Investor Pitch: 3 Neural Networks for Genre Classification

Camille Knot, Mia Manabat,
Christine Van Kirk

How often do you want to rapidly classify music??

Increasing relevance of music classification.... such as Spotify, Apple Music, and Pandora music recommendation algorithms!

2-Genre MLP Classification

- Data: EchoNest Audio Features
- 8,000 data points total: 4,000 Rock and 4,000 Electronic
  - (Rock & Electronic most frequent in fma dataset)

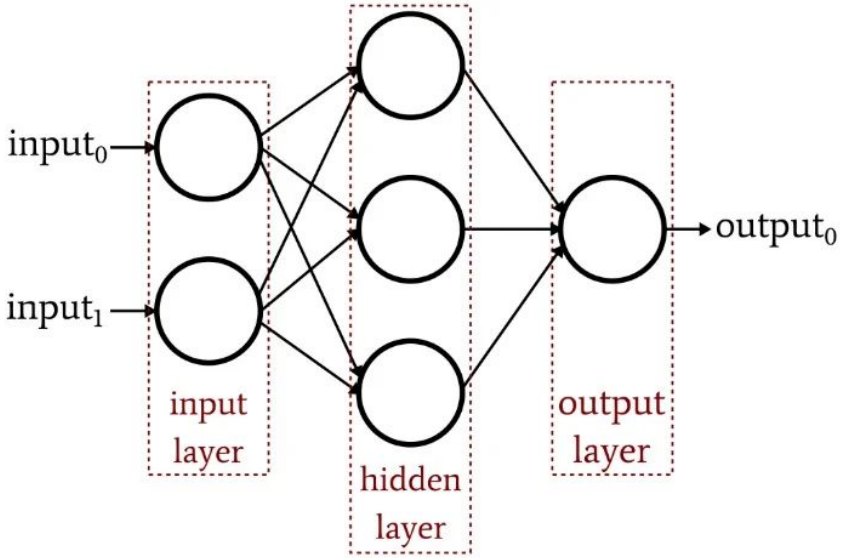
genre	acousticness	danceability	energy	instrumentalness	liveness	speechiness	tempo	valence
R	0.9883055496	0.2556609004	0.9797737929	0.9730057241	0.121342	0.0517397623	90.241	0.0340178672
R	0.9701348869	0.3529461312	0.0238515619	0.9571128852	0.1132607212	0.0321767642	53.758	0.035631561
R	0.9816573443	0.1422492483	0.9121220855	0.9672939775	0.3635095155	0.0875269884	91.912	0.0343252564
R	0.989141175	0.2259775281	0.7228348527	0.2630762793	0.0923705318	0.0534063523	94.322	0.0283472078
R	0.8866597102	0.2985183025	0.7443325625	0.9209504901	0.1395870184	0.0887810414	97.88	0.0735475474
R	0.6982777611	0.2858156458	0.2134937338	0.9556913003	0.0870359278	0.0640938189	125.645	0.1505991359
R	0.8155489042	0.1441250712	0.8927211866	0.9004301244	0.1047034668	0.1022936576	138.68	0.0349164186
R	0.8421128878	0.2852932287	0.5646886968	0.9516243456	0.1104813781	0.0406105474	166.552	0.254299214
R	0.3673926458	0.3296989706	0.8466394286	0.9091861794	0.1989439803	0.4626974726	70.887	0.0482427326

- 8-dimensional inputs, 1-dimensional output

2-Genre MLP Classification

- 8-dimensional inputs, 90-dimensional hidden neurons, 1 output neuron
- Dataset: 60% training, 20% validation, 20% testing
- Loss function: MSE

- **Final testing accuracy: 76% with 0.01 learning rate and 20 epochs**



8-Genre Classification RNN

Model Summary:

```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_6 (LSTM)               (None, 20, 64)            347136

 lstm_7 (LSTM)               (None, 64)                33024

 dense_6 (Dense)             (None, 64)                4160

 dropout_3 (Dropout)         (None, 64)                0

 dense_7 (Dense)             (None, 8)                 520

=================================================================
Total params: 384,840
Trainable params: 384,840
Non-trainable params: 0
_________________________________________________________________
```

