

# ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

## ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

### ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

#### Εκφώνηση 3ης Εργαστηριακής Άσκησης:

#### Αναγνώριση Είδους και Εξαγωγή Συναισθήματος από Μουσική

Εξηγήστε περιεκτικά και επαρκώς την εργασία σας. Κώδικας χωρίς σχόλια δεν θα βαθμολογηθεί. Επιτρέπεται η συνεργασία εντός ομάδων των 2 ατόμων εφόσον φοιτούν στο ίδιο πρόγραμμα σπουδών (είτε ομάδες προπτυχιακών, είτε ομάδες μεταπτυχιακών). Κάθε ομάδα 2 ατόμων υποβάλλει μια κοινή αναφορά που αντιπροσωπεύει μόνο την προσωπική εργασία των μελών της. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός των βιβλίων και του εκπαιδευτικού υλικού του μαθήματος, πρέπει να το αναφέρετε. Η παράδοση της αναφοράς και του κώδικα της εργασίας θα γίνει ηλεκτρονικά στη [σελίδα του μαθήματος](#).

**Επισημαίνεται ότι απαγορεύεται η ανάρτηση των λύσεων των εργαστηριακών ασκήσεων στο github, ή σε άλλες ιστοσελίδες.**

Στην [ακόλουθη σελίδα](#) μπορείτε να βρείτε βοηθητικό κώδικα σχετικά με τα εργαστήρια. Στη σελίδα αυτή μπορείτε επίσης να υποβάλετε απορίες και ερωτήσεις προς τους βοηθούς του μαθήματος με μορφή issues. Ερωτήσεις αναφορικά με το εργαστήριο που θα γίνονται μέσω mail δεν θα λαμβάνουν απάντηση.

**Σημαντικό:** Κατά τη διάρκεια του εργαστηρίου θα υπάρχει ένα hands on κομμάτι, όπου θα επιλύσετε τα ερωτήματα της άσκησης. Στόχος είναι όταν φύγετε από το εργαστήριο να έχετε γράψει το μεγαλύτερο μέρος του απαιτούμενου κώδικα. Για αυτό είναι σημαντικό να έχετε ολοκληρώσει όλα τα βήματα της προπαρασκευής **ΠΡΙΝ** έρθετε στο εργαστήριο (Η εκπαίδευση των πλήρων μοντέλων που είναι χρονοβόρα θα πραγματοποιηθεί από το σπίτι, αλλά θα χρησιμοποιήσουμε ένα μικρό υποσύνολο των δεδομένων στο hands on κομμάτι)

### ΠΕΡΙΓΡΑΦΗ

Σκοπός της άσκησης είναι η αναγνώριση του είδους και η εξαγωγή συναισθηματικών διαστάσεων από φασματογραφήματα (spectrograms) μουσικών κομματιών. Σας δίνονται 2 σύνολα δεδομένων, το Free Music Archive (FMA) genre με 3834 δείγματα χωρισμένα σε 20 κλάσεις (είδη μουσικής) και τη βάση δεδομένων (dataset) multitask music με 1497 δείγματα με επισημειώσεις (labels) για τις τιμές συναισθηματικών διαστάσεων όπως valence, energy και danceability. Τα δείγματα είναι φασματογραφήματα, τα οποία έχουν εξαχθεί από clips 30 δευτερολέπτων από διαφορετικά τραγούδια.

Θα ασχοληθούμε με την ανάλυση των φασματογραφημάτων με χρήση βαθιών αρχιτεκτονικών με συνελκτικά νευρωνικά δίκτυα (CNN) και αναδρομικά νευρωνικά δίκτυα (RNN).

Η άσκηση χωρίζεται σε 5 μέρη:

- 1) Ανάλυση των δεδομένων και εξοικείωση με τα φασματογραφήματα.
- 2) Κατασκευή ταξινομητών για το είδος της μουσικής πάνω στη βάση δεδομένων (dataset) FMA.

- 3) Κατασκευή regression μοντέλων για την πρόβλεψη valence, energy και danceability πάνω στη Multitask βάση δεδομένων.
- 4) Χρήση προηγμένων τεχνικών εκπαίδευσης (transfer - multitask) learning για τη βελτίωση των αποτελεσμάτων
- 5) (Προαιρετικά) Υποβολή των μοντέλων στο Kaggle competition του εργαστηρίου και σύγκριση των αποτελεσμάτων [0]

Τα δεδομένα είναι διαθέσιμα στο [1]. Μπορείτε να κάνετε χρήση των kaggle kernels για να έχετε πρόσβαση σε δωρεάν GPUs: [2].

