Data Science Final Capstone Proposal

What is the problem you are attempting to solve?

Implementation of unsupervised and supervised machine learning techniques to classify sound commands. Random Forest, SVM and convolutional neural networks will be used to implement sound classification from a speech dataset of 15000 sounds with 30 classes of sounds. Summary statistics will be extracted from the 15000 raw audio files which consist of 1 s of sound sampled at 44 kHz. Sound classification is useful as a first step in automating voice commands in systems such as alexa and other voice recognition software. For example specific commands can be used to control or interface with a smart home system as opposed to either using an application or computer as the interface.

- How is your solution valuable?

Solutions to this type of problem provide a means to change how humans interact with computers and computing technologies. Voice recognition done with high accuracy will enable a wide range of applications to be developed. For example voice translation, speech to text, system and device control. Specifically the application that I am proposing is the use of classification of voice commands to control smart home systems. The resulting product from this project can be used as the backbone of a software interface that a user may simply speak into to control, lighting and heating levels, security preferences as well as home audio and entertainment systems.

- What is your data source and how will you access it?
- I will be using the google voice command dataset, which is freely available and downloadable through google. The dataset is available at the following website: https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html. The data is contained in a folder structure with 30 distinct folders of 500 audio files each. Each folder contains .wav files that are 1 s in duration sampled at 44 kHz.
- What techniques from the course do you anticipate using?
- I will be using SVM, Random Forests and Convolutional

Neural Networks. The summary statistics of the data will be used as features to be trained on the supervised learning techniques (SVM, Random Forest). The summary statistics include the mean, standard deviation , kurtosis, skewness, power, minimum and maximum values of the sound samples. In the unsupervised case 5 unique models will be used which consist of a 1D convolutional neural network and 4, 2D architectures which will use raw audio samples for the 1D case and short term fourier transformations, wavelet transformations and mel ceptral coefficients for the 2D case respectively. The use of CNN's will enable the extraction of significant features from the time/frequency domain representation of the audio signal. The use of multiple filters and layers will capture global as well as local features unique to the sound class representation.

- What do you anticipate to be the biggest challenge you'll face?
- Determining the correct features to use in the models and tuning the parameters to give the best performance