ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

Χειμερινό Εξάμηνο 2019-20

Εκφώνηση 3ης Εργαστηριακής Άσκησης:
Αναγνώριση Είδους και Εξαγωγή Συναισθήματος από Μουσική

ΠΕΡΙΓΡΑΦΗ

Σκοπός της άσκησης είναι η αναγνώριση του είδους και η εξαγωγή συναισθηματικών διαστάσεων από φασματογραφήματα 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 (spectrograms), 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 (dataset).
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].

ΒΙΒΛΙΟΘΗΚΕΣ PYTHON

- librosa, numpy, pytorch, scikit-learn

ΕΚΤΕΛΕΣΗ

Στην προπαρασκευή θα ασχοληθούμε με την αναγνώριση είδους μουσικής με βάση το φασματογράφημα (spectrogram). Όπως είδαμε και στο εργαστήριο 2 το φασματογραφήματα είναι μια οπτική αναπαράσταση του συχνοτικού περιεχομένου ενός σήματος, όπου η εξαγώμενη εικόνα αναπαριστά τις διαφορετικές ζώνες συχνοτήτων ως προς το χρόνο.

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 υποφακέλους, 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). 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 (labels). (υπόδειξη: συχνότητα στον κατακόρυφο άξονα, χρόνος στον οριζόντιο)

Βήμα 2: Συχρονισμός 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) σχετίζονται με δώδεκα διαφορετικές νότες (ημιτόνια) {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.

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. Σχολιάστε γιατί είναι αναγκαίο να γίνει αυτό.

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) 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).

β) 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 dataset.

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

δ) (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 dataset.

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

Υπόδειξη: Για την εκπαίδευση ενεργοποιήστε τη GPU.

Υπόδειξη: Για να επισπεύσετε τη διαδικασία ανάπτυξης και αποσφαλμάτωσης των μοντέλων σας προτείνονται οι ακόλουθες 2 τεχνικές (δείτε και τα [12], [13], [14])

- Εκπαίδευση και επαλήθευση για λίγες εποχές σε λίγα (4-5) batches: Στόχος αυτού είναι να βεβαιωθείτε ότι το δίκτυο μπορεί να τρέξει απροβλημάτιστα ένα κύκλο εκπαίδευσης (2-3 εποχών) σε ένα πολύ μικρό υποσύνολο των πραγματικών δεδομένων. Χρήσιμο για να πιάσουμε μικρά λάθη νωρίς.

- 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.

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** ----------

Βήμα 7: 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 συνελικτικά δίκτυα (CNN).

α) Στο σύνδεσμο [19] μπορείτε να εκπαιδεύσετε απλά συνελικτικά 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, να το εκπαιδεύσετε στο train + validation set και να 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], [18]

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

Υπόδειξη: Εκτελέστε το μοντέλο σε διαφορετικό kernel για να αποφύγετε προβλήματα μνήμης.

Υπόδειξη: Ισχύουν όλες οι υποδείξεις του Βήματος 5

Υπόδειξη: 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) κτλ. Αν το δίκτυο σας δε λειτουργεί όπως θα έπρεπε, είναι πιο πιθανό να οφείλεται σε κάποιο λάθος (bug) στον κώδικα από την κακή επιλογή παραμέτρων, ειδικά αν δεν αποκλίνουν πολύ τις προεπιλεγμένες (default) τιμές.

Βήμα 8: Εκτίμηση συναισθήματος - συμπεριφοράς με παλινδρόμηση

Σε αυτό το βήμα θα χρησιμοποιήσετε το multitask dataset (`'../input/data/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 το καλύτερο μοντέλο του Βήματος 5 και το μοντέλο του Βήματος 7 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 δε 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.

Τα βήματα 9α και 9β είναι προαιρετικά για τους μεταπτυχιακούς φοιτητές και υποχρεωτικά για τους προπτυχιακούς φοιτητές

Βήμα 9α: 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 το 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. 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)

δ) Αρχικοποιήστε ένα μοντέλο με αυτά τα βάρη για το πρόβλημα του ερωτήματος 10 και εκπαιδεύστε το για λίγες εποχές (fine tuning) στο multitask dataset. Για ευκολία μπορείτε να αναφέρετε τα αποτελέσματα μόνο για έναν από τους 3 άξονες.

ε) Συγκρίνετε τα αποτελέσματα με αυτά από το Βήμα 8

Βήμα 9β: Εκπαίδευση σε πολλά προβλήματα (Multitask Learning)

Στο Βήμα 10 εκπαιδεύσατε 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) Εκπαιδεύστε ένα μοντέλο στο multitask dataset χρησιμοποιώντας σαν συνάρτηση κόστους το άθροισμα από τα κόστη (losses) για το valence, energy και danceability. Μπορείτε να χρησιμοποιήσετε βάρη για να φέρετε τα επιμέρους κόστη στην ίδια τάξη μεγέθους.

γ) Συγκρίνετε τα αποτελέσματα με αυτά από το Βήμα 8

Βήμα 10 (Προαιρετικό): Υποβολή στο Kaggle

α) Επιλέξτε το καλύτερο μοντέλο σας για το multitask dataset και πραγματοποιήστε προβλέψεις για το valence, energy και danceability στα test δεδομένα.

β) Διαμορφώστε ένα αρχείο solution.txt στη μορφή

```
Id.fused.full.npy.gz,valence,energy,danceability
123212738.fused.full.npy.gz,0.153,0.961,0.013
```

γ) Υποβάλετε το solution.txt στο διαγωνισμό στο kaggle και δείτε τα αποτελέσματα στο leaderboard.

δ) Σχολιάστε πόσο κοντά είναι τα αποτελέσματά σας με αυτά που περιμένατε.

ΠΑΡΑΔΟΤΕΑ

1) Σύντομη αναφορά (σε pdf ή jupyter notebook) που θα περιγράφει τη διαδικασία που ακολουθήθηκε σε κάθε βήμα, καθώς και τα σχετικά αποτελέσματα. Τα αποτελέσματα πρέπει να συνοδεύονται και από ερμηνεία – σχολιασμό.
2) Κώδικας Python (συνοδευόμενος από σύντομα σχόλια). Προσπαθήστε να κάνετε vectorized υλοποιήσεις.

Συγκεντρώστε τα (1) και (2) σε ένα .zip αρχείο το οποίο πρέπει να αποσταλεί μέσω του mycourses.ntua.gr πριν από τη διεξαγωγή του εργαστηρίου.

ΧΡΗΣΙΜΟΙ ΣΥΝΔΕΣΜΟΙ

[0] https://www.kaggle.com/c/multitask-music-classification-2020/overview
[1] https://www.kaggle.com/c/multitask-music-classification-2020/data
[2] https://www.kaggle.com/c/multitask-music-classification-2020/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://ruder.io/multi-task/
[7] https://arxiv.org/pdf/1706.05137.pdf
[8] https://gist.github.com/efthymisgeo/42f1936d417c5a85209069339333bca8
[9] https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1
[10]https://medium.com/@george.drakos62/how-to-select-the-right-evaluation-metric-for-machine-learning-models-part-3-classification-3eac420ec991
[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/
[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/