## BIBΛΙΟΘΗΚΕΣ PYTHON

- librosa, numpy, pytorch, scikit-learn

## ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ

- [0] <https://www.kaggle.com/c/multitask-affective-music-lab-2022/overview>
- [1] <https://www.kaggle.com/c/multitask-affective-music-lab-2022/data>
- [2] <https://www.kaggle.com/c/multitask-affective-music-lab-2022/kernels>
- [3] <https://www.coursera.org/lecture/machine-learning-projects/transfer-learning-WNPap>
- [4] <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>
- [5] <http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf>
- [6] <http://runder.io/multi-task/>
- [7] <https://arxiv.org/pdf/1706.05137.pdf>
- [8] <https://github.com/slp-ntua/patrec-labs/tree/main/lab3>
- [9] <https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1>
- [10] <https://becominghuman.ai/how-to-evaluate-the-machine-learning-models-part-3-ff0dd3b76f9>
- [11] <https://towardsdatascience.com/metrics-for-imbalanced-classification-41c71549bbb5>
- [12] <https://twitter.com/karpathy/status/1013244313327681536>
- [13] <http://karpathy.github.io/2019/04/25/recipe/>
- [14] [https://www.reddit.com/r/MachineLearning/comments/5pidk2/d\\_is\\_overfitting\\_on\\_a\\_very\\_small\\_data\\_set\\_a/](https://www.reddit.com/r/MachineLearning/comments/5pidk2/d_is_overfitting_on_a_very_small_data_set_a/)
- [15] <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [16] <https://colah.github.io/posts/2014-07-Understanding-Convolutions/>
- [17] <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>
- [18] <https://blog.xrds.acm.org/2016/06/convolutional-neural-networks-cnns-illustrated-explanation/>
- [19] <https://cs.stanford.edu/people/karpathy/convnetjs/>

## ΕΚΤΕΛΕΣΗ

Στην προπαρασκευή θα ασχοληθούμε με την αναγνώριση είδους μουσικής με βάση το [φασματογράφημα](#) (spectrogram). Όπως είδαμε και στο εργαστήριο 2 το φασματογράφημα είναι μια οπτική αναπαράσταση της μεταβολής του συχνοτικού περιεχομένου ενός σήματος με το χρόνο (time-frequency distribution), όπου η εξαγόμενη εικόνα αναπαριστά την ενέργεια του σήματος για διαφορετικές ζώνες συχνοτήτων και χρονικά παράθυρα.

### **Βήμα 0: Εξοικείωση με Kaggle kernels**

Ανοίξτε ένα (private) Kaggle kernel στη σελίδα [2].

Τα δεδομένα μπορούν να φορτωθούν όπως φαίνεται από το [notebook](#).

Τρέξτε την εντολή `os.listdir("../input/data/data/")` για να εξερευνήσετε τους υποφάκελους, δοκιμάστε να ενεργοποιήσετε και να απενεργοποιήσετε τη GPU και κάντε commit τις αλλαγές σας.

### **Βήμα 1: Εξοικείωση με φασματογραφήματα στην κλίμακα mel**

Τα δεδομένα που θα χρησιμοποιήσετε στην προπαρασκευή είναι ένα υποσύνολο του Free Music Archive (FMA) dataset. Το FMA είναι μια βάση δεδομένων από ελεύθερα δείγματα (clips) μουσικής με επισημειώσεις ως προς το είδος της μουσικής.

Έχουμε εξάγει τα φασματογραφήματα και τις επισημειώσεις τους στο φάκελο `../input/data/data/fma_genre_spectrogram`.

Το αρχείο `../input/data/data/fma_genre_spectrograms/train_labels.txt` περιέχει γραμμές του στη μορφή `"spec_file label"`.

α) Διαλέξτε δύο τυχαίες γραμμές με διαφορετικές επισημειώσεις (labels). Τα αντίστοιχα αρχεία βρίσκονται στο φάκελο `../input/data/data/fma_genre_spectrograms/train`,

β) Διαβάστε τα αρχεία και πάρτε το φασματογράφημα σε κλίμακα mel σύμφωνα με τις οδηγίες του [1].

γ) Απεικονίστε τα φασματογραφήματα για τα διαφορετικά labels με χρήση της συνάρτησης `librosa.display.specshow`. Σχολιάστε τι πληροφορία σας δίνουν και τις διαφορές για δείγματα που αντιστοιχούν σε διαφορετικές επισημειώσεις. (Υπόδειξη: συχνότητα στον κατακόρυφο άξονα, χρόνος στον οριζόντιο)

### **Βήμα 2: Συγχρονισμός φασματογραφημάτων στο ρυθμό της μουσικής (beat-synced spectrograms)**

α) Τυπώστε τις διαστάσεις των φασματογραφημάτων του Βήματος 1.

