11320EE 655000 Machine Learning

Homework1

Deadline: Mar. 24, 2025 (Mon.) 23:59

TA: Yanqing Wang
s112061518ee@gapp.nthu.edu.tw

Grading Policy:

1. In the handwriting assignment, please submit the pdf file. (HW1_student_id_Handwriting.pdf)
2. In the programming assignment, the report (HW1_student_id_Programming.pdf), train.csv, test.csv and code (HW1.py) should be compressed into a ZIP file and uploaded to eeclass website. Also, please write a Readme.txt 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. But 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. Assume your student ID is 123456789:

1. HW1_123456789_Handwriting.pdf
2. HW1_123456789.zip
   - HW1_123456789_Programming.pdf
   - Readme.txt
   - HW1.py (only .py)
   - train.csv
   - test.csv

You need to upload HW1_123456789_Handwriting.pdf and HW1_123456789.zip to the eeclass website.

Part 1. Handwriting assignment

1. (10%) According to Eq. (2) and Eq. (3), please show that

$$\mathbb{E}[x_n x_m] = \mu^2 + I_{nm}\,\sigma^2. \tag{1}$$

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

$$\mathbb{E}[x] \;=\; \int_{-\infty}^{\infty} \mathcal{N}\big(x \mid \mu, \sigma^2\big)\, x\, dx \;=\; \mu. \tag{2}$$

$$\mathbb{E}[x^2] \;=\; \int_{-\infty}^{\infty} \mathcal{N}\big(x \mid \mu, \sigma^2\big)\, x^2\, dx \;=\; \mu^2 + \sigma^2. \tag{3}$$

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

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

2. (10%) The uniform distribution for a continuous variable x is defined by

$$\mathrm{U}(x \mid a, b) \;=\; \frac{1}{b-a}, \qquad a \le x \le b. \tag{6}$$

Verify that this distribution is normalized, and derive its mean and variance.

3. (10%) The predictive distribution takes the form

$$p\big(t \mid \mathrm{x}, \mathbf{t}, \alpha, \beta\big) \;=\; \mathcal{N}\Big(t \;\Big|\; \mathbf{m}_N^{\mathrm{T}}\phi(\mathrm{x}),\; \sigma_N^2(\mathrm{x})\Big). \tag{7}$$

where the variance $\sigma_N^2(\mathrm{x})$ of the predictive distribution is given by

$$\sigma_N^2(\mathrm{x}) \;=\; \frac{1}{\beta} \;+\; \phi(\mathrm{x})^{\mathrm{T}}\mathbf{S}_N\,\phi(\mathrm{x}). \tag{8}$$

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

$$\big(\mathrm{M} + \mathrm{v}\,\mathrm{v}^{\mathrm{T}}\big)^{-1} \;=\; \mathrm{M}^{-1} \;-\; \frac{\big(\mathrm{M}^{-1}\mathrm{v}\big)\,\big(\mathrm{v}^{\mathrm{T}}\mathrm{M}^{-1}\big)}{1 \;+\; \mathrm{v}^{\mathrm{T}}\mathrm{M}^{-1}\mathrm{v}}. \tag{9}$$

to show that the uncertainty $\sigma_N^2(\mathrm{x})$ associated with the linear regression function given by Eq. (8) satisfies

$$\sigma_{N+1}^2(\mathrm{x}) \;\le\; \sigma_N^2(\mathrm{x}). \tag{10}$$

4. (10%) The beta distribution, given by Eq. (11), is correctly normalized, so that Eq. (12) holds:

$$\mathrm{Beta}(\mu \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\,\mu^{a-1}\left(1-\mu\right)^{b-1}. \tag{11}$$

$$\int_0^1 \mu^{a-1}\left(1-\mu\right)^{b-1} d\mu = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}. \tag{12}$$

Make use of the result Eq. (12) to show that the mean and variance of the beta distribution Eq. (11) are given respectively by:

$$\mathbb{E}[\mu] = \frac{a}{a+b}. \tag{13}$$

$$\mathrm{var}[\mu] = \frac{a\,b}{\left(a+b\right)^2\left(a+b+1\right)}. \tag{14}$$

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)** classifier for **60 instances** with their features. There are a total of **483 instances** in `wine.csv`. The first column is the label (0, 1, 2) of type and the other columns are each feature's detailed values. Information about 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. (8%) Please split wine.csv into training data and test data and discuss why do we need to split the dataset into training and test dataset. When splitting, please randomly select 20 instances of each category as test data. Then save the training dataset as train.csv and test dataset as test.csv. (423 instances for training and 60 instances for testing.)

2. (20%) 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 it 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 90% overall. (Please screenshot the result and add corresponding comments in your code to describe how you obtain the posterior probability in your report.)

3. (8%) Please plot the PCA visualized result of test data and 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. (8%) Please discuss the effect of prior distribution on the posterior probabilities in your report.

5. (8%) Please discuss how the likelihood of each feature in each wine category influences final classification results. In this question, "contribution" refers to the relative influence on the classification decision results.

6. (8%) The confusion matrix can help us understand the performance of the classifier on different categories. Please calculate and plot the confusion matrix for your test data and briefly discuss your results in your report. An example of confusion matrix visualization result is shown in Fig. 1. (In this question, you can directly use the built-in function to calculate and plot confusion matrix.)

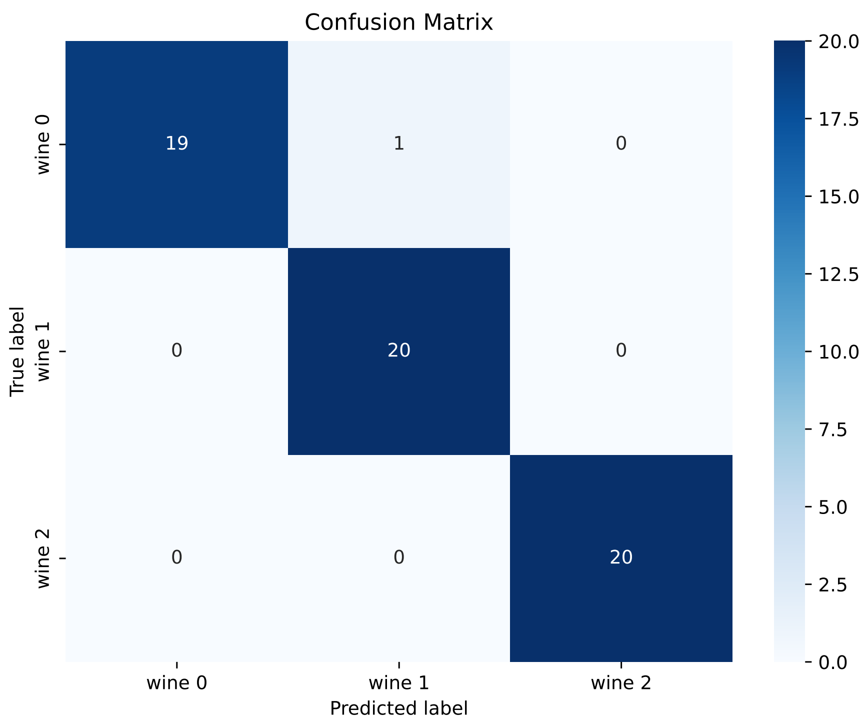

Figure 1: Confusion matrix visualization result

References

[1] Stefan Aeberhard and M. Forina. Wine. UCI Machine Learning Repository, 1992. DOI: https://doi.org/10.24432/C5PC7J.