

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

Εκφώνηση 2ης 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.

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) Οπτικοποίηση (visualization) αναπαραστάσεων μουσικών κομματιών με χρήση αλγορίθμων μείωσης διαστασιμότητας (dimensionality reduction).

Τα δεδομένα είναι [διαθέσιμα εδώ](#). Μπορείτε να κάνετε χρήση των kaggle kernels για να έχετε πρόσβαση σε δωρεάν GPUs και να χρησιμοποιήσετε [αυτό το kernel](#) ως οδηγό για να αναπτύξετε τη λύση σας (επιλέξτε Copy & Edit για να δημιουργήσετε έναν δικό σας «κλώνο»).

BIBΛΙΟΘΗΚΕΣ PYTHON

- librosa, numpy, pytorch, scikit-learn

ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ

- [0] <https://www.kaggle.com/pxaris/lab2-data-loading-tutorial>
- [1] <https://www.kaggle.com/datasets/geoparslp/patreco3-multitask-affective-music>
- [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] <https://www.ruder.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-eb8b2c2ca1>
- [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://x.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/
- [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://cs.stanford.edu/people/karpathy/convnetjs/>
- [19] <https://arxiv.org/abs/2104.01778>
- [20] <https://towardsdatascience.com/dimensionality-reduction-for-data-visualization-pca-vs-tsne-vs-umap-be4aa7b1cb29>

ΕΚΤΕΛΕΣΗ

Στην προπαρασκευή θα ασχοληθούμε με την αναγνώριση είδους μουσικής με βάση το [φασματογράφημα](#) (spectrogram). Το φασματογράφημα είναι μια οπτική αναπαράσταση της μεταβολής του συχνотικού 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

Επισκεφθείτε το Kaggle. Χρησιμοποιήστε [αυτό το kernel](#) ως οδηγό και επιλέξτε “Copy & Edit” για να δημιουργήσετε έναν δικό σας «κλώνο».

Τρέξτε τις εντολές:

```
import os  
os.listdir("../input/patreco3-multitask-affective-music/data/")
```

για να εξερευνήσετε τους υποφάκελους. Δοκιμάστε να ενεργοποιήσετε και να απενεργοποιήσετε τη GPU και κάντε Save 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 εξαγάγει τα φασματογραφήματα και τις επισημειώσεις τους στο φάκελο

```
../input/patreco3-multitask-affective-music/data/fma_genre_spectrogram.
```

Το αρχείο `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 `fma_genre_spectrograms/train`.

β) Διαβάστε τα αρχεία και πάρτε το φασματογράφημα σε κλίμακα mel ακολουθώντας το [0].

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).

δ) Ποια είναι η κλίμακα “Mel”, πώς δημιουργήθηκε και γιατί την χρησιμοποιούμε στην επεξεργασία μουσικών σημάτων; Σχολιάστε στην αναφορά σας.

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 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/patreco3-multitask-affective-music/data/fma_genre_spectrogram_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 στην αναγνώριση των αλλαγών του ηχοχρώματος και των οργάνων (μπορεί να θεωρήσει κάποιος ότι το χρωμογράφημα είναι ένα φασματογράφημα modulo την οκτάβα).

Επαναλάβετε τα Βήματα 1 (α, β, γ) και 2 για τα χρωμογραφήματα των αντίστοιχων αρχείων.

Βήμα 4: Φόρτωση και ανάλυση δεδομένων

Χρησιμοποιήστε τον βοηθητικό κώδικα στο [0] και στο [8]

α) Στον βοηθητικό κώδικα παρέχεται έτοιμη η 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 να 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 στην προηγούμενη άσκηση:

α) Προσαρμόστε τον κώδικα του LSTM (της προηγούμενης άσκησης) για να δέχεται ως είσοδο τα φασματογραφήματα από το Pytorch dataset του βήματος 4.

β) Για να επιταχύνετε 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 (π.χ. ένα batch) και να εκπαιδεύσουμε το δίκτυο για πολλές εποχές πάνω σε αυτό. Αυτό που περιμένουμε να δούμε είναι το σφάλμα εκπαίδευσης να πηγαίνει 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) 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.

