Predicting song genre from audio features

By Argenis Arriojas

Data description

Data description

- Data available at Kaggle.com
  - https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks
- This data has been taken from Spotify's Web API
- Several csv files available
- 3 of these csv files contain information about
  - Artists (over 1 million): `'artists.csv'`
    - Associates some artists to musical genres
  - Tracks (over 500 thousand): `'tracks.csv'`
    - Associated to one or more artists
    - Each track has 11 audio features provided by Spotify, track duration and popularity
    - Audio features: mode, acousticness, danceability, energy, instrumentalness, liveness, loudness, speechiness, tempo, valence, key
  - Genre (around 3 thousand): `'data_by_genres_o.csv'`
    - Contains reference audio features for each genre

The project

Given a track’s audio features,

can we predict the track’s musical genre(s)?

Data preprocessing

- Track duration and popularity were included as features
- Track duration was log-transformed and rescaled to lie in range [0,1]
- Most of the features are already normalized
- Those not normalized, have been rescaled to range [0, 1]
- Genres have been limited to 9 of the most popular in the USA
  - `'rock', 'pop', 'country', 'hip hop', 'easy listening', 'jazz', 'blues', 'reggae', 'folk'`
  - List extracted from: https://www.statista.com/statistics/442354/music-genres-preferred-consumers-usa/

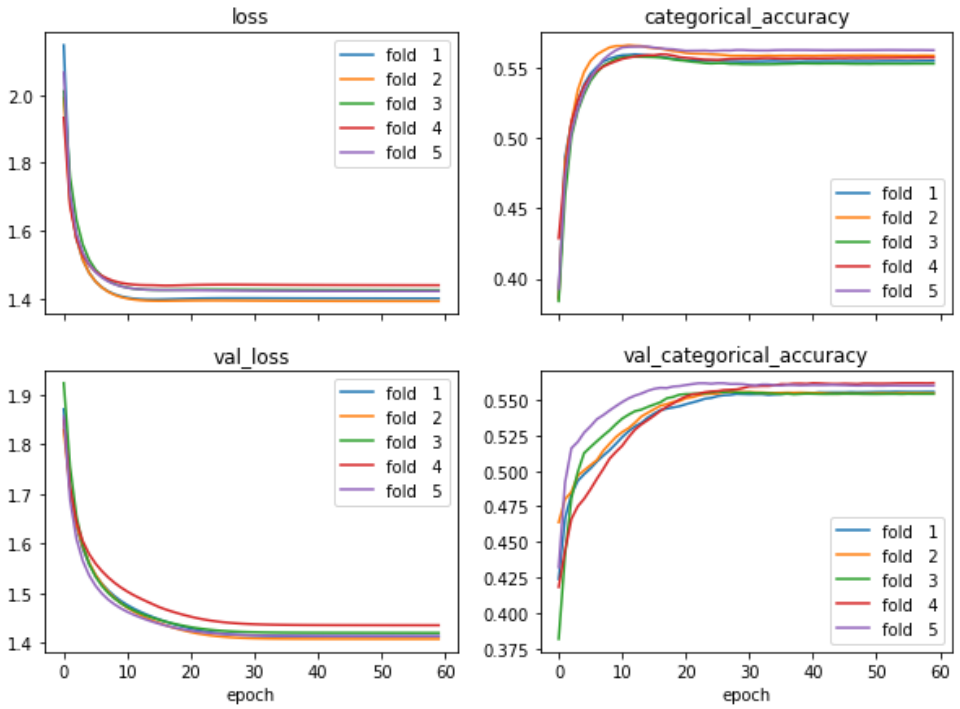
Training labels

- Obtaining labels for training process
  - Available track data is not directly associated to genre labels. We need to find a way to create training labels
  - Artists' data does contain relevant genre for each artist
  - We can use a track's artist(s) as a proxy to associate to musical genres
  - In this project we will focus on a small subset of genres
  - Tracks with no associated genre labels were excluded, resulting in around 80 thousand samples
  - Labels' data were one-hot encoded

