ASSIGNMENT 4

CS5304 - PLAYING WITH DEEP LEARNING

1. TASK

In this task, you'll be comparing a Logistic Regression Model with Deep Learning. On the given data set, build both a Logistic Regression Model and a Multi-Layer Perceptron and evaluate their performance. Please report the AUC and ROC curves for both models. In your MLP, make sure that you don't have more than 2 hidden layers and the largest hidden layer is not larger than 500 hidden nodes. This is primarily to ensure that your models converge on time.

Implement these models using TensorFlow and Scikit Learn. We will not be using Spark for this part of the assignment. For extra credit, calibrate the outputs of both these models and present both the pre- and post-calibration Brier loss and Brier curves.

2. DATA

We'll be using an anonymized data set with about 5M rows, 2 class labels and about 20,000 sparse boolean features. The validation and test sets contain roughly 2.5M rows each. You should get an AUC of 0.95 or better using Logistic Regression on this task. Your task is to build a Deep Learning model that is better than this. The data is in SVMLight format and the last 3 characters tell you whether it is training, validation, or test data partition.

3. EXTRA CREDIT

If you provide properly calibrated outputs for both the Logistic Regression and Deep Learning models, you'll get an additional 10 points for this assignment.