

**Τμήμα Μηχανικών Η/Υ & Πληροφορικής,  
Πανεπιστήμιο Ιωαννίνων**

**Μεταπτυχιακό μάθημα: «Μηχανική Μάθηση»**

(Ημερομηνία παράδοσης: έως 13/5/2019)

**Άσκηση 1 : Πρόβλημα Ταξινόμησης**

Κατεβάστε το αρχείο **Digits.rar** ([www.cs.uoi.gr/~arly/courses/ml/Digits.rar](http://www.cs.uoi.gr/~arly/courses/ml/Digits.rar)) το οποίο αφορά ένα σύνολο παραδειγμάτων χειρόγραφων χαρακτήρων - ψηφία από 0 έως 9. Κάθε παράδειγμα αναφέρεται σε μία εικόνα μεγέθους [8x8] *pixels*, δηλ. είναι ένα διάνυσμα **64** χαρακτηριστικών, κάθε ένα από τα οποία αντιστοιχούν στην φωτεινότητα ενός pixel της εικόνας. (Να σημειωθεί ότι η φωτεινότητα έχει διακριτοποιηθεί σε 16 στάθμες, δηλ. είναι μία ακέραια τιμή μεταξύ 0 και 16).

Το αρχείο αποτελείται από σύνολο εκπαίδευσης (trainImages, trainLabels) αποτελούμενο από 3823 πρότυπα εκπαίδευσης και σύνολο ελέγχου (testImages, testLabels) αποτελούμενο από 1797 πρότυπα για έλεγχο και αξιολόγηση των μεθόδων.

Στόχος της Άσκησης είναι η ανάπτυξη ενός **συστήματος αυτόματης αναγνώρισης χειρόγραφων αριθμητικών ψηφίων** (Optical Character Recognition – **OCR**).

Χρησιμοποιώντας το σύνολο εκπαίδευσης να εξετάσετε τις παρακάτω μεθόδους ταξινόμησης:

- **Multi-class SVM Classifier** με *linear* και *RBF kernel function*. Μπορείτε να χρησιμοποιήσετε το περιβάλλον της **libsvm** (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>) η οποία υποστηρίζει αυτόματα περιπτώσεις πολλών (>2) κατηγοριών (*multi class*).
- Πολυεπίπεδα Νευρωνικά Δίκτυα (**MLPs**) με 1 κρυμμένο επίπεδο.
- Deep neural networks (stacked autoencoders) με 3 κρυμμένα επίπεδα.

Για κάθε μέθοδο:

- 1) Να βρείτε τις καλύτερες τιμές παραμέτρων χρησιμοποιώντας 10-fold cross-validation στο σύνολο εκπαίδευσης.
- 2) Για τις ανωτέρω τιμές παραμέτρων, να εκπαιδεύσετε ένα ταξινομητή σε όλο το σύνολο εκπαίδευσης.
- 3) Να υπολογίσετε την ακρίβεια του ταξινομητή που προκύπτει στο σύνολο ελέγχου.

**Άσκηση 2 : Πρόβλημα Παλινδρόμησης**

Κατεβάστε από την βάση δεδομένων **UCI Machine Learning Repository** (<http://archive.ics.uci.edu/ml/>) το σύνολο δεδομένων ‘**abalone.data**’ το οποίο αφορά σε πρόβλημα παλινδρόμησης (πρόβλεψης).

Θα πρέπει να κατασκευάσετε συστήματα παλινδρόμησης και να συγκρίνετε τις παρακάτω μεθόδους παλινδρόμησης χρησιμοποιώντας 10-fold cross-validation:

- **Linear Regression** model,
- **Polynomial Regression** χρησιμοποιώντας πολυώνυμο βαθμού [2-10],
- Μέθοδος **lasso** με διάφορες τιμές της regularization parameter  $\lambda$ ,
- Πολυεπίπεδα Νευρωνικά Δίκτυα (**MLPs**) με 1 κρυμμένο επίπεδο.
- **Gaussian Processes** θεωρώντας είτε *linear kernel*, είτε *Gaussian kernel* χρησιμοποιώντας ένα διάστημα τιμών της παραμέτρου  $\sigma$  για την αυτόματη αναζήτηση της βέλτιστης τιμής σε αυτό το διάστημα.