

Εργασία Εαρινού Εξαμήνου 2023-2024

- Ημερομηνία παράδοσης: Ημερομηνία εξέτασης του μαθήματος, ώρα 23:59μμ. Η εργασία συμμετέχει με βάρος 40% στον τελικό βαθμό.

- Η εργασία είναι **ατομική** και παραδίδεται μέσω της πλατφόρμας e-class. Αποδεκτές γλώσσες είναι οι Matlab και Python. Στα παραδοτέα συμπεριλαμβάνονται:

α) η τεκμηρίωση της εργασίας σε αρχείο pdf, στην πρώτη σελίδα της οποίας αναγράφεται το ονοματεπώνυμο του φοιτητή/φοιτήτριας και ο ΑΜ του/της. Θα μηδενιστούν οι εργασίες που δεν περιέχουν τεκμηρίωση ή στοιχεία φοιτητή/φοιτήτριας.

β) τα αρχεία source code σε ένα συμπιεσμένο αρχείο με όνομα source2023.zip (ή .rar ή άλλη σχετική κατάληξη).

γ) οποιαδήποτε άλλα συνοδευτικά αρχεία κρίνεται απαραίτητα σε ένα συμπιεσμένο αρχείο με το όνομα auxiliary2023.zip (ή .rar ή άλλη σχετική κατάληξη).

- Η εργασία θα είναι η ίδια και τον Σεπτέμβριο.

- Η αντιγραφή ή η χρήση generative bots οδηγεί σε μηδενισμό.

A) Καλείστε να υλοποιήσετε ένα σύστημα που προχωρά στην κατάτμηση μιας πρότασης σε λέξεις, χρησιμοποιώντας **υποχρεωτικά** έναν ταξινομητή background vs foreground της επιλογής σας. Δηλαδή, δοθείσης μιας ηχογράφησης ενός ομιλητή, το σύστημα επιστρέφει τα χρονικά όρια των λέξεων που ειπώθηκαν (σε δευτερόλεπτα). Επίσης, παρέχετε συνοδευτικό πρόγραμμα το οποίο αναπαράγει τις λέξεις που εντοπίστηκαν. Το πλήθος των λέξεων στην πρόταση δεν είναι εκ των προτέρων γνωστό, αλλά μπορείτε να υποθέσετε ότι υπάρχει μικρό διάστημα απουσίας ομιλίας μεταξύ λέξεων.

Θα πρέπει να υλοποιήσετε και να συγκρίνετε τις επιδόσεις των παρακάτω ταξινομητών: Least Squares, SVM, RNN και MLP τριών επιπέδων (προσδιορίστε τον αριθμό νευρώνων ανά επίπεδο). Η σύγκριση θα γίνει όπως προβλέπεται σε δυαδικά συστήματα ταξινόμησης.

B) Από τις λέξεις που προκύπτουν, υπολογίστε τη μέση θεμελιώδη συχνότητα του ομιλητή.

- Πρέπει να εξηγήσετε ποια δεδομένα χρησιμοποιήσατε κατά τον έλεγχο και την εκπαίδευση του συστήματος. Αν είναι δικά σας, πώς τα δημιουργήσατε και αν είναι open source, πώς αξιοποιούνται.
- Προσπαθήστε να μην εξαρτάται το σύστημα από τα χαρακτηριστικά της φωνής του ομιλητή, αλλά να είναι όσο το δυνατόν ανεξάρτητο ομιλητή.

Προσοχή!!!! Δεν μπορείτε να χρησιμοποιήσετε **συνελικτικά** νευρωνικά δίκτυα. Δεν είναι αποδεκτή η χρήση έτοιμων web services ή APIs για speech recognition. Δεν μπορείτε να χρησιμοποιήσετε **transfer learning** από ήδη εκπαιδευμένα δίκτυα. Οι αντίστοιχες λύσεις μηδενίζονται.

Καλή επιτυχία!