

3^η ΥΠΟΧΡΕΩΤΙΚΗ ΕΡΓΑΣΙΑ ΣΤΟ ΜΑΘΗΜΑ «ΥΠΟΛΟΓΙΣΤΙΚΗ ΝΟΗΜΟΣΥΝΗ – ΣΤΑΤΙΣΤΙΚΗ ΜΑΘΗΣΗ»

Να γραφεί πρόγραμμα σε οποιαδήποτε γλώσσα προγραμματισμού το οποίο να υλοποιεί μείωση διάστασης δεδομένων βασισμένη σε φασματική ανάλυση γράφου (graph embedding, LPP, LLE, IsoMAP, etc) **στις 2 διαστάσεις** και ακολούθως ομαδοποίηση με spectral clustering.

Οι μέθοδοι αυτές θα εκπαιδευτούν, θα δοκιμαστούν και θα συγκριθούν σε δύο διαφορετικά προβλήματα (βάσεις δεδομένων) επιλογής σας από τα παρακάτω:

Βάση Δεδομένων

Για την εργασία μπορεί να χρησιμοποιηθούν δύο εκ των παρακάτω βάσεων

A. η βάση δεδομένων MNIST ή Cifar-10 ή SVHN που υπάρχουν στις παρακάτω διευθύνσεις:

<http://yann.lecun.com/exdb/mnist/>

<https://www.cs.toronto.edu/~kriz/cifar.html>

<http://ufldl.stanford.edu/housenumbers/>

B. Εναλλακτικά μπορεί να χρησιμοποιηθεί οποιαδήποτε από τις βάσεις δεδομένων που υπάρχουν στις ιστοσελίδες:

<http://www.cs.toronto.edu/~roweis/data.html>

<http://www.cs.cmu.edu/~cil/v-images.html>

<https://www.kaggle.com/datasets>

Για την ομαδοποίηση μας ενδιαφέρει να δούμε για διαφορετικό αριθμό ομάδων την συμπεριφορά του αλγορίθμου. Δηλαδή αν καταφέρνει να βρει ομάδες οι οποίες να αποτελούνται από στοιχεία μιας κλάσης. Ακολουθώς μπορείτε να κατηγοριοποιήσετε τα δεδομένα ελέγχου με βάση τα κέντρα που βρήκατε από την φάση της ομαδοποίησης.

Έκθεση αποτελεσμάτων

Θα πρέπει να γραφεί έκθεση στην οποία να περιγράφονται: ο αλγόριθμος, να δίνονται χαρακτηριστικά παραδείγματα ομαδοποίησης και ποσοστά επιτυχίας στην ομαδοποίηση.

ΗΜΕΡΟΜΗΝΙΑ ΠΑΡΑΔΟΣΗΣ

17^η Ιανουαρίου 2021

Για κάθε ημέρα αργοπορημένης υποβολής της εργασίας και για πέντε ημέρες μειώνεται η βαθμολογία κατά 10%. Μετά από την παράδοση της εργασίας θα ακολουθήσει προφορική εξέταση πάνω στην εργασία, στην οποία θα περιλαμβάνεται **και προφορική εξέταση του κώδικα**.

<p>Να ετοιμάσετε παρουσιάσεις των 3 εργασιών σας οι οποίες θα πραγματοποιηθούν την τελευταία εβδομάδα διαλέξεων.</p>

e-mail: tefas@ csd.auth.gr, Τηλ. 2310-991932