

## Εργασία Εαρινού Εξαμήνου 2024-2025

**Γενικές οδηγίες:** Ημερομηνία παράδοσης: 30/5/2025, ώρα 23:59μμ. Η εργασία συμμετέχει με βάρος 25% στον τελικό βαθμό, δηλαδή, αν παραδοθεί, γράφεται 3 από τα 4 ισοδύναμα θέματα των γραπτών εξετάσεων (όποια επιλέξετε). Αν δεν παραδοθεί, τότε στις γραπτές εξετάσεις εξετάζεστε και στα 4 θέματα. Επίσης, η εργασία είναι **ατομική** και παραδίδεται μόνο μέσω της πλατφόρμας e-class και δεν θα βαθμολογηθούν εκπρόθεσμες εργασίες ή εργασίες που αποστέλλονται με email. Αποδεκτές γλώσσες είναι οι Matlab και Python. Η εργασία θα είναι η ίδια και τον Σεπτέμβριο. Η λογοκλοπή κάθε είδους οδηγεί σε μηδενισμό.

**Παραδοτέα:** **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, και **β)** τα αρχεία κώδικα σε ένα συμπίεσμένο αρχείο με όνομα source2025.zip (ή .rar ή άλλη σχετική κατάληξη). Δεν παραδίδονται αρχεία δεδομένων διότι αυτά είναι κοινά για όλες τις εργασίες και παρέχονται μέσω του e-class.

**Θέμα:** Καλείστε να υλοποιήσετε ένα σύστημα που προχωρά στην κατάτμηση μιας πρότασης σε τμήματα σήματος ομιλίας (foreground) και σήματος υποβάθρου (background), χρησιμοποιώντας **υποχρεωτικά** έναν ταξινομητή background vs foreground (ανά frame) και εφαρμόζοντας στη συνέχεια μετα-επεξεργασία της ακολουθίας αποφάσεων του ταξινομητή. Τελικώς, δοθείσης μιας ηχογράφησης, το σύστημα επιστρέφει τα χρονικά όρια των τμημάτων σήματος ομιλίας και των τμημάτων σήματος υποβάθρου (σε δευτερόλεπτα). Το format των αποτελεσμάτων είναι ένα αρχείο csv, της μορφής:

```
Audiofile, start, end, class
File1, 0, 2.2, background
File1, 2.2, 4, foreground
File1, 4,10, background
...
```

Θα πρέπει να υλοποιήσετε και να σχολιάσετε τις επιδόσεις των παρακάτω ταξινομητών: Least Squares και MLP τριών επιπέδων (αποφασίστε το πλήθος νευρώνων ανά στρώμα).

Για την εκπαίδευση, παρέχεται στο link

[https://drive.google.com/drive/folders/1A-\\_ybw6sVtPYjrzkOa1rACvvCygPqhWn?usp=share\\_link](https://drive.google.com/drive/folders/1A-_ybw6sVtPYjrzkOa1rACvvCygPqhWn?usp=share_link)  
ένας κατάλογος αρχείων train, με υποκαταλόγους ομιλίας (speech) και υποβάθρου (noise). Μπορείτε να χρησιμοποιήσετε ένα υποσύνολο των αρχείων αυτών. Για την τελική δοκιμή παρέχεται ένας κατάλογος test με ένα μεικτό αρχείο, καθώς και τα σχετικά transcriptions σε json format (υποκατάλογος transcriptions). Από τα json αρχεία, θα χρειαστεί να διατηρήσετε μόνο τα δεδομένα που αφορούν στο test αρχείο που δίνεται και να αγνοήσετε την υπόλοιπη πληροφορία. Όλα τα αρχεία προέρχονται από γνωστά, δημόσια σύνολα δεδομένων (<https://www.openslr.org>, συλλογές MUSAN και CHiME) και ίσως χρειαστεί η τμηματική επεξεργασία τους, αναλόγως του τρόπου ανάλυσης που θα επιλέξετε. Ανεξαρτήτως προσέγγισης, προσπαθήστε να ποσοτικοποιήσετε κατάλληλα τις επιδόσεις των ταξινομητών.

**Σημειώσεις:** Δεν μπορείτε να χρησιμοποιήσετε άλλου είδους νευρωνικό δίκτυο. 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!**