Homework 1 - Handwriting

Deadline: 2021/10/27 10:10

Grading Policy

In handwriting assignments, you need to provide detailed derivations. Partial points will be given for correct reasoning. Please write your answers and derivations on A4 papers and hand in the homework in the class on 2021/10/27. No late submission will be accepted.

Problems

1. (10%) Derive equation 2.17.

$$w_1 = \frac{\sum_t r^t x^t - \overline{xr} N}{\sum_t (x^t)^2 - N\overline{x}^2}$$

   Hint: Take the derivative of the sum of squared errors with respect to the two parameters, set them equal to 0, and solve these two equations in two unknowns as:

$$E(w_1, w_0 | \mathcal{X}) = \frac{1}{N} \sum_{i=1}^{N} \left[r^t - (w_1 x^t + w_0) \right]^2$$

2. (10%) Assume a disease so rare that it is seen in only one person out of every million. Assume also that we have a test that is effective in that if a person has the disease, there is a 99 percent chance that the test result will be positive; however, the test is not perfect, and there is a one in a thousand chance that the test result will be positive on a healthy person. Assume that a new patient arrives and the test result is positive. What is the probability that the patient has the disease?

3. (10%) Show that as we move an item from the consequent to the antecedent, confidence can never increase: confidence(ABC→D) ≥ confidence (AB→CD).

4. (10%) Let us say that we toss a fair coin ten times. What is the probability that we see four to six heads? What is the probability that we see forty to sixty heads when we toss it one hundred times?

5. (10%) Write the log likelihood for a multinomial sample and show equation 4.6.

$$p_i = \frac{\sum_t x_i^t}{N}$$

6. (10%) Show equation 5.11.

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right)\right]$$

Hint: Consider multivariate normal distribution and use covariance matrix as

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \quad \text{with} \quad (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

Homework 1 - Programming

Python Tutorial and Naive Bayes Classification

Deadline: 2021/10/27 23:30

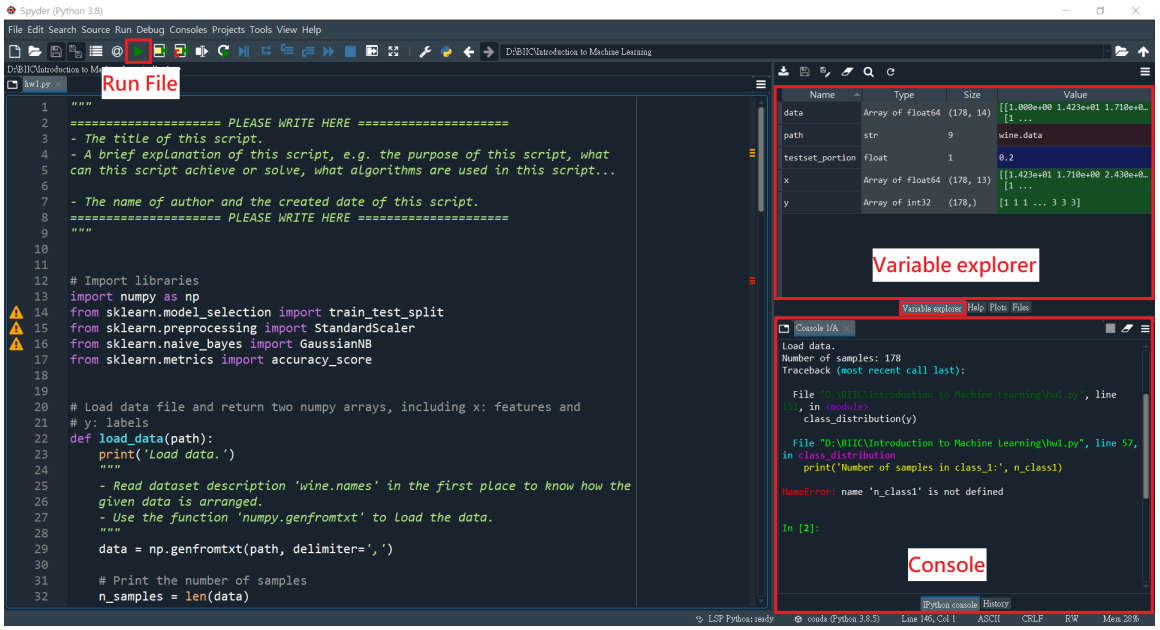
Part I. Python Tutorial

1. Download Anaconda Individual Edition installer and follow the installing instructions.
2. After the installation is complete, launch **Spyder**. Choose **File > Open…** and open the file **hw1.py** in the downloaded folder from eeclass. Then you will see the script.
3. In this homework, you will have to use libraries including *numpy* and *sklearn*. To install python libraries, please launch **Anaconda Prompt** and enter the following command:
   ```
   pip install <library_name>
   ```

   For example,
   ```
   pip install numpy
   ```

   You can also install multiple libraries at the same time:
   ```
   pip install numpy sklearn
   ```

4. After installing required libraries, please go back to **Spyder**. Click **Run File** in **Spyder** and you will see some messages shown in the **Console**.
5. Click **Variable explorer** and you will see the information of variables. Double click any variable and observe its data arrangement.



Part II. Naive Bayes Classification

Grading Policy

The programming section of homework consists of “implementation” and “report.” In the implementation, you have to complete the blocks enclosed by "PLEASE WRITE HERE" comments. If your submitted python file cannot be successfully compiled due to the mis-modification outside of the blocks mentioned above, some points will be deducted. In the report, answers are expected to be stated with your observations followed by explanations. Writing in English or Chinese are acceptable.
Please upload your python file and report named **hw1_StudentID.py** and **hw1_StudentID.pdf** to the eeclass. No late submission will be accepted.

Implementation

1. Open **hw1.py** in Spyder.
2. (1%) Write the information of this script at the beginning.
3. Load the given data file wine.data.
4. (2%) Print the number of samples in each class.
5. (5%) Split the data into training set and testing set.
6. (5%) Standardize the values of each feature dimension.
7. (5%) Train a Naive Bayes classifier.
8. (2%) Use the trained classifier to test the testing set.

Report

1. (10%) Try to apply the parameter **random_state** in the function **sklearn.model_selection.train_test_split** and choose different values to it.
    a. What can the parameter **random_state** do?
    b. Try to assign different values to **random_state** and observe the accuracies. Would **random_state** affect the accuracy? Why?
2. (10%) Try to modify the value of the parameter **testset_portion** in Main. Note that you should fix the **random_state** to do the following experiment.
    a. Assign **testset_portion** with 0.2, 0.5, 0.8, 0.9, 0.98 and compare their corresponding accuracies. What do you observe?
    b. What causes such results?