

Machine learning

Practicals

Κώστας Γιαννάκης

kostasg@ionio.gr

Τμ. Πληροφορικής, Ιόνιο Πανεπιστήμιο

Εργαστήριο

- Θα χρησιμοποιήσουμε R και RStudio
- Όσοι και όσες χρησιμοποιούν ήδη Python, θα μπορούν να υλοποιούν και να παραδίδουν εργασίες σε Python
- Θα ακολουθούμε τη θεωρία και θα βλέπουμε στην πράξη μεθόδους και αλγόριθμους Μηχανικής Μάθησης

Φόρμα σχολίων/ανάδρασης/προτάσεων

- ❖ Ανώνυμη φόρμα υποβολής σχολίων/ανάδρασης διδάσκοντα (Κώστα Γιαννάκη)
- ❖ Οποιαδήποτε στιγμή
- ❖ Γράφουμε με σεβασμό και με στόχο να βελτιώσουμε το μάθημα και τον διδάσκοντα
- ❖ <https://forms.office.com/e/908WhJKK5N>



Φόρμα ανώνυμης αξιολόγησης

Μπορείτε να γράψετε ελεύθερα την άποψή σας για το μάθημα, τον διδάσκοντα (Κώστα Γιαννάκη), την ύλη, τα εργαλεία κτλ. Και θετικά και αρνητικά σχόλια είναι καλοδεχούμενα.
Ευπρόσδεκτα ιδιαιτέρως είναι σχόλια που οδηγούν σε άμεση ενέργεια, πχ "Η ροή του λόγου του διδάσκοντα ήταν πολύ γρήγορη, ίσως λίγο πιο αργά θα ήταν καλύτερα", "Θα προτιμούσα λιγότερα/περισσότερα διαλείμματα", "θα ήθελα να εξηγήσεις τον αλγόριθμο Α όπως έκανες με τον αλγόριθμο Β", "άλλαξε ομάδα, η ΑΕΚ καταρρέει" κτλ.

Βοηθάτε έτσι εμένα και εν γένει τα μαθήματα να γίνονται καλύτερα :)

When you submit this form, it will not automatically collect your details like name and email address unless you provide it yourself.

Ευχαριστώ για το ενδιαφέρον!

1. Για ποιό μάθημα; (προαιρετικά)

Enter your answer

2. Αξιολόγηση - Σχόλιο - Πρόταση - Αίτημα

Enter your answer

3. Άλλο

Enter your answer

Submit

Never give out your password. [Report abuse](#)

Απαραίτητες προηγούμενες γνώσεις

- Βασικές έννοιες προγραμματισμού
- Βασικές έννοιες Στατιστικής και Πιθανοτήτων
- Βασικές έννοιες Δομών Δεδομένων
- Προαπαιτούμενο είναι η όρεξη και χρόνος για δουλειά!

Πλάνο εξαμήνου

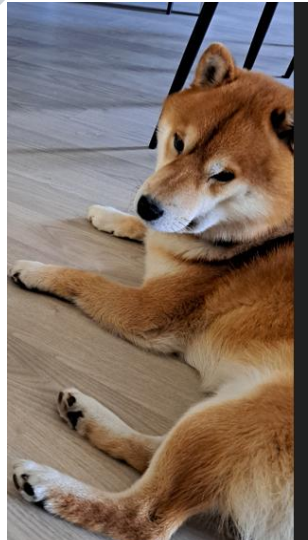
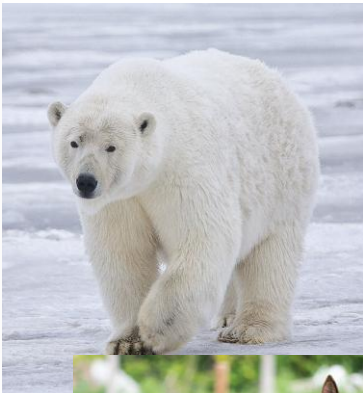
- Εισαγωγή στην R και στο RStudio
 - Εγκατάσταση, περιβάλλον, βασικές εντολές, αποθήκευση/φόρτωση αρχείων
- Εισαγωγή στη MM
 - Γρήγορη ανασκόπηση θεωρίας
 - Παραδείγματα MM
- Στατιστικό και μαθηματικό υπόβαθρο
 - αποστάσεις, κατανομές, πιθανότητες
- Προεπεξεργασία δεδομένων
 - Visualization, missing values κ.ο.κ
- Μέθοδοι και αλγόριθμοι MM
 - Supervised
 - classification - knn, dtree, linear regression, etc
 - Unsupervised
 - § clustering kmeans, hierarchical etc
- Μέτρα αξιολόγησης
 - P_{χ} , RMSE, confusion matrix etc
- Neural networks

