

Ronald Pacheco

Final Report: Urban Sound Classification

Problem Statement

Executive Summary

In this project, we are looking to create a deep learning supervised multiclass classification model to successfully classify common sounds of an urban environment. This can be useful in home security systems, where sounds can be identified and classified as burglary sounds (such as glass or door breaking, gunshot, etc), and common urban sounds (such as a car horn or ambulance siren).

Background

This project will involve 5,435 classified audio files distributed amongst 10 classes. Principal stakeholders in this project are Ronald Pacheco and Blake Arensdorf.

Success Criteria

This model will be evaluated with the accuracy metric, and will be successful once a model can accurately classify each sound in the test set in their respective class.

Data Source:

<https://datahack.analyticsvidhya.com/contest/practice-problem-urban-sound-classification/#ProblemStatement>

Data Wrangling

The data consisted of 5,435 observations. The data was separated into audio files, identified by an ID, and a csv file that contained the ID of the audio file and its class.

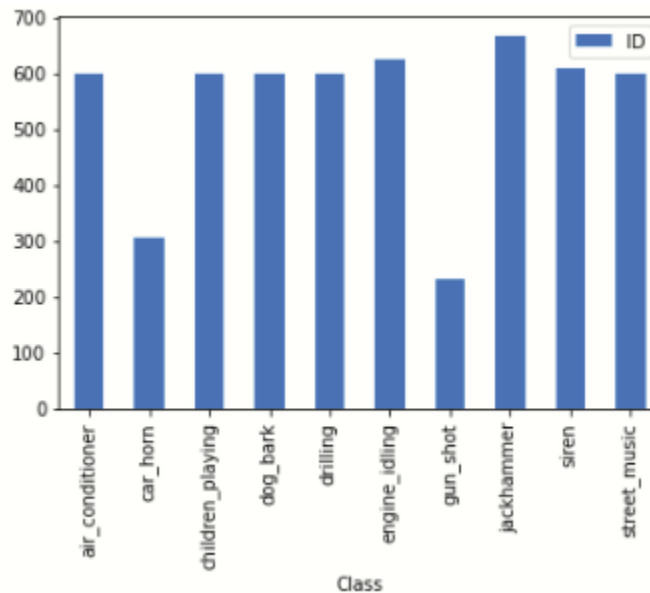
From each audio file, the nine following features were extracted:

- zcr: measures how many times the signal crosses 0
- tonnetz: Computes the tonal centroid features
- melspectrogram: Compute a Mel-scaled power spectrogram
- mfcc: Mel-frequency cepstral coefficients
- chorma_stft: Compute a chromagram from a waveform or power spectrogram
- spectral_contrast: Compute spectral contrast
- spectral_centroid: Indicates at which frequency the energy of a spectrum is centered upon
- spectral_rolloff: Represents the frequency at which high frequencies decline to 0
- spectral_bandwidth: The width of the band of light at one-half the peak maximum

Resulting in a new dataset of 5,435 rows and 185 rows.

Modeling

This dataset presented class imbalance for the classes 'car_horn' and 'gun_shot'.



