EE655000 Machine learning HW1

TA: Kang-yu Liu, Chao-yi Peng, Jun-Rei Wu

asdfghj49888@gmail.com
sky135410@yahoo.com.tw
heathentw@gmail.com

Grading policy

- In the handwriting assignment, you need to provide detailed derivations. Partial points will be credited when a wrong answer is accompanied by correct reasoning.

- In the programing assignment, the code, test data and report should be compressed into a **ZIP** file and upload to iLMS website. Also, please write a Readme file to explain how to run your code and discuss characteristics in your report. The Report format is not limited.

- The programming language that can be used on this assignment includes Python and Matlab.

- Discussions are encouraged, **but plagiarism is strictly prohibited.**

Part1. Handwriting homework assignment

You can find the corresponding problems from the textbook.

1. (30 points)
   Exercise 2.26
2. (40 points)
   Exercise 3.6
3. (30 points)
   Exercise 3.11

Part2. Computer assignment

It is known that different types of flower has different sepal and petal. Iris dataset [1] record the characteristic length of different Iris. The data set contains 3 classes of 50 instances each.

In this problem, you will implement the maximum a posteriori probability (MAP) of the classifier for 30 plants with their sepal information and petal information. Testing data is provided in the **testing.mat**.

Information:

- S_l : Sepal length in cm
- S_w : Sepal width in cm
- P_l : Petal length in cm
- P_w : Petal width in cm
- $label$: $\begin{cases} 0 \text{ , } Iris-setosa \\ 1 \text{ , } Iris-versicolor \\ 2 \text{ , } Iris-virginica \end{cases}$

Assume that S_l, S_w, P_l, and P_w are independent and the distribution of them is Gaussian distribution.

1. (80 points)
   To evaluate the posterior probabilities, you need to learn likelihood functions and prior distribution from the training dataset (**training.mat**). Then, you should calculate the error rate of the MAP detector by comparing to the label of each plant in the test data.
   (Please add corresponding comments in your code to describe how you obtain the posterior probability.)

2. (20 points)
   Please discuss characteristics and plot the visualized result of testing data in your report.

[1]http://archive.ics.uci.edu/ml/datasets/iris