8-Genre Classification RNN

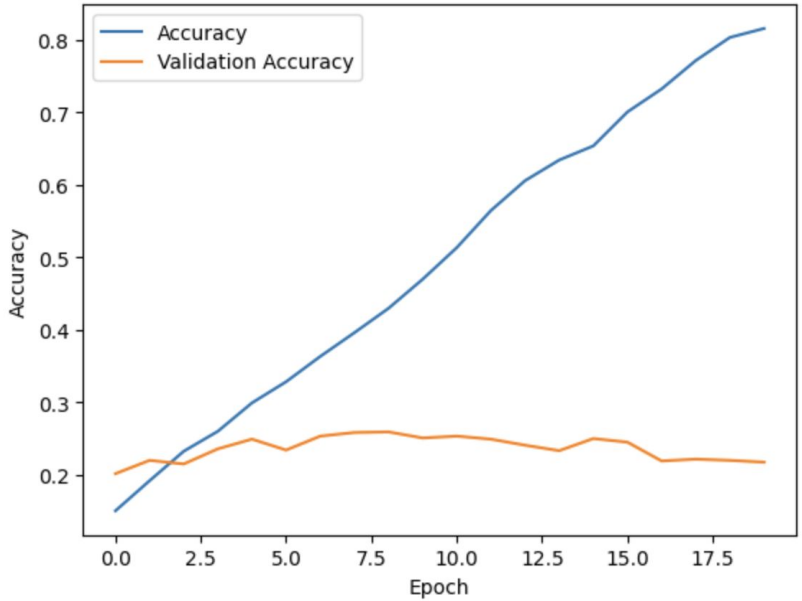
- Final classification accuracy: 28.66%
  - 20 epochs and a 0.001 learning rate
  - Better than random chance! (random chance = ⅛ = 12.5%)

- Accuracy of classifying the correct genre within the top 3 classifications: 60.75%

- *Next slides: Model performance*

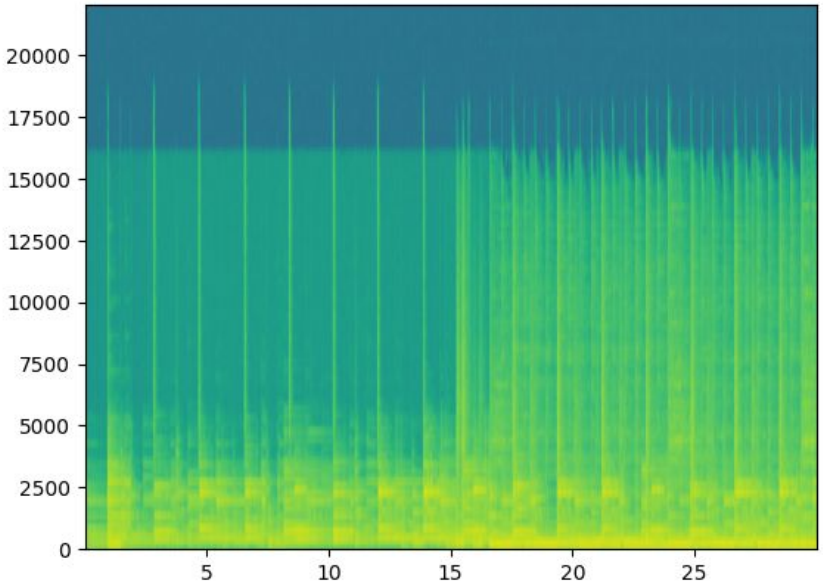
8-Genre Classification RNN

Figure 2. Validation Accuracy vs. Training Accuracy of the RNN Classification Model during Training



