

## **Report of Project 1 – MSBD 6000B Deep Learning**

Logistic Regression is used as classifier for the prediction of the test data.

Logistic regression is a regression model where the dependent variable is categorical. Since in this project the output can take only two values (i.e. 0 or 1), logistic regression is one of the ideal choice model.

Realizing that maximizing training accuracy (train and test on the same data) will result an overfitting model, we need to train and test the model on different data. But we also know that a high variance estimate will then result because the accuracy scores may significantly change when the testing data is changed. Therefore, cross-validation is employed for better estimation of out-of-sample accuracy to evaluate the model accuracy. We can calculate the accuracy scores when using different folds as the testing data and takes average of all these testing accuracy as estimation of the out-of-sample accuracy.

The mean of the validation score in this project is 0.92576.