

FE520 Assignment 5

Dan Wang, Zhiyuan Yao

November 2018

1 Linear Regression Practice (40 Pts)

1. Create two random arrays x_1 and x_2 , the values of array are from 0 to 100, and its size = 1000, sort these arrays from small to large.
2. Create the corresponding y , $y = x_1 * 3 + x_2 * 4 + \epsilon$, where $\epsilon \sim N(0, 2)$.
3. Combine x_1 and x_2 as x using pandas or numpy.
4. Let the combining x as your input x , y is your target response to x . Solve the coefficients Θ using mathematics way and output the coefficients. (Referring to my manuscript)
5. Using sklearn linear regression model to solve the coefficients. Compare the difference with last question.

2 Logic Regression Practice (40 Pts)

1. Look at the documentation of ([make_classification](#)) , to make a dataset with binary class, sample size = 1000.
2. Randomly set 80% of your data set as training set, and the rest as your test set.
3. Training your data set using logic regression.
4. Test your model using the trained model with your test set, and output the accuracy.
5. (Bonus 10pts) Data visualization: plot (scatter plot) different classes using different colors, plot the regression to cut off the two classes.
6. (Bonus 10 pts) Using mathematics method to compute the coefficients which is derived in my uploaded paper.

3 Softmax Regression Practice (20 Pts)

Change the two classes in Q2 into 10 classes, and repeat the steps in Question 2.

4 More Practice in sklearn (0 pt)

Due to time limitation, we can't cover most of detail in class. You are encouraged to learn more about those algorithms from online resources like books, blogs, and MOOCs. Then practise with Python.

Actually, the interfaces for implementations of majority of machine learning algorithms in scikit-learn have the similar processes with linear regression and logistic regression:

1. Processing the data (dividing the training data and testing data)
2. Importing different models from sklearn, (i.e. SVM, SVR, Decision Tree, Random Forest, etc.)
3. Training your model.
4. Test your model.
5. See the performance of different classifiers and regressors.

Submission Requirement:

For all the problems in this assignment you need to design and use Python 3, output and present the results in nicely format. Please submit a written report (pdf), where you detail your results and copy your code into an Appendix. You are required to submit a single python file and a brief report and the output as csv format. Your grade will be evaluated by combination of report and code. You are strongly encouraged to write comment for your code, because it is a convention to have your code documented all the time. In your python file, you need contain both function and test part of function. Python script must be a '.py' script, Jupyter notebook '.ipynb' is not allowed. Do NOT copy and paste from others, all homework will be firstly checked by plagiarism detection tool.