8-Genre Classification CNN

- Data: spectrograms converted from mp3 files



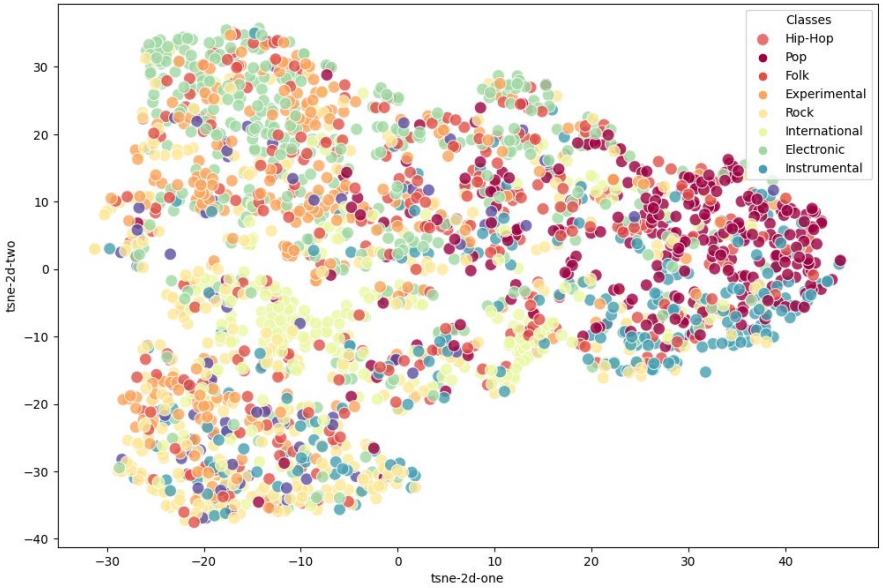
8-Genre Classification CNN

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1         [-1, 8, 496, 369]             224
         MaxPool2d-2         [-1, 8, 248, 184]               0
            Conv2d-3        [-1, 32, 248, 184]           2,336
         MaxPool2d-4         [-1, 32, 124, 92]               0
            Linear-5                [-1, 128]      46,727,296
            Linear-6                  [-1, 8]           1,032
================================================================
Total params: 46,730,888
Trainable params: 46,730,888
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 2.09
Forward/backward pass size (MB): 27.88
Params size (MB): 178.26
Estimated Total Size (MB): 208.24
----------------------------------------------------------------
```

8-Genre Classification CNN

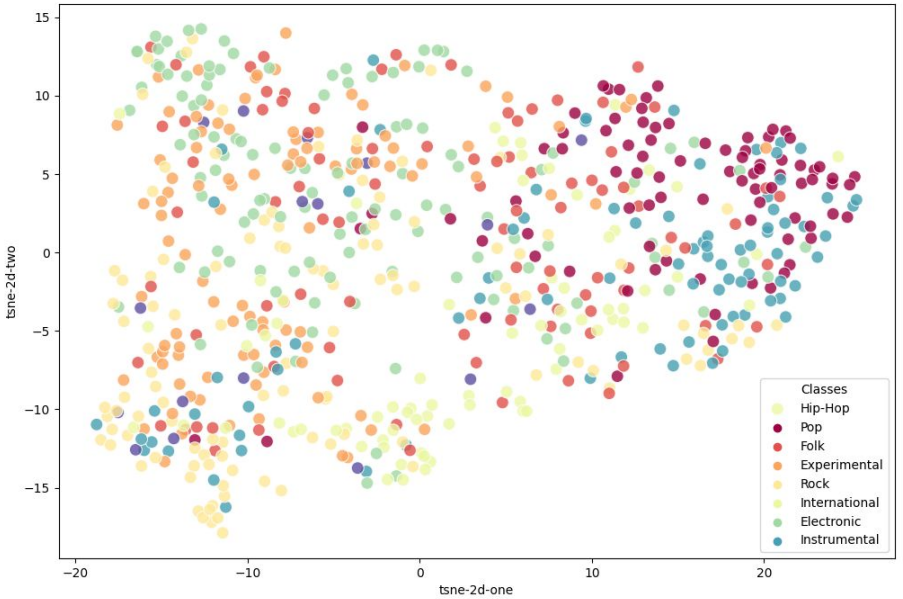
On Training Dataset:

Accuracy: 32.24%



On Testing Dataset:

Accuracy: 37.23%



Our Plans for the Future

- Improving the accuracy of each of these models
- Combining these models to achieve a better model which covers the complexity of the music songs
- Testing on more genres of music!!