- Πόσα χρονικά βήματα έχουν;
- Είναι αποδοτικό να εκπαιδεύσετε ένα LSTM πάνω σε αυτά τα δεδομένα;
- Γιατί;

β) Ένας τρόπος να μειώσουμε να τα χρονικά βήματα είναι να συγχρονίσουμε τα φασματογραφήματα πάνω στο ρυθμό. Για αυτό το λόγο παίρνουμε τη διάμεσο (median) ανάμεσα στα σημεία που χτυπάει το beat της μουσικής. Τα αντίστοιχα αρχεία δίνονται στο φάκελο `../input/data/data/fma_genre_spectrograms_beat`. Επαναλάβετε τα βήματα

του Βήματος 1 για αντίστοιχα beat-synced spectrograms και σχολιάστε τις διαφορές με τα αρχικά.

### **Βήμα 3: Εξοικείωση με χρωμογραφήματα**

Τα χρωμογραφήματα ([chromagrams](#)) απεικονίζουν την ενέργεια του σήματος μουσικής για τις ζώνες συχνοτήτων που αντιστοιχούν στις δώδεκα διαφορετικές νότες της κλίμακας κλασικής μουσικής {C, C#, D, D#, E, F, F#, G, G#, A, A#, B} και μπορούν να χρησιμοποιηθούν ως εργαλείο για την ανάλυση της μουσικής αναφορικά με τα αρμονικά και μελωδικά χαρακτηριστικά της ενώ επίσης είναι αρκετά εύρωστα και στην αναγνώριση των αλλαγών του ηχοχρώματος και των οργάνων (μπορεί να θεωρήσει κάποιος ότι το χρωμογράφημα είναι ένα φασματογράφημα modulo την οκτάβα).

Επαναλάβετε τα υποερωτήματα από τα Βήματα 1 και 2 για τα χρωμογραφήματα των αντίστοιχων αρχείων.

### **Βήμα 4: Φόρτωση και ανάλυση δεδομένων**

Χρησιμοποιήστε το βοηθητικό κώδικα [8]

α) Στο βοηθητικό κώδικα σας παρέχεται έτοιμη μια υλοποίηση ενός PyTorch Dataset η οποία διαβάζει τα δεδομένα και σας επιστρέφει τα δείγματα. Μελετήστε τον κώδικα και τα δείγματα που επιστρέφει και σχολιάστε τις λειτουργίες που εκτελούνται.

β) Στον κώδικα που σας δίνουμε συγχωνεύουμε κλάσεις που μοιάζουν μεταξύ τους και αφαιρούμε κλάσεις που αντιπροσωπεύονται από πολύ λίγα δείγματα.

γ) Σχεδιάστε δύο ιστογράμματα που θα δείχνουν πόσα δείγματα αντιστοιχούν σε κάθε κλάση, ένα πριν από τη διαδικασία του βήματος 4β και ένα μετά.

### **Βήμα 5: Αναγνώριση μουσικού είδους με LSTM.**

Με τη βοήθεια του κώδικα που υλοποιήσατε στη δεύτερη άσκηση

α) Προσαρμόστε τον κώδικα του LSTM που υλοποιήσατε στη 2η άσκηση για να δέχεται ως είσοδο τα φασματογραφήματα από το Pytorch dataset του βήματος 4.

β) Για να επισπεύσετε τη διαδικασία ανάπτυξης και αποσφαλμάτωσης των μοντέλων σας, στη συνάρτηση `train()` που εκπαιδεύει το μοντέλο σας προσθέστε μια boolean παράμετρο `overfit_batch`. Όταν η `overfit_batch` είναι `False` το δίκτυο εκπαιδεύεται κανονικά. Όταν είναι `True`, θα πραγματοποιεί υπερεκπαίδευση του δικτύου σε ένα μικρό σύνολο από batches (3-4).

- Υπερεκπαίδευση του δικτύου σε ένα batch: Μια καλή πρακτική κατά την ανάπτυξη νευρωνικών είναι να βεβαιωθούμε ότι το δίκτυο μπορεί να εκπαιδευτεί (τα gradients γυρνάνε πίσω κτλ). Ένας γρήγορος τρόπος για να γίνει αυτό είναι να επιλέξουμε τυχαία ένα πολύ μικρό υποσύνολο των δεδομένων (ένα batch) και να εκπαιδεύσουμε το δίκτυο για πολλές εποχές πάνω σε αυτό. Αυτό που περιμένουμε να δούμε είναι το σφάλμα εκπαίδευσης να πάει στο 0 και το δίκτυο να κάνει overfit (δείτε και τα [12], [13], [14]).

γ) εκπαιδεύστε ένα LSTM [15] δίκτυο, το οποίο θα δέχεται ως είσοδο τα φασματογραφήματα του συνόλου εκπαίδευσης (train set) και θα προβλέπει τις διαφορετικές κλάσεις (μουσικά είδη) του συνόλου δεδομένων (dataset).

δ) εκπαιδεύστε ένα LSTM δίκτυο, το οποίο θα δέχεται ως είσοδο τα beat-synced spectrograms (train set) και θα προβλέπει τις διαφορετικές κλάσεις (μουσικά είδη) του συνόλου δεδομένων.

