Music Genre Classification from Lyrics

Team members:
Animi Reddy
Sri Keshav
Raghuchandra
Sushman

Objective

To predict the genre of a song based on the lyrics.

Data collection

The data was collected from the following website: https://www.kaggle.com/gyani95/380000-lyrics-from-metrolyrics

There are around 380,000+ lyrics in the data set from a lot of different artists from a lot of different genres arranged by year.

Data preprocessing

Our text preprocessing will include the following steps:

- Convert multiline text to single line text by replacing “\n” with “\t”.
- Convert all text to lower case.
- Replace all these symbols '[/(){}\[\]\|@,;]' by space in text.
- Remove all these symbols '[^0-9a-z #+_]' from text.
- Remove stop words.

Model

We have used LSTM recurrent neural network models in Python using Keras deep learning library.

Steps in LSTM modelling:

- Vectorize lyrics, by turning each text into either a sequence of integers or into a vector.
- Limit the data set to the top 50,000 words.
- Set the max number of words in each text at 250.

- Truncate and pad the input sequences so that they are all in the same length for modeling.
- Converting categorical labels to numbers.
- The first layer is the embedded layer that uses 100 length vectors to represent each word.
- SpatialDropout1D performs variational dropout in NLP models.
- The next layer is the LSTM layer with 100 memory units.
- The output layer must create 8 output values, one for each class.
- Activation function is softmax for multi-class classification.
- Because it is a multi-class classification problem, categorical_crossentropy is used as the loss function.

Evaluating the model

Accuracy for the SVM model:

Accuracy for the RNN model: