

C S 487/519 Applied Machine Learning

Basic Python programming

1 Objective

In this *individual* homework, you are required to get familiar with programming using basic Python knowledge.

2 Requirements

2.1 Tasks

Write a python program to do the following,

- (1) (10 points) Read in the Iris dataset using functions in Pandas package. (The Iris dataset can be downloaded from [here](#) and the description of the data can be downloaded from [here](#).)
- (2) (20 points) Calculate and print the number of rows and columns that this dataset contains.
- (3) (20 points) Get all the values of the last column and print the distinct values of the last column.
- (4) (25 points) When the last column has value “Iris-setosa”, calculate the number of rows, the average value of the first column, the maximum value of the second column, and the minimum value of the third column.
- (5) (23 points) Draw a scatter plot with the data of the first column and the second column (y axis represents the second column and x axis represents the first column). Show the points in different colors and shapes when the last columns values are different.
- (6) (2 points) Write a readme file `readme.txt` with detailed instructions to run your program.

2.2 Other requirements

- Your Python code should be written for **Python version 3.5.2 or higher**.
- Please write proper **comments** in your code to help the instructor and teaching assistants to understand it.
- Please properly organize your Python code (e.g., create proper classes, modules).
- You can put your code to Jupyter Notebook or a `.py` file.

3 Submission instructions

Put all your files (Python code, readme file, report, etc.) to a zip file named `hw.zip` and upload it to Canvas.

4 Grading criteria

- (1) The score allocation has been put beside the questions.
- (2) FIVE points will be deducted if files are not submitted in the required format.
- (3) If the total points are more than 100. Your grades will be scaled to the range of [0,100].
- (4) Please make sure that you test your code thoroughly by considering all possible test cases. Your code may be tested using more datasets.