**Software Project Milestone One**

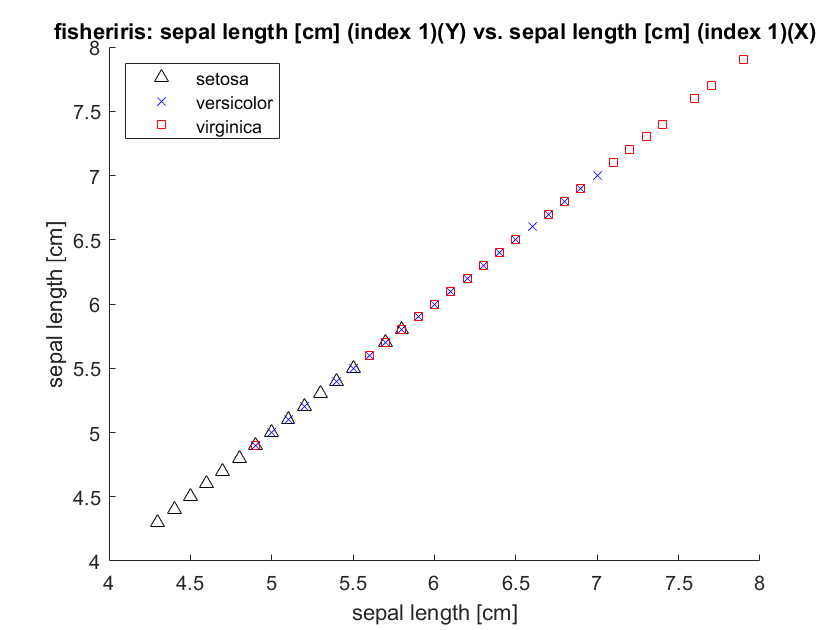
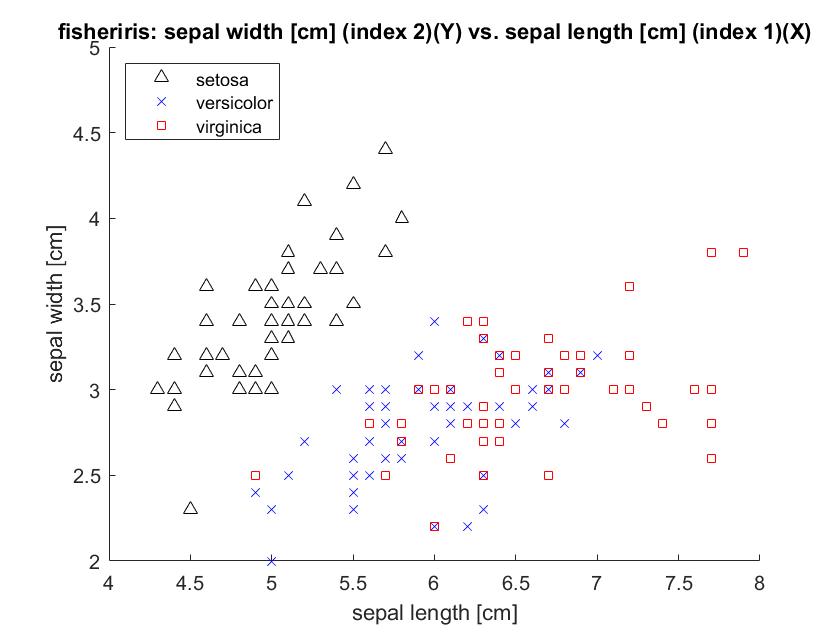
**Milestone one: initial dataset evaluation**

The project includes intermediate milestones. The first milestone involves taking a look at the data using scatter plots (specified as y value vs. x value). For the 4 features in the Iris dataset, provide scatter plots for 1 vs. 1, 2 vs. 3, 3 vs. 1, 4 vs. 1, 2 vs. 2, 2 vs. 3, 2 vs. 4, 3 vs. 3, 3 vs. 4, 4 vs. 4 (10 plots). Use different markers for different classes. Use the plotting ranges in Table A.

**Table A**. Plotting ranges for Iris dataset features.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Lowest value** | **Highest value** |
| Sepal length [cm] | 4 | 8 |
| Sepal width [cm] | 2 | 5 |
| Petal length [cm] | 1 | 7 |
| Petal width [cm] | 0 | 3 |

Make sure you label each axis and provide a legend of what the markers correspond to. The title should at least include the dataset name. Examples of the first two scatter plots are given below in Fig. A. Observe from Fig. Aa that sepal length cannot be used to perfectly separate any two classes because for each pair of classes there exists feature values that are the same.

1. (b)

**Figure A**. Scatter plots for pairs of features in iris dataset.

In addition to providing the 10 scatter plots, answer the following questions with regards to the data.

1. Can you perfectly separate the setosa and versicolor classes using only one feature? If so, which feature or features would work?
2. Can you perfectly separate the setosa and virginica classes using only one feature? If so, which feature or features would work?
3. Can you perfectly separate the versicolor and virginica classes using only one feature? If so, which feature or features would work?

**Solution**

The requested scatter plots are given in Figs. 1-10. In addition, here are answers to the three questions.

1. Can you perfectly separate the setosa and versicolor classes using only one feature? If so, which feature or features would work?
   1. Pedal length
   2. Pedal width
2. Can you perfectly separate the setosa and virginica classes using only one feature? If so, which feature or features would work?
   1. Petal length
   2. Petal width
3. Can you perfectly separate the versicolor and virginica classes using only one feature? If so, which feature or features would work?
   1. No, there are overlapping datapoints within each plot, aka more than two features may be necessary to separate the two classes.

A graph of a number of numbers

Description automatically generated with medium confidence

**Fig. 1** sepal length (1) vs. sepal length (1), iris dataset.

A graph of different colored dots

Description automatically generated

**Fig. 2** sepal width (2) vs. sepal length (1), iris dataset.

A graph of different colored lines

Description automatically generated

**Fig. 3** petal length (3) vs. sepal length (1), iris dataset.

A graph of different sizes and colors

Description automatically generated

**Fig. 4** petal width (4) vs. sepal length (1), iris dataset.

A graph with numbers and symbols

Description automatically generated with medium confidence

**Fig. 5** sepal width (2) vs. sepal width (2), iris dataset.

A graph of different colored dots

Description automatically generated

**Fig. 6** petal length (3) vs. sepal width (2), iris dataset.

A graph of different colored lines

Description automatically generated

**Fig. 7** petal width (4) vs. sepal width (2), iris dataset.

A graph with numbers and points

Description automatically generated with medium confidence

**Fig. 8** petal length (3) vs. petal length (3), iris dataset.

A graph with different colored lines

Description automatically generated

**Fig. 9** petal width (4) vs. petal length (3), iris dataset.

A graph with colored dots and numbers

Description automatically generated

**Fig. 10** petal width (4) vs. petal width (4), iris dataset.