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 συνελικτικά 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»).

ΒΙΒΛΙΟΘΗΚΕΣ 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://aws.amazon.com/what-is/transfer-learning/
[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-ebe8b2c2ca1
[10] https://becominghuman.ai/how-to-evaluate-the-machine-learning-models-part-3-ff0dd3b76f9
[11] https://www.kaggle.com/code/marcinrutecki/best-techniques-and-metrics-for-imbalanced-dataset
[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://www.datacamp.com/tutorial/introduction-t-sne

ΕΚΤΕΛΕΣΗ

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

Βήμα 0: Εξοικείωση με Kaggle kernels

Επισκεφθείτε το Kaggle. Χρησιμοποιήσετε 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 υποφακέλους. 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 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 (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** ----------