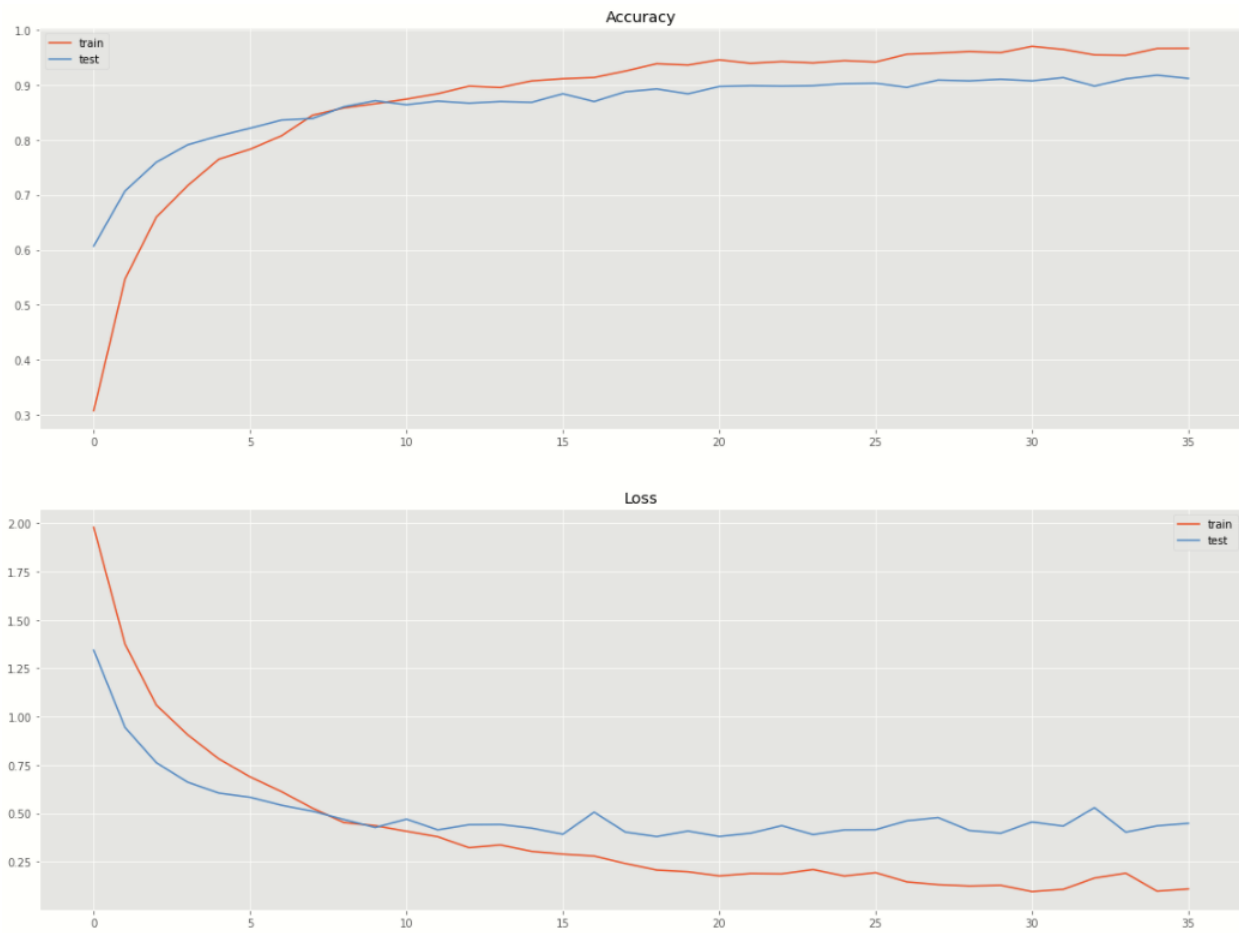
However, we decided to run the neural network with this imbalance to have a benchmark for future improvements.

For this neural network model, we first needed to standardize our features, and create our test and train sets.

