

# ΕΡΓΑΣΙΑ ΜΕ FOLDING

Διδάσκων: Ιωάννης Γ. Τσούλος

2022

## 1 ΟΔΗΓΙΕΣ

1. Κατεβάστε από το uci (<https://archive.ics.uci.edu/ml/index.php>) 4 dataset της επιλογής σας. Τα δύο αρχεία να είναι για μάθηση συναρτήσεων ( regression datasets) και 2 αρχεία για κατηγοριοποίηση δεδομένων (classification datasets).
2. Σαν μέθοδος αρχικοποίησης να χρησιμοποιηθεί η αρχικοποίηση σε πολύ χαμηλές τιμές πχ στο διάστημα  $[-0.01,0.01]$
3. Υλοποιήστε τις ακόλουθες τεχνικές Folding
  - (a) N - Folding. Σε αυτήν την περίπτωση το αρχικό σύνολο δεδομένων διαμερίζεται σε N ζευγάρια. Σε κάθε ζευγάρι το ένα μέλος είναι το TRAIN SET και το άλλο είναι το TEST SET. Για παράδειγμα για 10-Folding σε κάθε ζευγάρι το 10% θα είναι το TEST SET και το 90% θα είναι το TRAIN SET.
  - (b) Leave one out. Σε αυτήν την περίπτωση το 1 πρότυπο αποτελεί το TEST SET και όλο το υπόλοιπο DATASET είναι το TRAIN SET. Αν για παράδειγμα έχουμε ένα dataset με 50 πρότυπα θα δημιουργηθούν 50 συνδυασμοί και σε κάθε συνδυασμό τα 49 πρότυπα θα είναι το TRAIN SET και αυτό που απομένει θα είναι το TEST SET.
4. Σαν μέθοδος εκπαίδευσης να χρησιμοποιηθεί η Gradient Descent με ρυθμό μάθησης που υπολογίζεται με γραμμική αναζήτηση.
5. Χρησιμοποιήστε τις ακόλουθες τιμές για το N: 2,5,10
6. Καταγράψτε τα σφάλματα ελέγχου και κατηγοριοποίησης για κάθε μια περίπτωση καθώς και τα συμπεράσματά σας.

## 2 ΓΕΝΙΚΑ

1. Γλώσσα υλοποίησης μπορεί να είναι οποιαδήποτε, όπως C++, JAVA, PYTHON κτλ.

2. Θεωρούμε πως τα σύνολα δεδομένων έχουν χωριστεί περίπου στην μέση. Τα μισά από αυτά αποτελούν το TRAIN SET και τα υπόλοιπα το TEST SET.
3. Κάθε τεστ πρέπει να γίνει 30 φορές και να πάρετε μέσο όρο για το TRAIN ERROR, TEST ERROR και CLASSIFICATION ERROR. Σε κάθε εκτέλεση χρησιμοποιήστε διαφορετική γεννήτρια τυχαίων αριθμών (διαφορετική παράμετρο στην κλήση της srand).
4. Πρέπει να παραδώσετε ένα συμπιεσμένο αρχείο που θα περιέχει τα συμπεράσματά σας και τον κώδικα που χρησιμοποιήσατε μέχρι τις 10/1/2023 στο email: itsoulos@uoi.gr
5. Θα πρέπει να παρουσιάσετε είτε δημόσια στην τάξη είτε μέσω MsTeams την εργασία σας.
6. Η εργασία αυτή καταλαμβάνει το 70% του βαθμού σας στην Μηχανική Μάθηση.
7. Για οποιαδήποτε απορία ή διευκρίνιση μπορείτε να με βρίσκετε Δευτέρα και Τρίτη δια ζώσης 12:00 - 16:00 και δικτυακά μέσω MsTeams