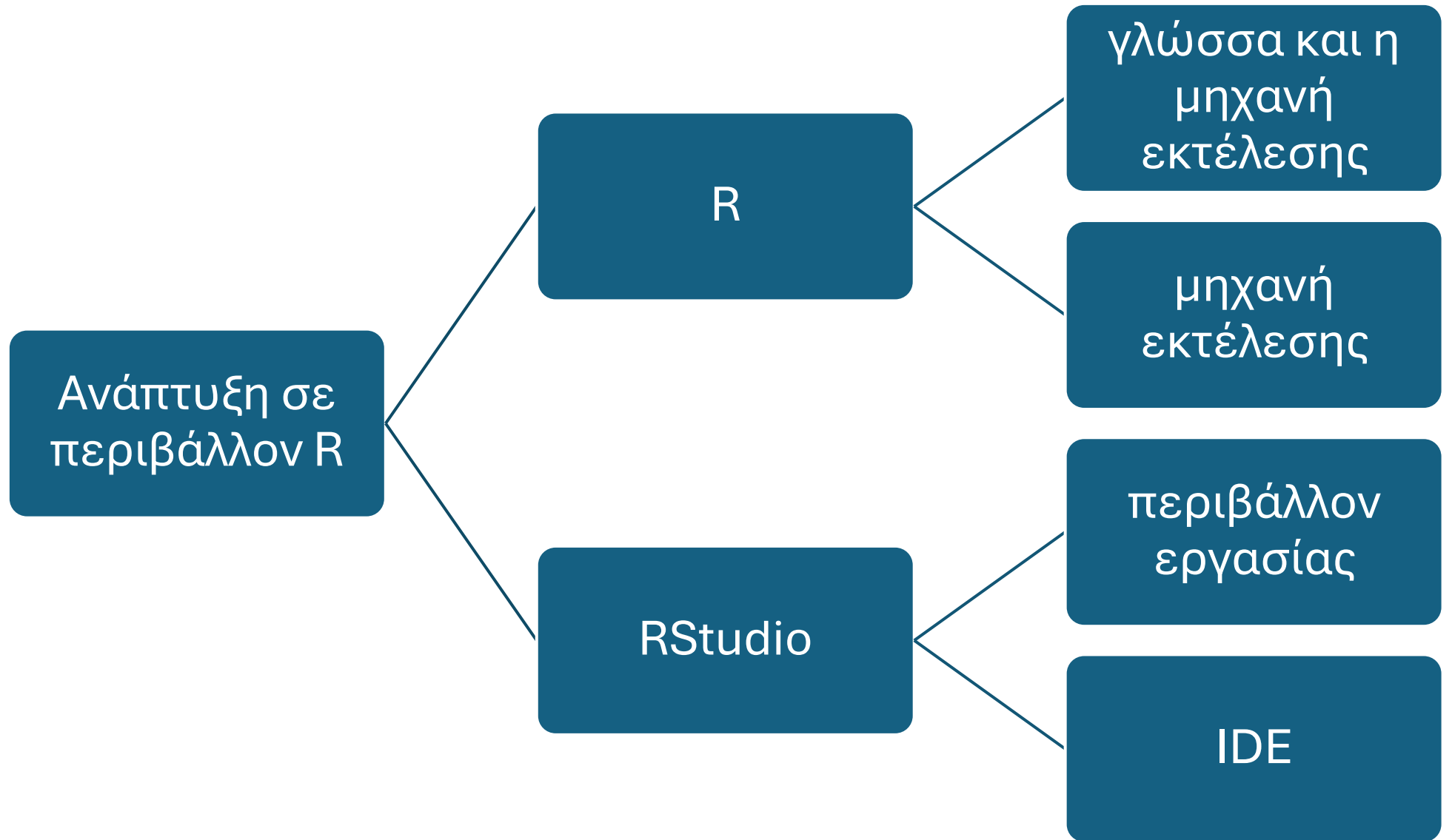
Training

Inference

Testing

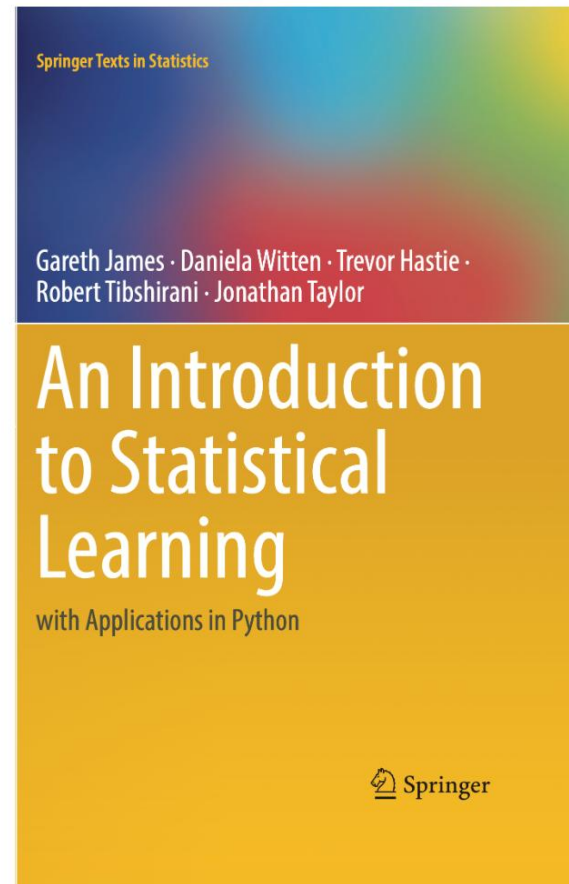
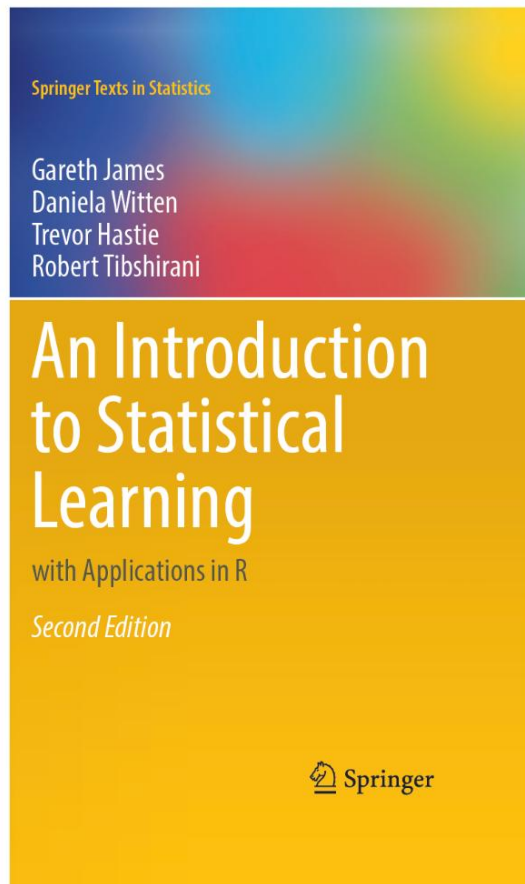
- Μεγάλο μέγεθος
- Μυώδης
- Καφέ, λευκό, ή ασπρόμαυρο
- Μεγάλο τρίχωμα
- Δυνατά σαγόνια
- Ζουν σε δάση ή σε πάγο
- Παμφάγα ή σαρκοφάγα ή φυτοφάγα





