**CMPE 257 - LAB2**

Suriya Palanikumar

Part 4:

Discuss/explain how you improve the accuracy of using different ML techniques.

***Improving the accuracy of Part 1(Movie Recommender):***

In case of Movie Recommender(Part 1), we didn’t use any classification or regression inorder to predict the accuracy. But in general, on using PCA, the accuracy could be improved in the following ways.

* Picking the number of components which provides us with the variance we expect.
* In order to do so, we can have the number of components as hyperparameters and tune it using cross validation.

***Improving the accuracy of Part 2(Kaggle Competition Dataset):***

In order to improve the accuracy of keras model, the following approaches could be made.(I tried a few inorder to improve the accuracy from 78% to 85%). Please find below the approaches.

* Data Augmentation

Incase of Images as inputs, data Augmentation provides big gain by

randomly shifting and rotating existing images which improves the generalization of the model.

* Rescaling of Data

Rescaling the data(with Normalization or Standardization) in order to

match with the parameter of activation function.

* Data Preprocessing and Transformations

Looking for outliers and transforming the data to the same shape

would help in achieving higher accuracy.

* Regularizations

Neural networks tend to overfit the data without regularization. With

dropout, overfitting problem could be solved.

* Trying out different Optimizers

Trying out different optimizers such as Adam, RMSprop etc., would help to

improve the accuracy.

* Early Stopping

Learning could be stopped once performance started to degrade. This would save a lot of time and also it is a type of regularization which curb the overfitting problem.

* Batch Size

Batch Size could be selected based on the best params in the grid search. Smaller batch size with larger training epochs will provide us with good accuracy.

* Model Performance

Tuning the parameters by checking if the model is underfitting or

overfitting will help in improving the accuracy.

***Improving the accuracy of Part 3(NLP):***

In general, NLP problems can achieve increased accuracy with the following approaches:

* Eliminating noise from text. Because Noisy corpus would probably decrease the accuracy of the model.
* Eliminating words which occur in low frequency.
* Text Preprocessing such as removal of stopwords and punctuation.
* Combination of words using N\_grams provides higher accuracy.
* Lemmatization and Stemming techniques map the words to some root words which would generalize the model.
* Selection of relevant Classification models such as Naive Bayes provides better accuracy.
* Applying different Word embedding techniques would help us to understand the dataset better and hence better accuracy could be achieved.
* Hyperparameter tuning of the classification model such as RandomizedCV and GridCV could help us to improve the accuracy.