Programming

Midterm Exam

09:10-12:00, Jan. 10, 2020

Student ID & Name:

1. (25%) Answer the following questions:

   (a) Give two simple functions to evaluate the sum of each column for a matrix. (Do not use **for(...)**)

   (b) What is the function that can be used for comparion of computation time?

   (c) Try to use function in the family "APPLY" and the defined function to evaluate the following quantity for each row of a matrix?

   $$\sum_{i=1}^{n} \frac{|y_i - \bar{y}|}{n}, \text{ where } \bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}.$$

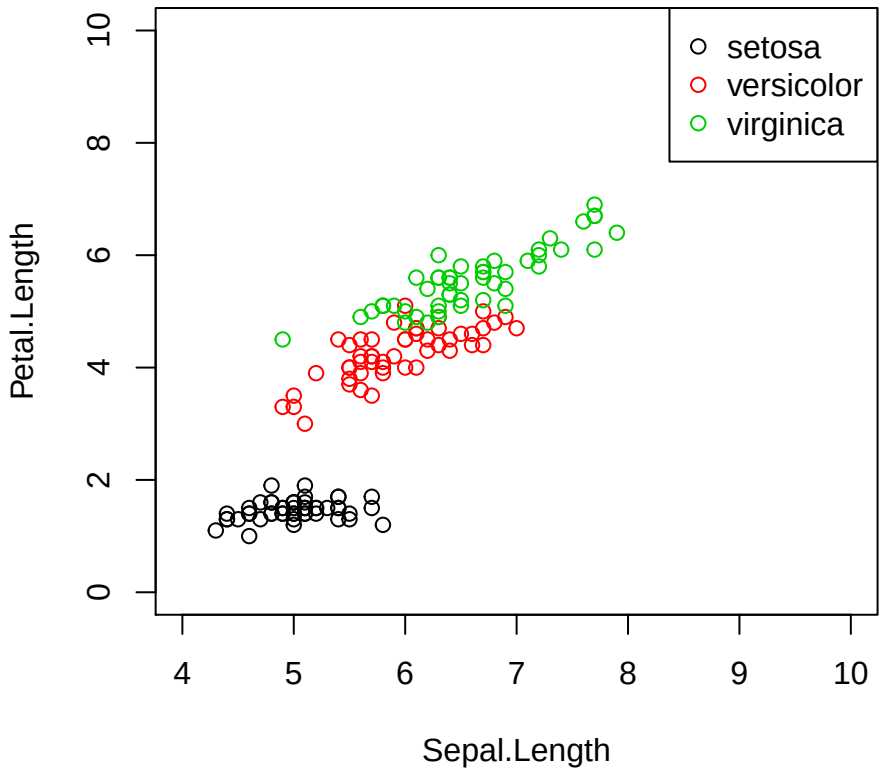
   By using the code, what is the value for each row of a matrix:

```
X <- matrix(c(-0.45 , 0.77,  0.16,  1.22,  1.75,  0.24,
          0.27,  1.74,  0.07, -0.87, -0.90,  1.07,
          -0.06,  0.33, -0.06,  0.09,  0.10, -0.04), 3, 6, byrow = TRUE)
```

   (d) Draw the curve by the following equations and identify what the graph is.

   $$y_1 = \sqrt{(1 - x^2)} \text{ and } y_2 = -\sqrt{(1 - x^2)}.$$

   (e) Use the dataset `iris`. Write a code to draw the plot:



2. (30%) Download the package "titanic" and `library("titanic")`. Use `data(titanic_train)` to the following questions:

   (a) Extract the column names of the dateset.
   (b) How many people are there in the dataset? The column "Survived" is Passenger Survival Indicator with dead 0 and survival 1. How many survived people are there?
   (c) Draw two pie charts for the columns of `Survived` and `Sex`.
   (d) Provide the value of percentage of each gender in both survival group and dead group.
   (e) Try to write a code with `function(...)` to creat the dummy variables for an arbitrary discrete variable.
   (f) Use `titanic_train$Pclass` and `titanic_train$Sex` to test your function in (e). Report the sum of each column of dummy variables.

3. (15%) Let `x <- seq(-1.5, 1.5, 0.01)` and `y <- seq(-1.5, 1.5, 0.01)`.

   (a) • Use a matrix to collect all possible combinations of `x` and `y` into a matrix satisfying

   $$x^2y^3 - (x^2 + y^2 - 1)^3 \geq 0.$$

   (b) Draw a plot including all possible combinations in (a). What is the graph?

4. (30%) Download the package "nlme" and `library(nlme)`. Begin with the data: `data(Fatigue)`. There are three columns, which are "Path", "cycles", and "relLength". **Path** means the ID number of a unit, **cycles** means the measurement time point, and **relLength** means the crack length over cycles. Use the `Fatigue` data to the following questions:

   (a) How many units are there in the dataset?

   (b) Try to plot the degradation path for all of the units, which means the x-axis is the cycles and the y-axis is the relLength. Note that there are n (from (a)) paths in the same plot, and use `type = "b"` in `plot(...)`. (The label names of x-axis and y-axis should be meaningful.)

   (c) Define the failure threshold to be 1.4. Add an horizontal line at 1.4 to the plot in (b) by different format (eg, different style of line or different color).

   (d) Try to provide the number of failures in the dataset during the testing period.

   (e) Use a transformation $y = \log(\text{relLength})$, and fit a linear regression for each path with the following model:

   $$y_i = a_i * t, i = 1, \ldots, n,$$

   where i is the index of the path and t is the measurement cycles. Report all values of a_i for all units and evaluate the mean of a_i.

   ```
   fit <- lm(...)
   fit$coefficients
   ```

   (f) Let the coefficient in (e) to be a random variable. Then, $a_1, a_2, ..., a_n$ are realizations of the random variable. Try to use the one sample t-test to test if the population mean of a_i is equal to 5?