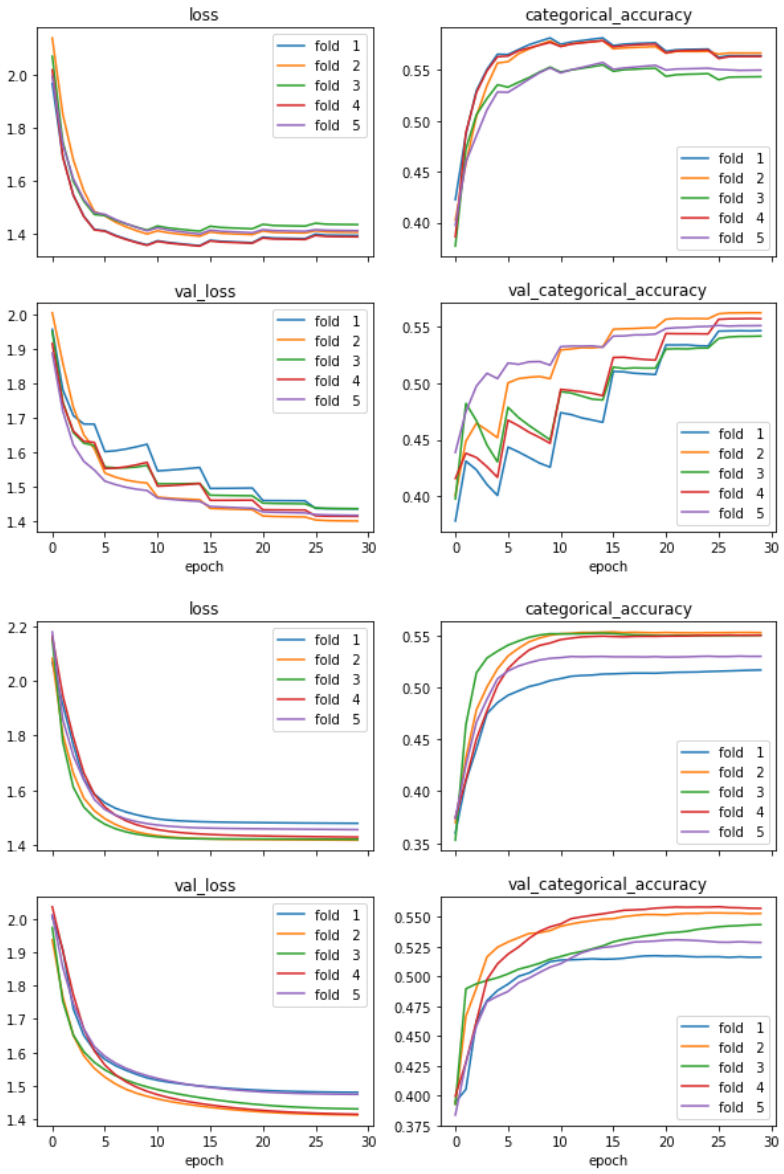
Model selection and implementation

- Several model configurations were considered, with 1, 2 and 3 hidden layers
- A learning rate schedule with exponential decay was used, starting with 1e-4
- Used 5-fold cross-validation on the labeled dataset
- The model selected showed good predictive power over all validation sets
- Parameters for the model selected:
  - Initial learning rate: 0.0001
  - Batch size: 5
  - Hidden layers: 1 layer with 10 nodes
  - Learning rate is halved every 5 epochs
  - Training was performed for 60 epochs

Results



- Performed 5-fold cross-validation
- Model accuracy saturates at around 55%
- Model performance remains consistent across all validation sets



About consistency of the model

- Figures on the left show different realizations of the experiment with the same model parameters as before
- In the upper figure, learning rate schedule is performed in staircase fashion
- Although there is some variance on the accuracy of the validation sets, these remain above 50%

Further improvements

- Clean training labels to make sure genres are accurately assigned to tracks
  - Exclude tracks with more than one artist. This would help remove uncertainty in training labels. This may introduce biases in genres where multiple artist are more frequent
- Consider incorporating reference audio features available at `'data_by_genres_o.csv'`
- For the categorical variable ‘key’, consider translating from a single numerical input to multiple inputs with one-hot encoding. This may improve how this feature shapes the weights in the model