**HW1**

**Advanced Business Analytics**

**Instructions:**

1. **You are required to submit the Jupyter notebook that you used to get the results.**
2. **You should clearly state the deliverables in your analysis. Deliverables are highlighted in red**
3. **[30 in Total] Practice Goal: OLS estimation/Cross validation/ MLE estimation for linear model**

Dataset “hw1\_housing\_data” contains the housing value information of owner-occupied homes in different areas of Boston along with some characteristics of those areas. The description of the variables in this data is as follows:

|  |  |
| --- | --- |
| CRIM | Per capita crime rate by town |
| CHAS | Charles River dummy variable (= 1 if tract bounds river; 0 otherwise) |
| NOX | Nitric oxides concentration (parts per 10 million) |
| RM | Average number of rooms per dwelling |
| AGE | The proportion of owner-occupied units built prior to 1940 |
| DIS | Weighted distances to five Boston employment centers |
| PTRATIO | The pupil-teacher ratio by town |
| MEDV | Median value of owner-occupied homes in $1000's |

We want to use those characteristics to predict the price of the housing. The model you use for prediction follows

,

1. Use observation ID 1-400 (you will use the last 100 data points as holdout sample for cross-validation) to estimate this model using Ordinary Least Square (OLS) method and report the coefficients and standard errors of the coefficients. [8]
2. Explain the interpretation of , , based on your estimation in part 1. [6]
3. Predict the values for the last 100 data points using the model you estimated in the last part. Plot the actual values versus predicted values. Comment on how well the model does for out of sample prediction [5]
4. Estimate the model using Maximum Likelihood Estimation method (MLE) using observation ID 1-400 and report the coefficients. When using the minimize() function, choose method to be SLSQP. Comment on how close the coefficients are from OLS and MLE methods [11]

1. **[15 in Total] Practice Goal: Law of total probability/Bayes theorem/Consumer heterogeneity**

Suppose you run a fast-food restaurant and you expect that the number of minutes between two drive-through customers follows exponential distribution:

Where is the parameter in this distribution and gives the number of minutes between two drive-through customers.

You expect that can be different depending on whether it is lunch/dinner time or not. Thus, you expect that can be either (lunch/dinner time) or with (other time). Your restaurant opens for 10 hours a day and lunch/dinner time would be 4 hours a day.

1. During lunch/dinner time, is it more likely to observe or ? Show your derivation clearly. [5]
2. Without knowing whether it is lunch/dinner time or not, is it more likely to observe or ? Show your derivation clearly. [5]
3. The number of minutes between two drive-through customers is 2.1, what is the probability of this happened in the lunch/dinner time? Show your derivation clearly. [5]