Machine Learning HW1

TA: 邵柏翔 b99872599@gmail.com
張君豪 howard88315@gapp.nthu.edu.tw

Deadline: 2023/03/20 (WED.) 23:59

Grading Policy :

1. In the handwriting assignment, please submit the pdf file. (HW1_student_id_ Handwriting.pdf)
2. In the programming assignment, the code, train.csv, test.csv and report (HW1_student_id_ Programming.pdf) should be compressed into a ZIP file and uploaded to eeclass 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.
3. You are required to finish this homework with Python 3. Moreover, built-in machine learning libraries or functions (like sklearn.linear_model) are NOT allowed to use.
   You can use dimension reduction functions such as *sklearn.decomposition.PCA* for better visualization in discussion.
4. Discussions are encouraged, but plagiarism is strictly prohibited (changing variable names, etc.). You can use any open source with clearly mentioned in your report. If there is any plagiarism, you will get 0 in this homework.

Submission :

Please follow the following format and naming rules when submitting files.

1. HW1_student_id_ Handwriting.pdf
2. HW1_student_id.zip

```
|----HW1_student_id_ Programming.pdf
|----Readme.txt
|----HW1.py (only .py)
|----train.csv
|----test.csv
```

- You need to upload HW1_student_id_ Handwriting.pdf and HW1_student_id.zip to eeclass website.

Part 1. Handwriting assignment :

1. (10%)

   According to Eq.1 and Eq.2 please show that

$$\mathbb{E}[x_n x_m] = \mu^2 + I_{nm}\sigma^2 \tag{3}$$

   x_n and x_m are two data points which sampled from a Gaussian distribution with mean μ and variance σ^2, and I_{nm} = 1 if n=m otherwise I_{nm} = 0. Hence prove the results Eq.4 and Eq.5 .

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x \, \mathrm{d}x = \mu. \tag{1}$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x^2 \, \mathrm{d}x = \mu^2 + \sigma^2. \tag{2}$$

$$\mathbb{E}[\mu_{\mathrm{ML}}] = \mu \tag{4}$$

$$\mathbb{E}[\sigma^2_{\mathrm{ML}}] = \left(\frac{N-1}{N}\right)\sigma^2 \tag{5}$$

2. (10%)

   Let a and b be two independent random vectors, so that p(a, b) = p(a)p(b). Show that the mean of their sum y = a+b is given by the sum of the means of each of the variable separately. Also show that the covariance matrix of y is given by the sum of the covariance matrices of a and b.

3. (10%)

   We know that as the size of the dataset increased, the uncertainty associated with the posterior distribution of the model parameters will be reduced. Make use of the matrix identity

$$\left(\mathbf{M} + \mathbf{v}\mathbf{v}^{\mathrm{T}}\right)^{-1} = \mathbf{M}^{-1} - \frac{\left(\mathbf{M}^{-1}\mathbf{v}\right)\left(\mathbf{v}^{\mathrm{T}}\mathbf{M}^{-1}\right)}{1 + \mathbf{v}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{v}}$$

   to show that the uncertainty σ^2_N(x) associated with the linear regression function given by (3.59) satisfies

$$\sigma^2_{N+1}(\mathbf{x}) \leqslant \sigma^2_N(\mathbf{x})$$

4. (10%)

If we have a linear model like

$$y(x, \mathbf{w}) = w_0 + \sum_{i=1}^{D} w_i x_i$$

And also have a sum-of-squares error function of the form

$$E_D(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2$$

Now suppose a Gaussian noise $\in_i$ with zero mean and variance σ^2 is added to each input variable x_i separately. By utilizing E[$\in_i$] = 0 and E[$\in_i \in_j$] = $\delta_{ij}\sigma^2$ to show that minimizing ED averaged over the noise distribution is equivalent to minimizing the sum-of-squares error for noise-free input variables with the addition of a weight-decay regularization term, in which the bias parameter w0 is omitted from the regularizer.

Part 2. Programming assignment :

This dataset [1] is the result of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

That is, there are 3 types of wines and 13 different features of each instance. In this problem, you will implement the Maximum A Posteriori probability (MAP) of the classifier for 60 instances with their features.

There are a total 483 instances in wine.csv. The first column is the label (0, 1, 2) of type and other columns are the detailed values of each feature. Information of each feature:

1. Alcohol
2. Malic acid
3. Ash
4. Alcalinity of ash
5. Magnesium
6. Total phenols
7. Flavanoids
8. Non Flavonoid phenols
9. Proanthocyanins
10. Color intensity

11. Hue
12. OD280/OD315 of diluted wines
13. Proline

Assume that all the features are independent and the distribution of them is Gaussian distribution.

1. (5%) Please split wine.csv into training data and test data. When splitting, please randomly select 20 instances of each category as testing data. Then save the training dataset as train.csv and testing dataset as test.csv. (423 instances for training and 60 instances for testing.)

2. (25%) To evaluate the posterior probabilities, you need to learn likelihood functions and prior distribution from the training dataset. Then, you should calculate the accuracy rate of the MAP detector by comparing to the label of each instance in the test data. Note that the accuracy rate will be different depending on the random result of splitting data, but it should exceed 95% overall. (Please screenshot the result and describe how you obtain the posterior probability in your report.)

3. (15%) Please plot the visualized result of testing data in your report. Please briefly describe the role of PCA and how it works in your report. (You can directly use the built-in PCA function to get visualized result.)

4. (15%) Please discuss the effect of prior distribution on the posterior probabilities in your report.

[1] https://archive.ics.uci.edu/ml/datasets/Wine