

Άσκηση στην κατηγοριοποίηση δεδομένων

Ιωάννης Γ. Τσούλος

Ημερομηνία παράδοσης: 21/11/2022

1 Εκτίμηση απόδοσης KNN

1. Κατεβάστε από την ενότητα **ΠΡΟΓΡΑΜΜΑΤΑ** το συμπιεσμένο αρχείο “Δεδομένα για το δίκτυο MLP” και αποσυμπίστε το. Αυτό θα δημιουργήσει τον φάκελο MLPDATA.
2. Να δοκιμάσετε την απόδοση του αλγορίθμου KNN στα προβλήματα που υπάρχουν στον φάκελο MLPDATA χρησιμοποιώντας διάφορες τιμές της παραμέτρου K. Να αναφέρετε το λάθος και την τιμή του K και αν παρατηρήσατε σε κάποιο πρόβλημα σημαντική αλλαγή και για ποια τιμή του K.
3. Η μέθοδος KNN χρησιμοποιεί Ευκλείδεια απόσταση για τον υπολογισμό αποστάσεων. Αλλάξτε την τεχνική απόστασης σε κάποια άλλη από την βιβλιογραφία και αναφέρετε την πιθανή επίδραση που παρατηρήσατε στα προβλήματα του φακέλου MLPDATA.
4. Πέρα από τα αρχεία που βρίσκονται στον φάκελο MLPDATA δοκιμάστε την επιτυχία της μεθόδου KNN σε 2 προβλήματα της επιλογής σας που βρίσκονται στην ενότητα <https://archive.ics.uci.edu/ml/datasets.php?format=&task=cla&att=&area=&numAtt=&numIns=&type=&sort=nameUp&view=table> (είναι δεδομένα κατηγοριοποίησης)

2 Κανονικοποίηση δεδομένων

1. Επαναλάβετε τα ερωτήματα της προηγούμενης εργασίας αφού κανονικοποιήσετε τα δεδομένα με Min-Max. Καταγράψτε τις διαφοροποιήσεις που παρατηρήσατε.
2. Προσθέστε στην κατηγορία Dataset την μέθοδο `normalizeZscore()` με την οποία τα δεδομένα θα κανονικοποιούνται με βάση την κανονικοποίηση Z-Score. Κανονικοποιήστε τα δεδομένα του προηγούμενου ερωτήματος και καταγράψτε τις πιθανές διαφοροποιήσεις που θα παρατηρήσετε.

3 Μάθηση συνάρτησης

1. Δημιουργήστε 200 τυχαία σημεία της συνάρτησης $f(x) = x \sin(x^2)$ στο διάστημα $[-4,4]$.
2. Εφαρμόστε την μέθοδο KNN για μάθηση της συνάρτησης με 2,5,10 γείτονες
3. Για κάθε περίπτωση γειτόνων από τις 2,5,10 κάνετε ένα γράφημα των πραγματικών δεδομένων σε σχέση με αυτό που μαθαίνει ο KNN αλγόριθμος.

4 Παρατηρήσεις - Βιβλιογραφία

1. Οι εργασίες είναι ατομικές.
2. Η αποστολή θα γίνει στο email itsoulos@uoi.gr
3. Ιστοσελίδες με υλικό:
 - (a) https://www.w3schools.com/python/python_ml_knn.asp
 - (b) https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm
 - (c) https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm
 - (d) https://eclass.emt.ihu.gr/modules/document/file.php/MS-C-TIE197/%CE%A7%CE%91%CE%9D%CE%99%CE%91%CE%A3%20./HANIAS_DATA_MINING3.pdf