Machine Learning Task 1

(1) Be familiar with PCA and linear regression algorithms.

(2) Complete the programming tasks introduced in tutorials 1-5.

(3) Write a linear regression algorithm by yourself.

- Use Boston House Dataset (506 samples and 13 feature variables), predict the value of prices of the house using the given features.

- Split input data into training and testing sets (the testing set includes 10% of the samples).

- The output of the algorithm should include the model learned from the whole training set, the average mean squared error (MSE) on training set via 10-fold cross validation. For ridge regression model, determine the value of hyper-parameter via 10-fold cross validation.

(4) Write the experimental process and results into an experimental report. (strict content, abundant pictures and words).

- See if there are any missing values in the data.

- Exploratory data analysis (e.g., plot the distribution of the target variable, create a correlation matrix that measures the linear relationships between the variables and plot the correlation matrix, show the mean squared error (MSE) on testing set, plot a scatter plot between the original house price and predicted house prices).

(5) The source code should be submitted as an appendix (Just submit the source code of the regression algorithm).