CS7052-Machine Learning

Workshop 4: Linear Models for Regression and Classification

You will learn:

- To practice Linear models for Regression and classification with several datasets
- Simple linear regression, Ridge regression, Lasso regression, Logistic regression
- To learn more about data visualisation using matplotlib

Task 1- Linear models for regression and classification

Open Muller & Guido's book from the reading list. Open your Jupyter notebook.

Follow the instructions on pages 47-70 in chapter 2.

Muller and Guido's book comes with accompanying code, which you can find on https://github.com/amueller/introduction_to_ml_with_python.

You can download the code and then open corresponding file to chapter 2 (02-supervised-learning.ipynb) in your Jupyter Notebook.

Make sure you understand the meaning of each line of code, make some changes to improve your understanding.

Questions:

W4.1. What is the training set score of Ridge Regression for extended Boston dataset when alpha is 2?

W4.2. What is the test score of Ridge Regression for extended Boston dataset when alpha is 2?

W4.3. How many features are used when using Lasso Regression with alpha = 0.1 for linear regression of the extended Boston dataset?

W4.4. How many features are used in Ridge Regression for extended Boston dataset?

W4.5. Is that all the number of features?

W4.6. What is the training accuracy of Logistic Regression on cancer dataset when c= 10?

W4.7. What is the test accuracy of Logistic Regression on cancer dataset when c= 10?

Task 2 - To learn more about data visualisation using matplotlib

Refer to chapter 7 of Nelli's book and draw a plot similar to My First Plot in figure 7.17 with a fifth data point labled as 'fifth'.

W4.8. Upload your code to draw the above plot.

Show the output to your tutor when you are done.