2023 Spring, INE5008 Machine Learning with Data Mining, by Kichun Lee

HW 2

Due 1 pm, April 5, 2023

**PART #1**

Refer to the attached *data\_1.txt*. Build a linear regression model as follows:

,

where is an independent and identically distributed (nice) noise.

Then, numerically compute the expectation and standard deviation of the predicted value, , for the first row of the input dataset using 2000 bootstrap datasets. (Either R or python scripts will be okay.)

**PART #2**

Write down a script in Python or R to compute MAPE (Mean Absolute Percentage Error) with two parameters of ground-truth Y and predicted Y.

**PART #3**

(open question) Suppose you are fitting your input data with a polynomial regression. The order of the polynomial regression model is a major parameter you should choose. Describe how to find a proper order of the polynomial regression model.

**PART #4**

(open question) Describe, at least 3, strategies to avoid both underfitting and overfitting.