Regression and Classification

Assignment 2

2023-11-11

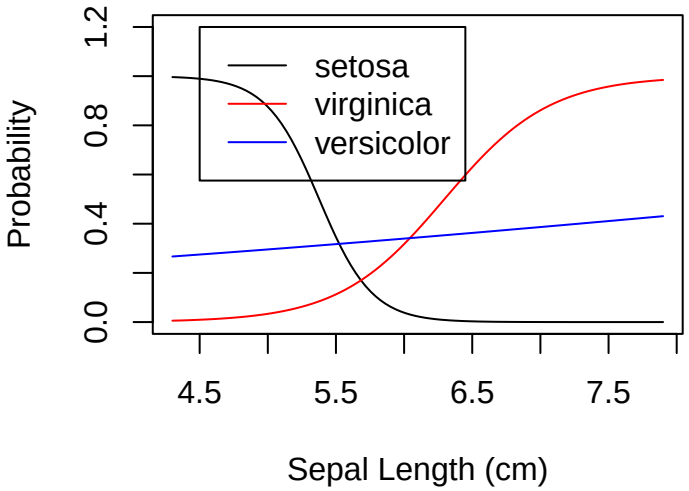
Total Points 10 points

- Do your analysis using **R-markdown**. Compile it as pdf file. Then print it. Staple it and submit it.
- Make sure you write your Name, Email ID and Roll Number on the top of your assignment.
- You can submit it with the front desk security.
- **Deadline** : 19-Nov-2023 - 11:59 pm

Q1 (5 points)

Following logistic regression analysis is based on `iris` data. It tries to predict the species of the flower from the sepal length.

1. If sepal length of a flower is 6.5 cm. From the following analysis can you predict the species of the flower? (2.5 points)
2. Does `Sepal.Length` have any effect on the identification of the versicolor? If you think it does - explain. If you think it does not - explain. However, if you explain both ways, no marks will be given. (2.5 points)



Q2 (5 points)

The linear discriminant rule for class k is presented as:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k,$$

the best decision rule is

$$G(x) = \underset{k}{argmax} \;\; \delta_k(x).$$

Consider the following summary statistics from two (independent) data sets:

$$\bar{X}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \bar{X}_2 = \begin{pmatrix} 5 \\ 7 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad n_1 = 20, \quad n_2 = 40$$

(i) Compute the linear discriminant function. Use that linear discriminant function to classify the point $x = [1 \;\; 4]$ as either π_1 or π_2 (2.5 points)

(ii) Under which theoretical assumptions do you expect your method to be reliable? (2.5 points)

$\log(2) = 0.693$ and $\log(3) = 1.099$