

Ari I.

As a foray into Deep Learning, I trained a music genre classification NN with the GTZAN dataset using Keras. I've trained the model with 80 epochs, with Keras determining the test accuracy to be 72.5%. I used the NN to predict the genres of 30-second clips from 10 songs, some of which were genres that the NN was not trained for (such as the Subnautica track being ambient electronic music). The predictions revolved strongly around 50% accuracy for my given set of tracks. Given that there are 10 genres to pick from, I'll say that using NN is better than picking at random. :)

Here's an average bout of predictions:

The predicted genre for `stayin_alive_the_begees.wav` is disco!  
The predicted genre for `feels_like_home.wav` is blues!  
The predicted genre for `cliffs_of_dover_eric_johnson.wav` is rock!  
The predicted genre for `gangplank_galleon.wav` is jazz!  
The predicted genre for `accoustic_guitar.wav` is disco!  
The predicted genre for `moonlight_sonata.wav` is jazz!  
The predicted genre for `1812_overture.wav` is disco!  
The predicted genre for `subnautica_below_zero.wav` is hiphop!  
The predicted genre for `bluegrass_riff.wav` is pop!  
The predicted genre for `in_a_sentimental_mood.wav` is blues!

where green is correct, blue is tangentially correct, and red is very incorrect.

It makes sense why the NN could not predict the Subnautica track correctly- the NN was not trained for ambient electronic music. Hiphop is definitely the closest genre that it could choose from, so I'd consider it correct. The 1812 Overture is also tricky because the clip of it centers on a moment when live cannons are shot. This begs the question: what genre of music most often has cannons?

In many of the code executions, the reds and the greens are flipped. I imagine this is the gradient descents' local minima at play. One minima may be accurate for one set of songs, while another minima is accurate for another set of songs. Nonetheless, I found that the predictions for "In a Sentimental Mood" and "Stayin Alive" were the most consistently accurate throughout executions.

Here's my observations on parameter changes, in search of higher accuracy:

- + Reducing the # of NN layers from 5 to 2 drastically reduced accuracy
- + Increasing the # of NN layers for 5 to 7 led to significantly less accurate results. Is this 'overfitting?'
- + Increasing the # of epochs past 30 led to slightly better results
- + Adding Dropout layers between Dense layers in a NN with 7 Deep layers reduced accuracy. I figured this would help with my overfitting hypothesis. It didn't.

I believe that, given a larger dataset to train from, and songs whose genres better conformed to the genres in the dataset, the NN would have produced more accurate results.