**Objective-1: Data Exploration and Visualization**

* 1. For each of the three flower categories. i.e., generating a 1-D histogram. Compare the three histograms side by side and explain your observations.

Setosa versicolor virginica

**Observations**: Most of the petal width of setosa are around 0.2 to 0.25 inches and only around 15% of all distribute rest of the range. The average and the median will be a lot of different.

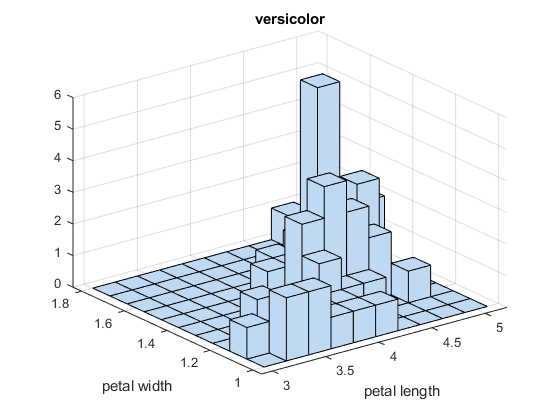
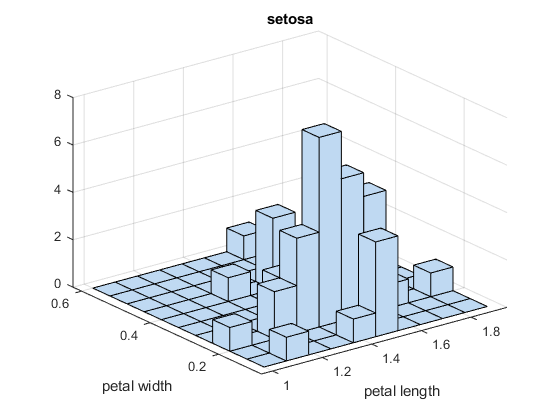
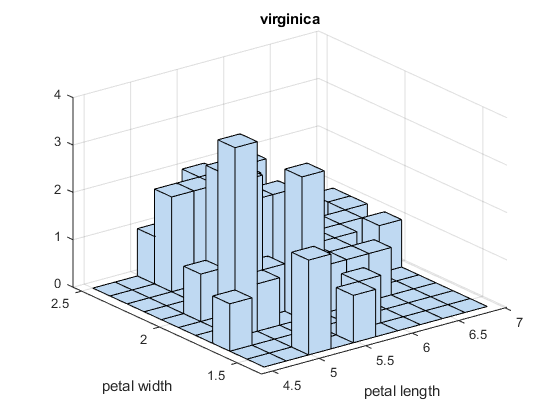
The petal width of versicolor are gather at 1.3~1.5, but there are several at different inches. The average of versicolor is similar with the median, around 1.4 inch.

The petal width of versicolor has a peak at 1.8 to 2 inches, and the average is close to about 2.2.

Versicolor and virginica widely distribute over length, but setosa centralizes at one width.

* 1. For each of the three categories, generate a 2D hist as the following and explain your observations if any:

Setosa Versicolor Virginica

 ****

**Observations**:

Setosa: Even though the petal width of Setosa gather at around 0.2 to 0.25, the distribution of petal length looks like normal distribution. Many kinds of petal length could be happened with the petal width which is around 0.2. Moreover, if a petal width is wider than 0.3 inches, the petal length will still be aroung 1.6 inchese.

Versicolor: When a petal width of versicolor increases, the petal of length increase. The wider petal, the longer petal length.

Virginica: A petal width cannot determine a petal length. Iris virginicas evenly distribute over the area.

1.3 For each category, use the boxplot() to generate the boxplotsfor all four features, as the following.

Setosa



Versicolour



Virginica



1.4 For each category in the dataset, randomly select 50% flower samples. With all the selected samples, try to generate the scatter plots:

Gpltomatrix



1.5 Visualization of similarity matrix. Represent every sample with a four-dimension vector (sepal length, sepal width, petal length, petal width).

****

**Objective-2: classification and evaluation.**

2.1

Source:

%Get one kind of iris

setosa\_all = iris( ismember(iris.Irissetosa, 'Iris-setosa' ),: );

%randomly pick 25 indices from 1 to 50.

rand\_row\_indices = randperm (size (setosa\_all, 1), size(setosa\_all, 1)/2);

%Separate the data into two subsets.

setosa\_train =setosa\_all(rand\_row\_indices,:);

setosa\_test =setosa\_all(setdiff(1:end,rand\_row\_indices),:);

2.2

MaxNumCategories : Control the number of categories. (Default is 10)

MergeLeaves : Whether merge the leaves when the risk is >= the parent node. (Default is on)

MinParentSize : It has at least number of nodes which will be observed.

