Problem set III

1. Expand the basic code for building a DNN on the Pima Indian Diabetic Dataset to include:

(a) pre-process the data by scaling/standardizing the 8 columns

(b) Split the entire dataset into three parts instead of two as we currently do. One is train, two is validation, and then a test set. Build DNN model with train data, tune hyper-parameters with validation data, and finally evaluate performance on the test data.

(c) Make Epoch versus train set accuracy, and validation set accuracy

(d) Report results using nice ROC curves, report AUC values. Feel free to use code form our course, or from elsewhere

2. Train, tune and evaluate DNN model for the Black Friday sales Dataset. https://www.kaggle.com/sdolezel/black-friday. It’s a regression problem and you need only one neuron in the final layer!!! Try all the above steps (for problem 1) on this one too.