Υποδείξεις:

- Για την εκπαίδευση χρησιμοποιήστε και σύνολο επαλήθευσης (validation set).
- Ενεργοποιήστε τη GPU.
- Χρησιμοποιήστε Adam optimizer.
- Χρησιμοποιήστε 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 στα ακόλουθα δύο σύνολα αξιολόγησης (test sets):

- `fma_genre_spectrograms_beat/test_labels.txt`
- `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 γίνεται να προκύπτει μεγάλη απόκλιση ανάμεσα στο accuracy / F1 score και τι σημαίνει αυτό;
- Πότε γίνεται να προκύπτει μεγάλη απόκλιση ανάμεσα στο micro/macro F1 score και τι σημαίνει αυτό;
- Υπάρχουν προβλήματα όπου το precision μας ενδιαφέρει περισσότερο από το recall και αντίστροφα; Είναι μια καλή τιμή accuracy / F1 αρκετή 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 -----

Βήμα 7.1: 2D CNN

Ένας άλλος τρόπος για την κατασκευή ενός μοντέλου για την επεξεργασία ηχητικών σημάτων είναι να δούμε το φασματογράφημα σαν εικόνα και να χρησιμοποιήσουμε συνελκτικά δίκτυα (CNN).

α) Στο σύνδεσμο [18] μπορείτε να εκπαιδεύσετε απλά συνελκτικά δίκτυα και να δείτε την εσωτερική λειτουργία του δικτύου οπτικοποιώντας τις ενεργοποιήσεις (activations) των επιμέρους επιπέδων του δικτύου χωρίς προγραμματιστικό κόπο. Εκπαιδεύστε ένα δίκτυο στο MNIST και παρατηρήστε τη λειτουργία των ενεργοποιήσεων κάθε επιπέδου. Σχολιάστε τις επιμέρους λειτουργίες, τι φαίνεται να μαθαίνει το δίκτυο και δώστε κατάλληλα screenshots στην αναφορά.

β) Υλοποιήστε ένα 2D CNN με 4 επίπεδα (layers) που θα επεξεργάζεται το φασματογράφημα σαν μονοκάναλη εικόνα. Εκπαιδεύστε το χρησιμοποιώντας τα training & validation sets και αναφέρετε τα αποτελέσματα στο test set. Κάθε επίπεδο θα πραγματοποιεί τις εξής λειτουργίες (operations) με αυτή τη σειρά:

- 1) 2D convolution
- 2) Batch normalization
- 3) ReLU activation
- 4) Max pooling

γ) Εξηγήστε τη λειτουργία και τον ρόλο των convolutions, batch normalization, ReLU και Max pooling. Παραπέμπουμε στις αναφορές [16], [17].

δ) Πραγματοποιήστε τη διαδικασία overfit_batch του βήματος 5β για να βεβαιωθείτε ότι το δίκτυο μπορεί να εκπαιδευτεί.

ε) Χρησιμοποιήστε αυτή την αρχιτεκτονική για την αναγνώριση μουσικού είδους με **φασματογραφήματα** και συγκρίνετε με το μοντέλο 5α.

Υποδείξεις:

- Εκτελέστε το μοντέλο σε διαφορετικό kernel για να αποφύγετε προβλήματα μνήμης.
- Ισχύουν όλες οι υποδείξεις του Βήματος 5.
- Μην σπαταλήσετε πολύ χρόνο στη ρύθμιση των υπερπαραμέτρων (hyperparameters) του δικτύου. Απλά δείτε κάποιες έτοιμες online υλοποιήσεις από CNNs και βάλτε κάποιες “λογικές” τιμές (πχ kernel size ~ 3 ή 5). Αν το δίκτυο σας δεν λειτουργεί όπως θα έπρεπε, είναι πιο πιθανό να οφείλεται σε κάποιο λάθος (bug) στον κώδικα παρά σε κακή επιλογή παραμέτρων, ειδικά αν δεν αποκλίνουν πολύ από τις προεπιλεγμένες (default) τιμές.

Βήμα 7.2: AST - Audio Spectrogram Transformer

Εκτός από τα CNNs, μπορούμε να χρησιμοποιήσουμε αρχιτεκτονικές Transformer για την επεξεργασία ηχητικών σημάτων, όπως το μοντέλο “Audio Spectrogram Transformer”.