CrossVal : Separate data into several parts. Use one them to be the test data and rest of them to be the train data.

MaxNumSplits : The maximum number which splitting times could accept.

MinLeafSize : Each leaf has at least the number per leaf

NumVariablesToSample: The number of samples that the function randomly selects for training.

2.3

Apply the learned tree to predict the labels of testing samples.

(i) success cases, i.e. samples for which your model can correctly predict their classes;

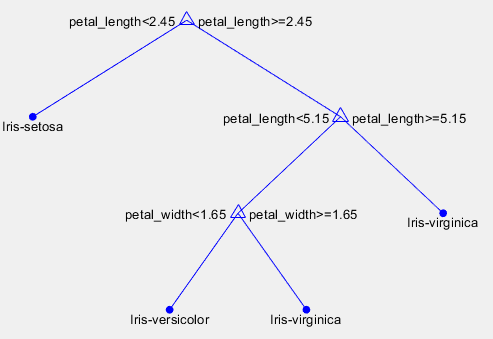
(ii) failure cases, i.e. samples for which your model fails to predict the correct labels.

(I) success cases

After using these samples trained, it has 98.67% accuracy to predict the testing samples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sepal length | Sepal width | Petal length | Petal width | Category |
| 4.6 | 3.4 | 1.4 | 0.3 | Iris-setosa |
| 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |
| 5.4 | 3.9 | 1.3 | 0.4 | Iris-setosa |
| 4.8 | 3 | 1.4 | 0.3 | Iris-setosa |
| 5.4 | 3.4 | 1.7 | 0.2 | Iris-setosa |
| 4.5 | 2.3 | 1.3 | 0.3 | Iris-setosa |
| 4.8 | 3.4 | 1.6 | 0.2 | Iris-setosa |
| 4.8 | 3.1 | 1.6 | 0.2 | Iris-setosa |
| 5.4 | 3.4 | 1.5 | 0.4 | Iris-setosa |
| 5.1 | 3.3 | 1.7 | 0.5 | Iris-setosa |
| 5.2 | 4.1 | 1.5 | 0.1 | Iris-setosa |
| 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 5.5 | 3.5 | 1.3 | 0.2 | Iris-setosa |
| 4.4 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 5.2 | 3.4 | 1.4 | 0.2 | Iris-setosa |
| 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| 5.3 | 3.7 | 1.5 | 0.2 | Iris-setosa |
| 5.4 | 3.9 | 1.7 | 0.4 | Iris-setosa |
| 4.4 | 2.9 | 1.4 | 0.2 | Iris-setosa |
| 5 | 3.6 | 1.4 | 0.2 | Iris-setosa |
| 5.1 | 3.8 | 1.9 | 0.4 | Iris-setosa |
| 5.1 | 3.8 | 1.5 | 0.3 | Iris-setosa |
| 5.7 | 4.4 | 1.5 | 0.4 | Iris-setosa |
| 4.9 | 3 | 1.4 | 0.2 | Iris-setosa |
| 5.5 | 2.5 | 4 | 1.3 | Iris-versicolor |
| 5.5 | 2.4 | 3.7 | 1 | Iris-versicolor |
| 5.4 | 3 | 4.5 | 1.5 | Iris-versicolor |