ε) εκπαιδεύστε ένα LSTM δίκτυο, το οποίο θα δέχεται ως είσοδο τα χρωμογραφήματα (train set) και θα προβλέπει τις διαφορετικές κλάσεις (μουσικά είδη) του συνόλου δεδομένων.

ζ) (extra credit) εκπαιδεύστε ένα LSTM δίκτυο, το οποίο θα δέχεται ως είσοδο τα ενωμένα (concatenated) χρωμογραφήματα και φασματογραφήματα (train set) και θα προβλέπει τις διαφορετικές κλάσεις (μουσικά είδη) του συνόλου δεδομένων.

**Υπόδειξη:** Για την εκπαίδευση χρησιμοποιήστε και σύνολο επαλήθευσης (validation set).

**Υπόδειξη:** Για την εκπαίδευση ενεργοποιήστε τη GPU.

**Υπόδειξη:** Χρησιμοποιείτε Adam optimizer

**Υπόδειξη:** Χρησιμοποιείτε την κλάση [Subset](#) του PyTorch, ώστε να μπορείτε να εκπαιδεύσετε τα μοντέλα σας σε μικρότερα υποσύνολα και να επιταχύνετε τις διαδικασίες debugging/tuning.

Παράδειγμα [χρήσης](#).

## **Βήμα 6: Αξιολόγηση των μοντέλων**

Αναφέρετε τα αποτελέσματα των μοντέλων από το Βήμα 5 στο δύο ακόλουθα σύνολα αξιολόγησης (test sets)

- `"../input/data/data/fma_genre_spectrograms_beat/test_labels.txt"`
- `"../input/data/data/fma_genre_spectrograms/test_labels.txt"`

Συγκεκριμένα

α) υπολογίστε το accuracy

β) υπολογίστε το precision, recall και F1-score για κάθε κλάση

γ) υπολογίστε το macro-averaged precision, recall και F1-score για όλες τις κλάσεις

δ) υπολογίστε το micro-averaged precision, recall και F1-score για όλες τις κλάσεις

Αναφέρετε την ερμηνεία των μετρικών αυτών και σχολιάστε ποια από αυτές τις μετρικές θα επιλέγατε για την αξιολόγηση ενός ταξινομητή σε αυτό το πρόβλημα. Συγκεκριμένα εστιάστε στις ερωτήσεις

- Τι δείχνει το accuracy / precision / recall / f1 score;
- Τι δείχνει το micro / macro averaged precision / recall / f1 score;
- Πότε μπορεί να έχω μεγάλη απόκλιση ανάμεσα στο accuracy / f1 score και τι σημαίνει αυτό;
- Πότε μπορεί να έχω μεγάλη απόκλιση ανάμεσα στο micro/macro f1 score και τι σημαίνει αυτό;

- Υπάρχουν προβλήματα όπου το precision με ενδιαφέρει περισσότερο από το recall και αντίστροφα; Είναι ένα καλό accuracy / f1 αρκετό σε αυτές τις περιπτώσεις για να επιλέξω ένα μοντέλο;

**Υπόδειξη:** Χρησιμοποιήστε τη συνάρτηση `sklearn.metrics.classification_report`

**Υπόδειξη:** Δείτε τα [9], [10], [11]

-----  
----- **ΤΕΛΟΣ ΠΡΟΠΑΡΑΣΚΕΥΗΣ** -----  
-----

## ΠΑΡΑΔΟΤΕΑ

1. Γραπτή αναφορά σε .pdf και γραμμένη στο Latex / Word / Google Docs etc. (OXI .ipynb / .html)
2. Κώδικας python για την υλοποίηση σε .py (OXI .ipynb / .html). Αν χρησιμοποιήσατε notebooks για την υλοποίηση θα πρέπει να κάνετε export σε .py πριν την υποβολή.

Για την υποβολή, δημιουργείτε ένα .zip (OXI rar, / tar.\* / .7z) με τα παραδοτέα (1) και (2) και με όνομα ("patrec1\_ΑριθμόςΜητρώου1ουΣυνεργάτη\_ΑριθμόςΜητρώου2ουΣυνεργάτη.zip").

Για τις ομαδικές υποβολές, πρέπει να υποβάλλει ένας εκ των δύο συνεργατών και να αναγράφονται τα στοιχεία και των δύο στην πρώτη σελίδα της αναφοράς καθώς και στο όνομα του zip, όπως αναφέρεται παραπάνω.

Αν δεν έχετε διαθέσιμο ιδρυματικό e-mail / δεν μπορείτε να πραγματοποιήσετε την υποβολή, μπορείτε να επικοινωνήσετε με τους βοηθούς του μαθήματος [geopar|efthygeo]