείτε με τις παραμέτρους ώστε να βγάλετε ένα δένδρο αποφάσεων που να μοιάζει όσο το δυνατόν περισσότερο με τον πραγματικό αλγόριθμο.

Στην έκθεσή σας:

- Γράψτε τί κάνατε στο στάδιο της προεπεξεργασίας για να περάσετε τα δεδομένα στο WEKA.
- Γράψτε ποιόν αλγόριθμο χρησιμοποιήσατε για την δημιουργία του δένδρου και με ποιές παραμέτρους (options) τον τρέξατε.
- Βάλτε ένα screen capture με το δένδρο αποφάσεων που βγάλατε.
- Προσθέστε περιγραφές και επεξηγήσεις κατά την κρίση σας.

ΜΕΡΟΣ Δ'

Στο τελικό δένδρο του προηγούμενου μέρους, για κάθε κόμβο A υπολογίστε το **Κέρδος Πληροφορίας (IG = Information Gain)**:

$$\text{Κέρδος}(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \text{Υπόλοιπο}(A)$$

$$\text{όπου: } \text{Υπόλοιπο}(A) = \sum_{i=1}^v \frac{p_i+n_i}{p+n} I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$$

$$\text{και: } I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n} \log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n} \log_2\left(\frac{n}{p+n}\right)$$

Για τη σχετική θεωρία δείτε την ενότητα 18.3 του βιβλίου [1] (Russell & Norvig, 2004) ή/και την ενότητα 18.1.2 στο βιβλίο [2] (Βλαχάβας κ.ά., 2006).

Στην έκθεσή σας:

- Γράψτε τους υπολογισμούς για κάθε κόμβο.
- Προσθέστε περιγραφές και επεξηγήσεις κατά την κρίση σας.
- Προσθέστε ένα screen capture ή διάγραμμα με το δένδρο όπου δίπλα σε κάθε κόμβο έχετε σημειώσει την πληροφορία πριν, την πληροφορία μετά και το κέρδος πληροφορίας που υπολογίσατε (μόνο τα αποτελέσματα).

Βιβλιογραφία

- [1] Stuart Russell, Peter Norvig, "Τεχνητή Νοημοσύνη - Μια Σύγχρονη Προσέγγιση", 2η Αμερικανική Έκδοση, Κλειδάριθμος, 2004.
- [2] Ιωάννης Βλαχάβας, Πέτρος Κεφαλάς, Νικόλαος Βασιλειάδης, Φώτης Κόκκορας, Ηλίας Σακελλαρίου, "Τεχνητή Νοημοσύνη" Γ' Έκδοση, Β. Γκιούρδας Εκδοτική, 2006.
- [3] Remco R. Bouckaert, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann, Alex Seewald, David Scuse, "WEKA Manual", 2013.
- [4] Ian H. Witten, Eibe Frank, Mark A. Hall, "Data Mining - Practical Machine Learning Tools and Techniques", 3rd Edition, Morgan Kaufmann/Elsevier, 2011.

Παραδοτέα

Για την άσκηση αυτή υποβάλετε στο eclass ένα αρχείο zip ή rar που θα περιέχει:

- την έκθεσή σας (κατά προτίμηση σε pdf format)
- τα αρχεία τα .csv ή .arff που φτιάξατε στα Μέρη Β' και Γ' της άσκησης

Προθεσμία υποβολής: 10/6/2015.

ΚΑΛΗ ΕΠΙΤΥΧΙΑ

Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση **1.0**

- Έκδοση **1.0** διαθέσιμη [εδώ](#).

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Πατρων, Φακωτάκης Νικόλαος, Σγάρμπας Κυριάκος, Πέππας Παύλος, Μουστάκας Κωνσταντίνος. «Τεχνητή Νοημοσύνη II, Εργαστήριο 5». Έκδοση: 1.0. Πάτρα 2015.

Διαθέσιμο από τη δικτυακή διεύθυνση:

https://eclass.upatras.gr/modules/course_metadata/opencourses.php?fc=15

Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Διατήρηση Σημειωμάτων

- Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:
- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