| 5 | 2.3 | 3.3 | 1 | Iris-versicolor |
| 6 | 2.7 | 5.1 | 1.6 | Iris-versicolor |
| 6.7 | 3 | 5 | 1.7 | Iris-versicolor |
| 6.2 | 2.2 | 4.5 | 1.5 | Iris-versicolor |
| 5.9 | 3.2 | 4.8 | 1.8 | Iris-versicolor |
| 6.3 | 2.5 | 4.9 | 1.5 | Iris-versicolor |
| 6.3 | 3.3 | 4.7 | 1.6 | Iris-versicolor |
| 5.7 | 3 | 4.2 | 1.2 | Iris-versicolor |
| 6.2 | 2.9 | 4.3 | 1.3 | Iris-versicolor |
| 5.6 | 3 | 4.1 | 1.3 | Iris-versicolor |
| 5.6 | 2.9 | 3.6 | 1.3 | Iris-versicolor |
| 6.1 | 3 | 4.6 | 1.4 | Iris-versicolor |
| 5.6 | 2.7 | 4.2 | 1.3 | Iris-versicolor |
| 6.4 | 3.2 | 4.5 | 1.5 | Iris-versicolor |
| 6.6 | 2.9 | 4.6 | 1.3 | Iris-versicolor |
| 5.1 | 2.5 | 3 | 1.1 | Iris-versicolor |
| 5.7 | 2.9 | 4.2 | 1.3 | Iris-versicolor |
| 5.2 | 2.7 | 3.9 | 1.4 | Iris-versicolor |
| 5.6 | 2.5 | 3.9 | 1.1 | Iris-versicolor |
| 5.7 | 2.6 | 3.5 | 1 | Iris-versicolor |
| 6.7 | 3.1 | 4.7 | 1.5 | Iris-versicolor |
| 5.6 | 3 | 4.5 | 1.5 | Iris-versicolor |
| 7.2 | 3 | 5.8 | 1.6 | Iris-virginica |
| 6 | 2.2 | 5 | 1.5 | Iris-virginica |
| 6.4 | 2.7 | 5.3 | 1.9 | Iris-virginica |
| 6.1 | 2.6 | 5.6 | 1.4 | Iris-virginica |
| 7.2 | 3.2 | 6 | 1.8 | Iris-virginica |
| 6.7 | 3.3 | 5.7 | 2.5 | Iris-virginica |
| 6.2 | 3.4 | 5.4 | 2.3 | Iris-virginica |
| 6.5 | 3 | 5.8 | 2.2 | Iris-virginica |
| 6.7 | 3.1 | 5.6 | 2.4 | Iris-virginica |
| 6.8 | 3 | 5.5 | 2.1 | Iris-virginica |
| 6.4 | 2.8 | 5.6 | 2.2 | Iris-virginica |
| 6.7 | 3.3 | 5.7 | 2.1 | Iris-virginica |
| 7.9 | 3.8 | 6.4 | 2 | Iris-virginica |
| 7.1 | 3 | 5.9 | 2.1 | Iris-virginica |
| 6.4 | 3.2 | 5.3 | 2.3 | Iris-virginica |
| 6.8 | 3.2 | 5.9 | 2.3 | Iris-virginica |
| 5.8 | 2.7 | 5.1 | 1.9 | Iris-virginica |
| 7.2 | 3.6 | 6.1 | 2.5 | Iris-virginica |
| 7.4 | 2.8 | 6.1 | 1.9 | Iris-virginica |
| 7.6 | 3 | 6.6 | 2.1 | Iris-virginica |
| 4.9 | 2.5 | 4.5 | 1.7 | Iris-virginica |
| 7.7 | 3.8 | 6.7 | 2.2 | Iris-virginica |
| 6 | 3 | 4.8 | 1.8 | Iris-virginica |
| 7.7 | 2.8 | 6.7 | 2 | Iris-virginica |
| 6.5 | 3 | 5.2 | 2 | Iris-virginica |

