October 10th, 2016
Due Date: October 24th, 2016

Assignment 2

Supervised Learning, Linear/Logistic Regression, SVMs, Neural Networks.

1. You have been hired by a Boston Real Estate firm. You have been given the dataset called Boston house-price dataset (it is available in sklearn). Analyze the data (using whatever plots you consider relevant) and draw conclusions about the different variables in the dataset.
2. In your next task at this new firm, you are asked to create a price-predicting model for this dataset. Using linear regression create a model. Try to prune variables or to transform them to obtain the best results. You can also use regularization.
3. Town Hall is amazed by your good work, and they have asked you to create a classifier using the same dataset. This classifier will set whether a house needs social assistance or not. Your task is to label the data accordingly and choose the best classifier.
4. After all your success, you've decided to come back to Tucson, and help out the Pima Indians. Download the diabetes dataset (https://www.kaggle.com/uciml/pima-indians-diabetes-database) , create a classifier based on the variables, and write some recommendation based on your findings.

5. In class, we saw how to implement an autoencoder from scratch. All the code is available at https://github.com/leonpalafox/MLClass/tree/master/Chapter5NNs/AutoEncoderLibrary . Download it, run it and try different configurations to get the best results in the final classification.