α) Αντλώντας πληροφορίες από τη δημοσίευση που βρίσκεται στον σύνδεσμο [19], περιγράψτε συνοπτικά τον τρόπο λειτουργίας του μοντέλου στην αναφορά σας.

β) Χρησιμοποιήστε την έτοιμη υλοποίηση “ASTBackbone” που παρέχεται στο αρχείο “ast.py”, του βοηθητικού κώδικα [8], για την αναγνώριση μουσικού είδους με φασματογραφήματα και συγκρίνετε με τα αποτελέσματα που πήρατε με τα μοντέλα των ερωτημάτων 7.1-(ε) και 5-(α).

Υποδείξεις:

- Πραγματοποιήστε τη διαδικασία overfit_batch για να βεβαιωθείτε ότι το δίκτυο εκπαιδεύεται.
- Λόγω των μεγαλύτερων υπολογιστικών απαιτήσεων του μοντέλου, μπορείτε να επιλέξετε μικρότερο αριθμό εποχών σε σχέση με τα προηγούμενα μοντέλα, για την επιτάχυνση των πειραμάτων σας.

- Ορίστε κατάλληλα τις τιμές των παραμέτρων “input_fdim”, “input_tdim” και “feature_size” κατά την αρχικοποίηση του backbone μοντέλου και αφήστε τις υπόλοιπες (π.χ. “fstride”) στις προεπιλεγμένες τιμές.

Βήμα 8: Εκτίμηση συναισθήματος - συμπεριφοράς με παλινδρόμηση

Σε αυτό το βήμα θα χρησιμοποιήσετε το multitask dataset (“../input/patreco3-multitask-affective-music/data/multitask_dataset/train_labels.txt”).

Εδώ σας δίνονται τα φασματογραφήματα, καθώς και επισημειώσεις σε 3 άξονες που αφορούν το συναίσθημα του τραγουδιού. Οι επισημειώσεις είναι πραγματικοί αριθμοί μεταξύ 0 και 1:

- Valence (πόσο θετικό ή αρνητικό είναι το συναίσθημα), όπου αρνητικό κοντά στο 0, θετικό κοντά στο 1.
- Energy (πόσο ισχυρό είναι το συναίσθημα), όπου ασθενές κοντά στο 0, ισχυρό κοντά στο 1.
- Danceability (πόσο χορευτικό είναι το τραγούδι), όπου μη χορευτικό κοντά στο 0, χορευτικό κοντά στο 1.

α) Προσαρμόστε τα μοντέλα των Βήματος 5 και 7 (LSTM, CNN, AST) για [παλινδρόμηση](#) (regression) αλλάζοντας τη συνάρτηση κόστους.

β) Εκπαιδεύστε τα μοντέλα του 8α για την εκτίμηση του valence.

γ) Επαναλάβετε για την εκτίμηση του energy.

δ) Επαναλάβετε για την εκτίμηση του danceability.

ε) Η τελική μετρική είναι το μέσο [Spearman correlation](#) ανάμεσα στις πραγματικές (ground truth) τιμές και στις προβλεπόμενες τιμές για όλους τους άξονες.

Υπόδειξη: Προσοχή. Σε αυτό το σύνολο δεδομένων δεν σας παρέχονται οι επισημειώσεις για το test set, οπότε η εκτίμηση του πόσο καλά γενικεύει το μοντέλο θα πρέπει να γίνει παίρνοντας ένα υποσύνολο από τα δεδομένα που σας δίνονται.

Βήμα 9: Μεταφορά γνώσης (Transfer Learning)

Ένας τρόπος για τη βελτίωση των βαθιών νευρωνικών όταν έχουμε λίγα διαθέσιμα δεδομένα είναι η μεταφορά της γνώσης από ένα άλλο μοντέλο, εκπαιδευμένο σε ένα μεγαλύτερο dataset. Για αυτό τον λόγο:

α) Δείτε τα links [3], [4], [5]. Περιγράψτε με 2 προτάσεις τα βασικά συμπεράσματα του [5].

β) Επιλέξτε ένα μοντέλο από τα βήματα 5, 7.1 και 7.2. Εξηγήστε γιατί επιλέξατε αυτό το μοντέλο.

