1. Load the dataset “fisheriris” into the workspace. The dataset is provided in the attached Excel file.
2. Study the dataset in terms of

* Number of classes (types of flowers),
* Number of features (number of standard measurements: Sepal Length, Sepal Width, etc.),
* What the data represents, i.e., gain some intuition about the problem domain.

1. Based on your study, would you expect the features to perform well (i.e. to differentiate flower species) in this problem?

Compute the following quantities for each feature.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sepal length | Sepal Width | Petal length | Petal Width |
| Minimum |  |  |  |  |
| Maximum |  |  |  |  |
| Median |  |  |  |  |
| Mean |  |  |  |  |
| Variance |  |  |  |  |
| Within-class variance |  |  |  |  |
| Between-class variance |  |  |  |  |

Do you observe anything of interest from these statistics?

1. Compute and display the correlation coefficients similar to that shown below figure. Do you observe anything interesting from this display?

A red and blue squares

Description automatically generated

1. Display each of the four features versus the class label, similar to that shown below. What can you state about how well the features may perform in classification?

A graph with red lines

Description automatically generated A graph with red lines

Description automatically generated A graph with red lines

Description automatically generated A graph with red lines

Description automatically generated