Comparative Analysis of Classifiers on Audio, Image, and Text Datasets

Akshay Singh Rana 20152453 | **Harmanpreet Singh** 20164950 | **Himanshu Arora** 20152429

Course Code: IFT 6390 | November 12, 2019

1. Abstract

Recently, machine learning has found its applications in a diverse set of domains. Specifically, many real-world decision-making problems fall into the category of classification. This motivated us to explore the classification performance of popular machine learning algorithms on datasets spanning Audio, Image, and Text.

2. Datasets

Audio: We plan to conduct experiments on Freesound Dataset 2018 to automatically recognize sounds from a wide range of real-world environments. It contains 11,073 audio files annotated with 41 labels of the AudioSet Ontology. The task setup is a multi-class classification problem. We will use spectral frequency-based methods for feature extraction from the audio representations.

Image: For image classification, we plan to investigate the CIFAR-10 dataset. The CIFAR-10 dataset consists of 60,000 32x32 color images of 10 classes, with 6,000 images per class. We plan to explore machine learning approaches as well as Convolutional Neural Networks.

Text: We further explore the Sentiment140 dataset for sentiment classification of tweets. It contains 1,600,000 tweets which have been annotated with a positive, neutral, or negative sentiment and thus, can be used for learning sentiment detection. We plan to use machine learning as well as deep sequence-learning based classification methods along with word embeddings for feature extraction.

3. Algorithms

Naive Bayes classifier assumes all the features to be conditionally independent and hence can be extremely fast even on a high-dimensional distribution. In spite of overly simplified assumptions, it seems to perform quite well on real-world situations like text classification problems.

Logistic Regression is a linear classification method that learns the probability of a sample belonging to a certain class. We also plan to examine regularization techniques to avoid overfitting, resulting in a more generalized model.

Support Vector Machine is one of the most powerful classifiers that we will use in all our datasets. It is very effective in high-dimensional space and it can create both linear and non-linear decision boundaries.

Multilayer Perceptron can learn a complex non-linear function approximator using multiple hidden layers and is a great fit for our classification problems. We will do grid-search to explore the optimal hyper parameters like number of neurons, hidden layers, activation functions, etc.

ConvNets & LSTMs are deep neural network architectures that have achieved state-of-the-art results on image and text classification problems. These can prove to be useful in our use case as we have 2D images and sequential audio data.