Useful resources

- Available for free at <https://www.statlearning.com/>



1	Introduction	1
2	Statistical Learning	15
2.1	What Is Statistical Learning?	15
2.1.1	Why Estimate f ?	17
2.1.2	How Do We Estimate f ?	21
2.1.3	The Trade-Off Between Prediction Accuracy and Model Interpretability	24
2.1.4	Supervised Versus Unsupervised Learning	26
2.1.5	Regression Versus Classification Problems	28
2.2	Assessing Model Accuracy	29
2.2.1	Measuring the Quality of Fit	29
2.2.2	The Bias-Variance Trade-Off	33
2.2.3	The Classification Setting	37
2.3	Lab: Introduction to R	42
2.3.1	Basic Commands	43
2.3.2	Graphics	45
2.3.3	Indexing Data	47
2.3.4	Loading Data	48
2.3.5	Additional Graphical and Numerical Summaries	50
2.4	Exercises	52
3	Linear Regression	59
3.1	Simple Linear Regression	61
3.1.1	Estimating the Coefficients	61
3.1.2	Assessing the Accuracy of the Coefficient Estimates	63
3.1.3	Assessing the Accuracy of the Model	68
3.2	Multiple Linear Regression	71
3.2.1	Estimating the Regression Coefficients	72

Also useful

- <https://link.springer.com/book/10.1007/978-0-387-21736-9>
 - Available for free at <https://www.stat.cmu.edu/~brian/valerie/617-2022/0%20-%20books/2004%20-%20wasserman%20-%20all%20of%20statistics.pdf>

Σήμερα - basic R, plotting and intro to DTs

- Working directory, περιβάλλον RStudio
 - File -> New Script ->
- Μεταβλητές και βασικές πράξεις
- Εκτύπωση μεταβλητών στο terminal
- For loops, if statements
- Δημιουργία vector και βασικές πράξεις
- Δημιουργία matrix
- Δημιουργία dataframe
- Βασικές πράξεις και μετασχηματισμοί dataframe
 - colnames, rownames, rbind, cbind, data types, table
- Βοήθεια στο R
- Τυχαίο δείγμα από Κανονική Κατανομή (συμβολίζεται ως $N(\mu, \sigma^2)$, όπου μ η μέση τιμή και σ^2 η διακύμανση)
- Summary and descriptive statistics
- Οπτικοποίηση και γραφήματα (plotting)
- Διάβασμα αρχείου:
- Εξαγωγή αρχείου
- Αποθήκευση (όχι εξαγωγή!) μεταβλητών
- Seed (ή αλλιώς πώς 'ελέγχουμε' την τυχαιότητα)
- **Clustering**



Συνεχίζουμε από το προηγούμενο εργαστήριο

- Και μετά θα κάνουμε γρήγορη εισαγωγή στο clustering και εφαρμογή του k-means
 - Θα χρησιμοποιήσουμε συναρτήσεις από διάφορες βιβλιοθήκες

Ανοίγουμε με τη σειρά τα αρχεία:

1. lab_1.txt και external_file.csv από το προηγούμενο μάθημα
 - Αρχικός κατάλογος » Εργαστήριο » Εργαστήριο 2
2. R_preprocessing.pdf και lab_2.txt
3. R_intro_DTs.pdf και lab_3.txt
4. lab_3_additional.txt