γ) Εκπαιδεύστε αυτό το μοντέλο στο *fma_genre_spectrograms* dataset και αποθηκεύστε τα βάρη του δικτύου στην εποχή που έχει την καλύτερη επίδοση (checkpoint). Εξηγήστε ποια μετρική και σε ποιο σετ επιλέγετε για να κρίνετε πιο μοντέλο έχει την καλύτερη επίδοση. Γιατί;

δ) Αρχικοποιήστε ένα μοντέλο με αυτά τα βάρη για το πρόβλημα του ερωτήματος 8 και εκπαιδεύστε το για λίγες εποχές (fine-tuning) σε έναν από τους συναισθηματικούς άξονες της επιλογής σας (π.χ. valence).

ε) Συγκρίνετε τα αποτελέσματα με αυτά από το Βήμα 8.

Βήμα 10: Εκπαίδευση σε πολλαπλά προβλήματα (Multitask Learning)

Στο Βήμα 8 εκπαιδεύσαμε ξεχωριστά ένα μοντέλο για κάθε συναισθηματική διάσταση. Ένας τρόπος για να εκπαιδεύσουμε πιο αποδοτικά μοντέλα όταν μας δίνονται πολλές επισημειώσεις είναι η χρήση multitask learning.

- α) Δείτε τα links [3], [6], [7] και περιγράψτε με 2 προτάσεις τα βασικά συμπεράσματα του [7].
- β) Υλοποιήστε μια συνάρτηση κόστους η οποία θα λαμβάνει σαν είσοδο τα logits (outputs του μοντέλου) και τα targets (πραγματικές τιμές) για το valence, arousal και danceability και θα επιστρέφει το άθροισμα των επιμέρους losses για κάθε task. Μπορείτε να χρησιμοποιήσετε βάρη για να φέρετε τα επιμέρους κόστη στην ίδια τάξη μεγέθους.
- γ) Εκπαιδεύστε ένα μοντέλο στο multitask dataset χρησιμοποιώντας την παραπάνω συνάρτηση κόστους.
- δ) Συγκρίνετε τα αποτελέσματα με αυτά από το Βήμα 8.

Βήμα 11 (Προαιρετικό): Οπτικοποίηση κρυφών αναπαραστάσεων

Επιλέξτε τα αποθηκευμένα καλύτερα μοντέλα των βημάτων 5 (LSTM), 7.1 (CNN) και 7.2 (AST). Χρησιμοποιήστε το test set του *fma_genre_spectrograms* dataset και πάρτε την τελική αναπαράσταση (ένα διάνυσμα για κάθε τραγούδι), ακριβώς πριν το output layer (του Classifier).

α) Για κάθε μοντέλο, δημιουργήστε ένα νέο σύνολο δεδομένων που θα αποτελείται από το latent representation (διάνυσμα διάστασης “feature_size”) του κάθε τραγουδιού του test set (όπως αυτό εξάγεται από το τελευταίο layer του νευρωνικού πριν το output_layer) και από το αντίστοιχο label του (genre).

β) Δείτε τον σύνδεσμο [20]. Εφαρμόστε dimensionality reduction χρησιμοποιώντας PCA και προχωρήστε σε οπτικοποιήσεις του latent representation χώρου του κάθε μοντέλου. Πόσο πολύ διαχωρίζονται οι κλάσεις σε αυτόν τον χώρο; Συγκρίνετε τους χώρους που διαμορφώνουν τα διαφορετικά μοντέλα.

γ) Επαναλάβετε το παραπάνω χρησιμοποιώντας το μοντέλο του βήματος 5 που παρουσίασε την χειρότερη επίδοση. Παρατηρείτε κάποια ποιοτική διαφορά ανάμεσα σε αυτό και τα προηγούμενα μοντέλα;

δ) Δοκιμάστε περισσότερους αλγόριθμους dimensionality reduction (t-SNE κ.α.). Ποιος πιστεύετε ότι δουλεύει καλύτερα για τα δεδομένα σας;

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 ή/και .ipynb αρχείο που εξήχθη από το Kaggle που δουλέψατε την άσκησή σας (OXI .html).

Για την υποβολή, δημιουργήστε ένα .zip (OXI rar, / tar.* / .7z) με τα παραδοτέα (1) και (2) και με όνομα ("patrec2_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.