The decision tree which has 98.67% accuracy



(II) failure cases

After using these samples with MaxNumSplits=1 trained, it has 66.67% accuracy to predict the testing samples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sepal length | Sepal width | Petal length | Petal width | Category |
| 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |
| 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| 4.8 | 3.1 | 1.6 | 0.2 | Iris-setosa |
| 5.1 | 3.8 | 1.9 | 0.4 | Iris-setosa |
| 5.1 | 3.8 | 1.5 | 0.3 | Iris-setosa |
| 5.2 | 3.4 | 1.4 | 0.2 | Iris-setosa |
| 5 | 3.4 | 1.6 | 0.4 | Iris-setosa |
| 4.4 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 5 | 3.5 | 1.6 | 0.6 | Iris-setosa |
| 4.8 | 3 | 1.4 | 0.1 | Iris-setosa |
| 5.8 | 4 | 1.2 | 0.2 | Iris-setosa |
| 5.4 | 3.9 | 1.7 | 0.4 | Iris-setosa |
| 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| 4.3 | 3 | 1.1 | 0.1 | Iris-setosa |
| 5.1 | 3.5 | 1.4 | 0.3 | Iris-setosa |
| 4.8 | 3 | 1.4 | 0.3 | Iris-setosa |
| 4.9 | 3 | 1.4 | 0.2 | Iris-setosa |
| 5.5 | 4.2 | 1.4 | 0.2 | Iris-setosa |
| 5.3 | 3.7 | 1.5 | 0.2 | Iris-setosa |
| 4.4 | 2.9 | 1.4 | 0.2 | Iris-setosa |
| 4.7 | 3.2 | 1.6 | 0.2 | Iris-setosa |
| 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa |
| 4.6 | 3.4 | 1.4 | 0.3 | Iris-setosa |
| 4.5 | 2.3 | 1.3 | 0.3 | Iris-setosa |
| 5.1 | 3.3 | 1.7 | 0.5 | Iris-setosa |
| 4.9 | 2.4 | 3.3 | 1 | Iris-versicolor |
| 5.6 | 3 | 4.1 | 1.3 | Iris-versicolor |
| 6.1 | 2.8 | 4.7 | 1.2 | Iris-versicolor |
| 5.8 | 2.7 | 3.9 | 1.2 | Iris-versicolor |
| 5.8 | 2.7 | 4.1 | 1 | Iris-versicolor |
| 6.8 | 2.8 | 4.8 | 1.4 | Iris-versicolor |
| 5.5 | 2.4 | 3.7 | 1 | Iris-versicolor |
| 5 | 2 | 3.5 | 1 | Iris-versicolor |
| 6 | 3.4 | 4.5 | 1.6 | Iris-versicolor |
| 5.6 | 3 | 4.5 | 1.5 | Iris-versicolor |
| 5.6 | 2.9 | 3.6 | 1.3 | Iris-versicolor |
| 5.7 | 2.6 | 3.5 | 1 | Iris-versicolor |
| 5.6 | 2.7 | 4.2 | 1.3 | Iris-versicolor |
| 6 | 2.2 | 4 | 1 | Iris-versicolor |
| 5.7 | 3 | 4.2 | 1.2 | Iris-versicolor |
| 5.9 | 3.2 | 4.8 | 1.8 | Iris-versicolor |
| 5.5 | 2.5 | 4 | 1.3 | Iris-versicolor |
| 6.4 | 2.9 | 4.3 | 1.3 | Iris-versicolor |
| 5.7 | 2.8 | 4.5 | 1.3 | Iris-versicolor |
| 6.3 | 2.3 | 4.4 | 1.3 | Iris-versicolor |
| 6.1 | 2.8 | 4 | 1.3 | Iris-versicolor |
| 5.1 | 2.5 | 3 | 1.1 | Iris-versicolor |
| 6.5 | 2.8 | 4.6 | 1.5 | Iris-versicolor |
| 5.8 | 2.6 | 4 | 1.2 | Iris-versicolor |
| 6.3 | 3.3 | 4.7 | 1.6 | Iris-versicolor |
| 7.4 | 2.8 | 6.1 | 1.9 | Iris-virginica |
| 5.8 | 2.7 | 5.1 | 1.9 | Iris-virginica |
| 6.4 | 2.8 | 5.6 | 2.1 | Iris-virginica |
| 6.4 | 3.2 | 5.3 | 2.3 | Iris-virginica |
| 6.8 | 3.2 | 5.9 | 2.3 | Iris-virginica |
| 6.3 | 2.8 | 5.1 | 1.5 | Iris-virginica |
| 6.5 | 3 | 5.8 | 2.2 | Iris-virginica |
| 6.7 | 3.1 | 5.6 | 2.4 | Iris-virginica |
| 5.7 | 2.5 | 5 | 2 | Iris-virginica |
| 5.9 | 3 | 5.1 | 1.8 | Iris-virginica |
| 7.6 | 3 | 6.6 | 2.1 | Iris-virginica |
| 7.2 | 3.2 | 6 | 1.8 | Iris-virginica |
| 6 | 3 | 4.8 | 1.8 | Iris-virginica |
| 7.3 | 2.9 | 6.3 | 1.8 | Iris-virginica |
| 6 | 2.2 | 5 | 1.5 | Iris-virginica |
| 7.7 | 2.6 | 6.9 | 2.3 | Iris-virginica |
| 7.9 | 3.8 | 6.4 | 2 | Iris-virginica |
| 6.9 | 3.1 | 5.4 | 2.1 | Iris-virginica |
| 6.7 | 3 | 5.2 | 2.3 | Iris-virginica |
| 6.3 | 3.4 | 5.6 | 2.4 | Iris-virginica |
| 6.7 | 2.5 | 5.8 | 1.8 | Iris-virginica |
| 6.9 | 3.1 | 5.1 | 2.3 | Iris-virginica |
| 6.3 | 2.9 | 5.6 | 1.8 | Iris-virginica |
| 6.7 | 3.3 | 5.7 | 2.5 | Iris-virginica |
| 7.7 | 2.8 | 6.7 | 2 | Iris-virginica |

The decision tree which has only 66.67% accuracy

