# Report for assignment 5

Karkala Shashank Hegde

15th April 2020

# 1 Model Description

From repetitive experimentation, it was found that the most models work well on training data, but have very poor performance on unseen audio files. This showed signs on over-fitting. There during training, validation data was picked from untouched audio files. Therefore to generalize more to unseen audio files, a narrow-deep network was chosen over wider-shallow networks (both combinations were tested). Therefore 8 GRU layers were added in sequence with dropout = 0.5. Also, one-hot encoding with categorical cross entropy gave the best results, thus the model output is 3d with a sofmax at the end. Finally a sequence length of 200(5 seconds) was chosen after experimentaion. The training model diagram used is show in the figure below:

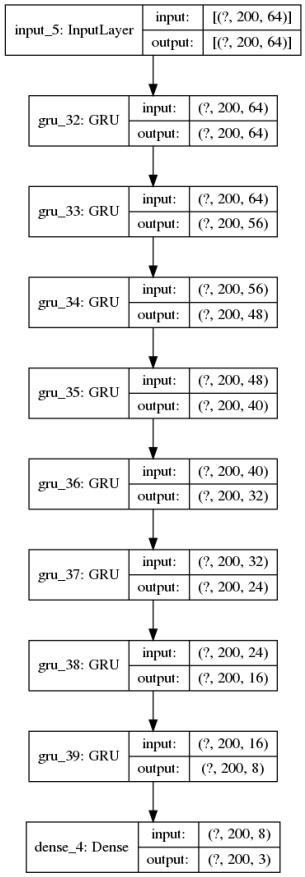


Figure 1: model

The Learning curve shows that the model has generalised reasonably, but there is still decreasing trend in validation accuracy. This could be handled with more data.

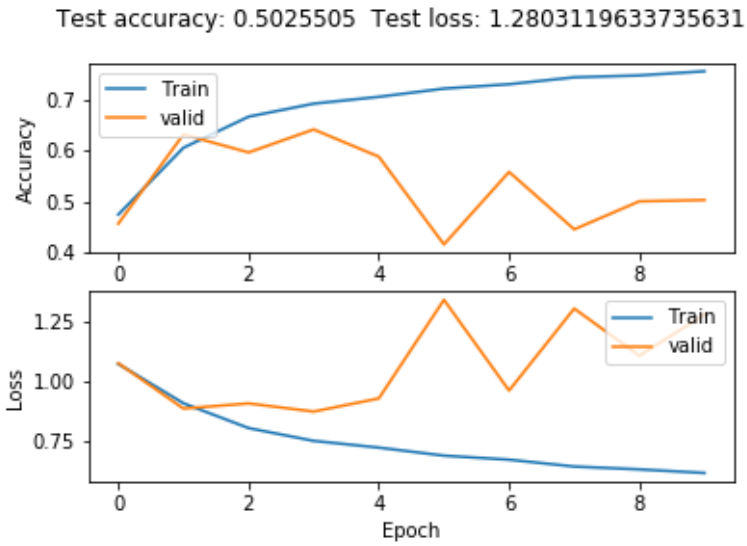


Figure 2: Learning Curve

Removing silence showed a higher performance on validation data. Therefore a db cutoff of 10 was used with librosa's split function to remove silence.

A callback was implemented to save the model that gave highest validation performance. This model is also used for the streaming model. The streaming model is submitted as "model.h5". To see how to use this model, follow the cod e used in the attached python notebook. There is a method called getStreamingModel() that can be called to get the streaming model and streamThroughFile() that can be called to stream the model through a file. Alternatively the "test_streaming_model.py" also contains the same code.

For hyperparameters,

- sequence length: 200 (5 seconds)
- regularization: l1=0.0001, l2=0.01
- loss: categorical cross-entropy
- optimizer: adam
- class imbalence: 30 files from each of the classes were used for training