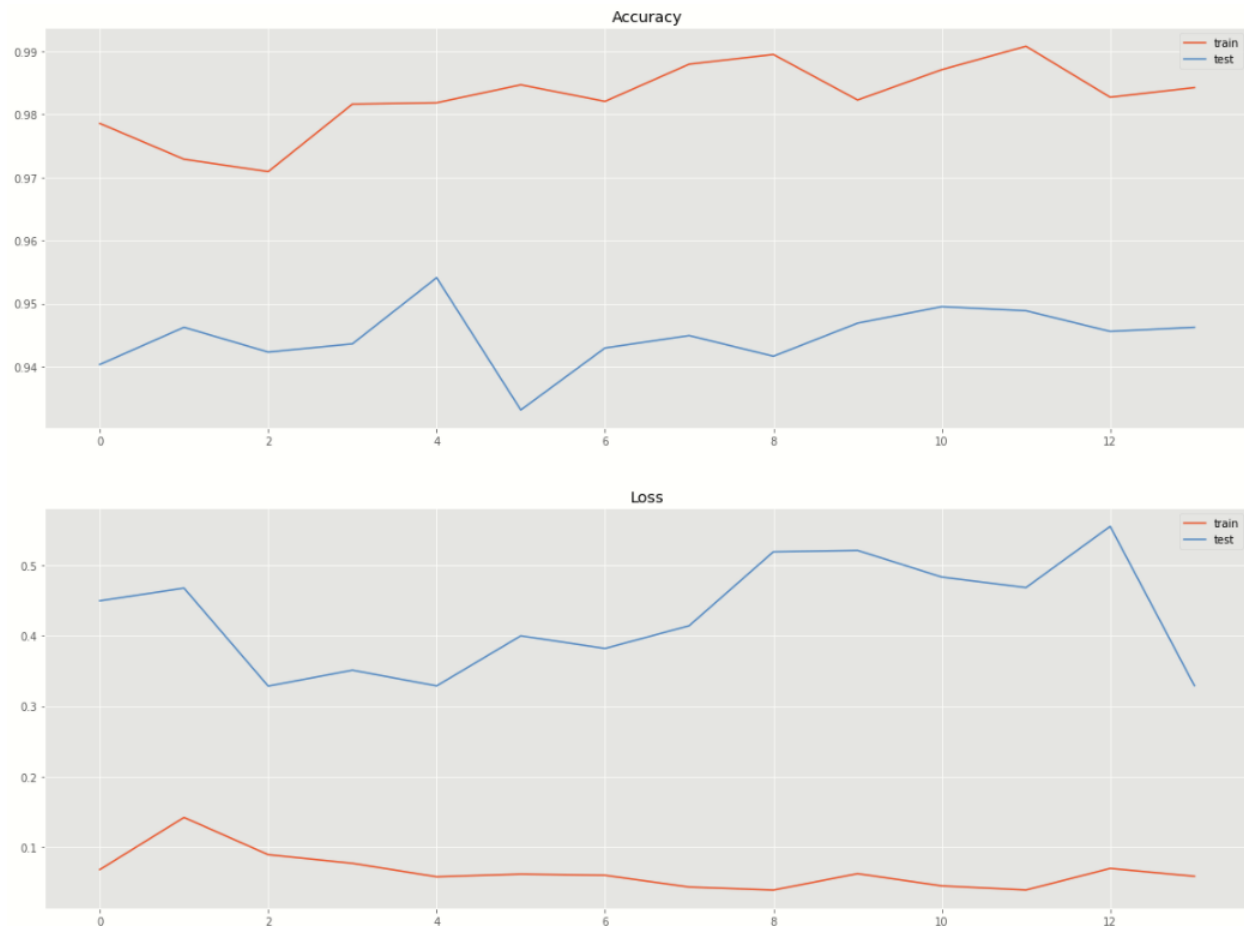
We created a Sequential model with 4 Dense layers and 2 Dropout layers:

```
model = Sequential()
model.add(Dense(182, input_shape=(X_train.shape[1],), activation = 'relu'))
model.add(Dense(256, activation = 'relu'))
model.add(Dropout(0.6))
model.add(Dense(128, activation = 'relu'))
model.add(Dropout(0.5))
model.add(Dense(len(y.columns), activation = 'softmax'))
model.compile(loss='categorical_crossentropy', metrics=['accuracy'], optimizer='adam')
```

Which gave us an initial accuracy of 91.83%:



We then proceeded to balance the data and run it through the same neural network, to which we got an accuracy of 94.03%, an improvement of over 2%:



Future Work

Filters can be applied to the sound files to try and reduce the amount of noise. This would give us a clearer audio that might be easier to correctly classify, improving the model's accuracy.

Recommendations

These classes can be grouped into different categories, which will represent the type of action required.

Based on the predicted class, a decision can be made regarding the type of activity happening in said residence. This can take into account other factors, such as time of day, crime rate in the area, type of crimes, etc.

Conclusions

This project serves as proof of concept for the business case. More audio files (such as glass breaking, door breaking, etc.) need to be gathered in order to have a successful model that can be useful in this application.