Statistical Learning EE 2102575
Chulalongkorn University
Semester II, 2023
Suwichaya Suwanwimolkul, Ph.D.

Homework 1 (25 scores)

1 Question 1

From Lecture 2, BLUE —
Derive the Fisher information that $\boldsymbol{y}$ carries about $\boldsymbol{\beta}$, if $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 I)$. (5 scores)

2 Question 2

Goal: Implement the IRLS algorithm

Here I provide the homework on Github Page ....
** After you finish, please submit your code alongside with your report online on MS-Team**

2.1 Your tasks (20 scores)

1. Implement WLS using the analytical formulation (Hint! See page 18/31 in lecture 3).... (6 scores)
2. Implement IRLS (Hint! see page 26-25 in lecture 3).... (6 scores)
3. The predicted IRLS produces the better testing results than OLS's on the following metrics: MAE, MSE, RMSE, R^2 Score (4 scores).
4. Provide the reasonable scatter plots comparing OLS, WLS, and IRLS (2 score).
5. Provide the reasonable scatter plots with the lines showing the general trend of the predictions OLS, WLS, and IRLS (2 score).

2.2 How to submit your python code

1. Please implement your work in the provided '.ipynb' file and save it with your id-name, *e.g.*, **6470160121-Somchai.ipynb**... so as your report **6470160121-Somchai-HW1.pdf**.
2. **Save each of your figures in '.png' format** with a name corresponding to each question...For example, in responding to **Question 1**, you plot a **histogram** of a prior to **verify** that it is a **Gamma** distribution, please name your file **Q1-Histogram-Verified-Gamma.png**; otherwise, your score could be missing.

```
print("================= OLS ========================")
print("MAE:",metrics.mean_absolute_error(y_test,y_pred_OLS))
print("MSE:",metrics.mean_squared_error(y_test,y_pred_OLS))
print("RMSE:",np.sqrt(metrics.mean_squared_error(y_test,y_pred_OLS)))
print("R2score:",metrics.r2_score(y_test,y_pred_OLS))
```

✓ 0.0s

```
================= OLS ========================
MAE: 0.9992677375917789
MSE: 2.1868189190905087
RMSE: 1.4787896804787721
R2score: 0.9134869106859319
```

Figure 1: Performance of OLS on Vehicle Dataset

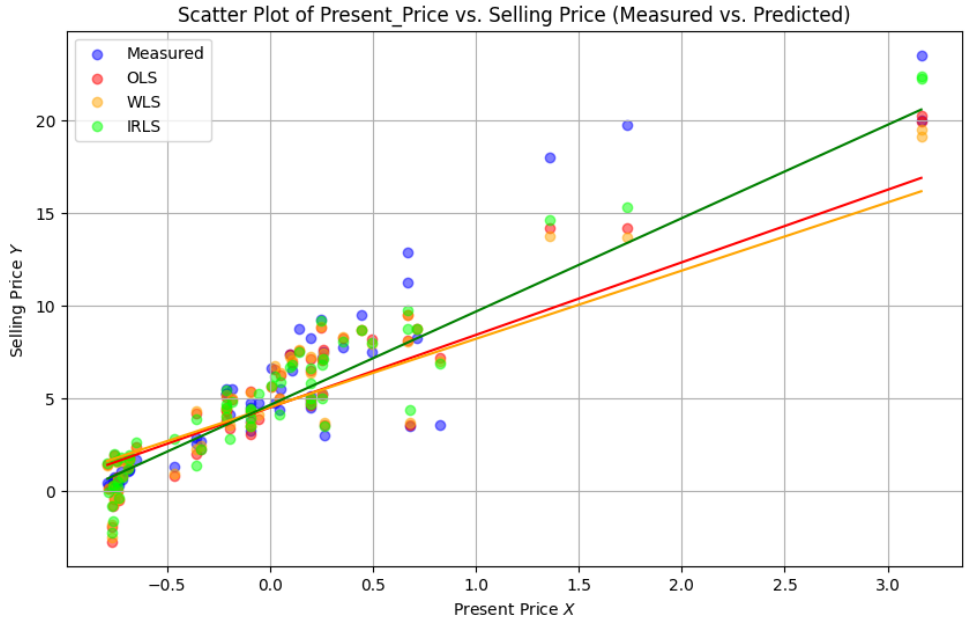

Figure 2: Hint of the performance of IRLS. It is expected to provide the estimation closer to GT than OLS.

3. I will **start marking** your 'ipynb' file, **if and only if** there are **plots and output showing**, and **without any error showing**.

4. Scores in each section will be given, **if and only if** your code can run successfully and correctly without any bugs.

5. To ensure above problems will not happen, you should provide a line to install all the required environments on the top of your 'ipynb' file. For example ...

   ```
   ! pip install numpy pandas tqdm matplotlib statsmodels seaborn  kaggle
   ```

6. The ipynb file should be submitted it alongside with your report online on MS-Team.

7. You can write in markdowns or attach figures to explain your python code ...

8. Please also copy and paste your codes, plots, and with explanation into your report.