10920ESS427000 Introduction to Machine Learning

Homework 3: Simple regression method & Tuning model complexity

Announced: **April 22**

Deadline: **May 6 at 11:59 pm(eeclass)**

TA: 邱瑞崑 {chiu.steve98@gmail.com}

Rule:

- Built-in machine learning libraries(Sklearn) is not allowed to use for this homework, you need to use basic mathematical operations in NumPy to do regression.
- You can use any libraries to read the data from the given dataset.
- Discussions are encouraged, but plagiarism is strictly prohibited and punishable!

Submission:

You should compress your code, dataset and report(hw3report_studentID.pdf) into a ZIP file(hw3_studentID.zip), and submit it on eeclass before the deadline. No follow the requirements will minus 10% of your homework.

Implementation[70%]

1. *Dataset Loading & splitting*[10%]:
   You need to split the original dataset into training data and testing data. (proportion: 7/3)
2. *Plot scatter figure including training data and testing data*[5%]:
   Just consider X2 house age, X3 distance to the nearest MRT station and X4 number of convenience stores corresponding to Y house price of unit area. (using blue color as training data, red color as testing data)
3. *Define loss function*(Mean Square Error)[5%]:
   You need to use basic mathematical operations in NumPy to define the loss function and use it in gradient method and least square method.
4. *Using gradient method*[15%]:
   Assume the model function is $Y = \beta_0 + \beta_1 X_2 + \beta_2 X_3 + \beta_3 X_4$, you need to find the parameter $\beta_0, \beta_1, \beta_2, \beta_3$ and test 10 times. Also, you should follow the setting: 7000 iterations, 100 iterations per epoch and print training loss and testing loss every epoch.
5. *Using least square method*[35%]:
   Assume there are eight model function as
   $Y = \beta_0' + \beta_1' X_2 + \beta_2' X_3 + \beta_3' X_4$,
   $Y = \beta_4 + \beta_5 X_2 + \beta_6 X_2^2 + \beta_7 X_3 + \beta_8 X_4$,

$Y = \beta_9 + \beta_{10}X_2 + \beta_{11}X_3 + \beta_{12}X_3^2 + \beta_{13}X_4$,

$Y = \beta_{14} + \beta_{15}X_2 + \beta_{16}X_3 + \beta_{17}X_4 + \beta_{18}X_4^2$,

$Y = \beta_{19} + \beta_{20}X_2 + \beta_{21}X_2^2 + \beta_{22}X_3 + \beta_{23}X_3^2 + \beta_{24}X_4$,

$Y = \beta_{25} + \beta_{26}X_2 + \beta_{27}X_2^2 + \beta_{28}X_3 + \beta_{29}X_4 + \beta_{30}X_4^2$,

$Y = \beta_{31} + \beta_{32}X_2 + \beta_{33}X_3 + \beta_{34}X_3^2 + \beta_{35}X_4 + \beta_{36}X_4^2$ and

$Y = \beta_{37} + \beta_{38}X_2 + \beta_{39}X_2^2 + \beta_{40}X_3 + \beta_{41}X_3^2 + \beta_{42}X_4 + \beta_{43}X_4^2$, you need to find all the parameters and get the training loss and testing loss for each model function.

Report[30%]

The format is not limited, but the following matters must be discussed in your report:

1. Plot scatter figure just including training data and testing data. (using blue color as training data, red color as testing data)(3 Figures) [5%]
2. Plot the figure including training data, testing data and model function based on gradient method. (using blue color as training data, red color as testing data)(3 Figures) [5%]
3. Is $(\beta_0, \beta_1, \beta_2, \beta_3)$ equal to $(\beta_0', \beta_1', \beta_2', \beta_3')$? Explain?[5%]
4. In least square method, which model function you will choose to predict Y house price of unit area? And explain why the model function you choose is better? Also, plot the figure including training data, testing data and the best model function in least square method. (using blue color as training data, red color as testing data)(3 Figures) [15%]