

Readme File for InClass Kaggle Competition (MSBD5001)

The code provided is part of the inclass Kaggle competition hosted for MSBD5001 (Fall 2018) students. This competition is about modeling the performance of computer programs. The given dataset describes a few examples of running SGDClassifier in Python. The features of the dataset describes the SGDClassifier as well as the features used to generate the synthetic training data. The task is to predict the training time of the SGDClassifier.

1. Getting Started

This assignment has used tensorflow for the above stated regression problem.

1.1. Data Description

1. **train1.csv** - 440 rows with 14 attributes (including 1 label attribute).
2. **test.csv** - 100 rows of information with 13 attributes for testing.

1.2. Prerequisites

I have used Python3 and its **numpy**, **pandas** and **tensorflow** libraries for this prediction problem. While the others are usually available when using Python3 via Anaconda, **tensorflow** needs to be downloaded separately (*pip install tensorflow*). Please make sure the above mentioned libraries are present before running this code.

1.3. Running the Code

Once all pre-requisites are set, the code can be run by keeping the *train1.csv* and *test.csv* in the same folder as the PY file. The output results will be obtained in the *submission.csv* file.

2. Authors

BANERJEE, Rohini - HKUST Student ID: 20543577

3. References

1. <https://www.kaggle.com/c/msbd5001-fall2018>