Pattern Recognition

Analysis of the Iris Dataset

Name: Aly Mohamed Aly Nasr

ID: 6835

Iris Dataset

The Iris Dataset classifies 150 samples of flowers into 3 different types:

- Setosa

- Versicolour

- Virginica

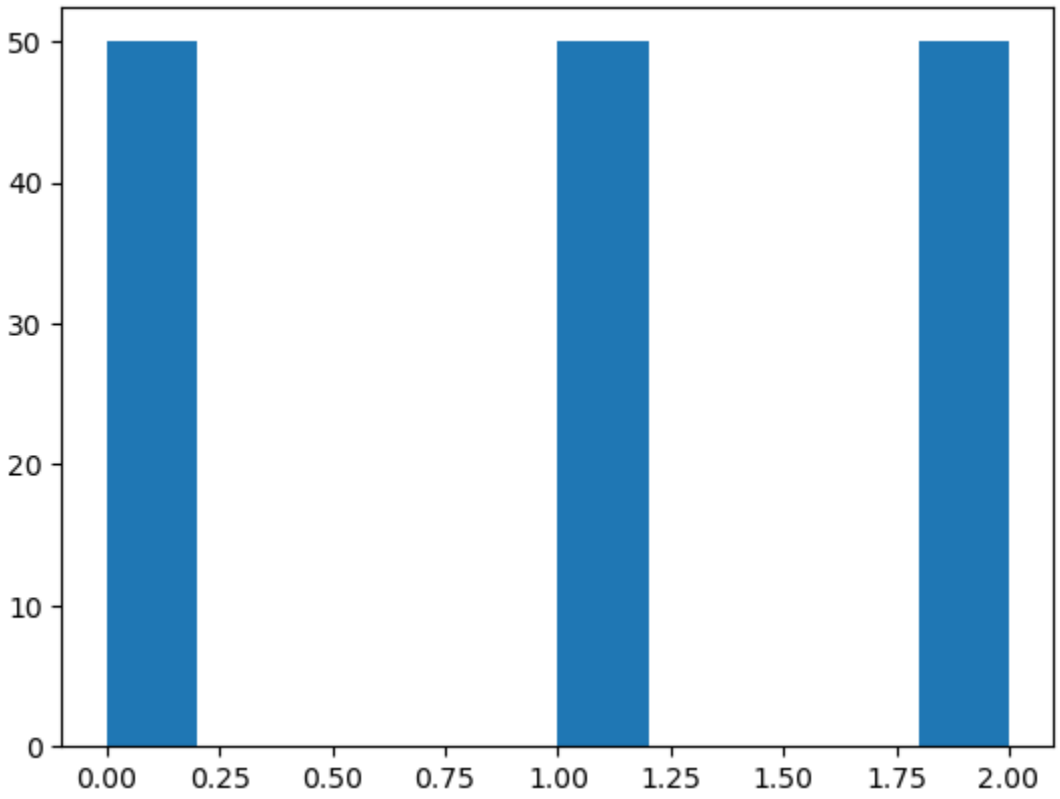
It has 4 attributes which are:

- sepal length
- sepal width
- petal length
- petal width

Which are all recorded in cm.

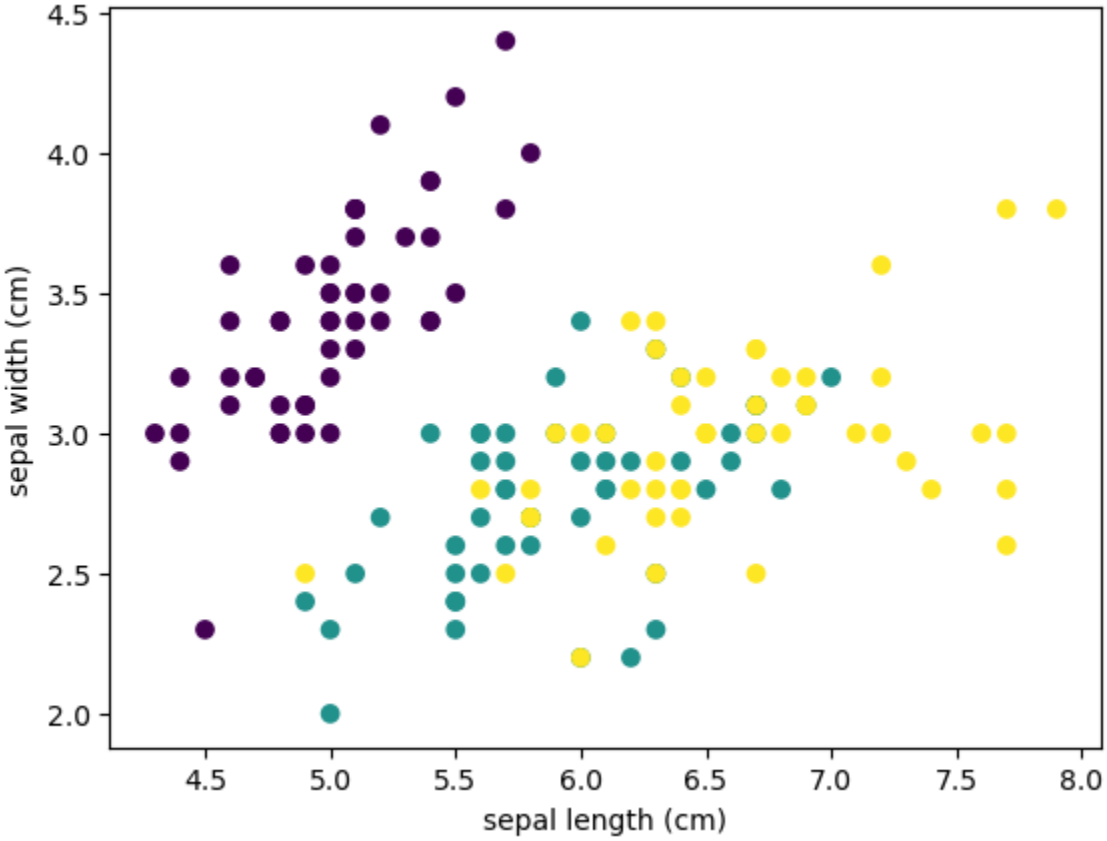
Visualization

It can be seen that each class has exactly 50 samples.

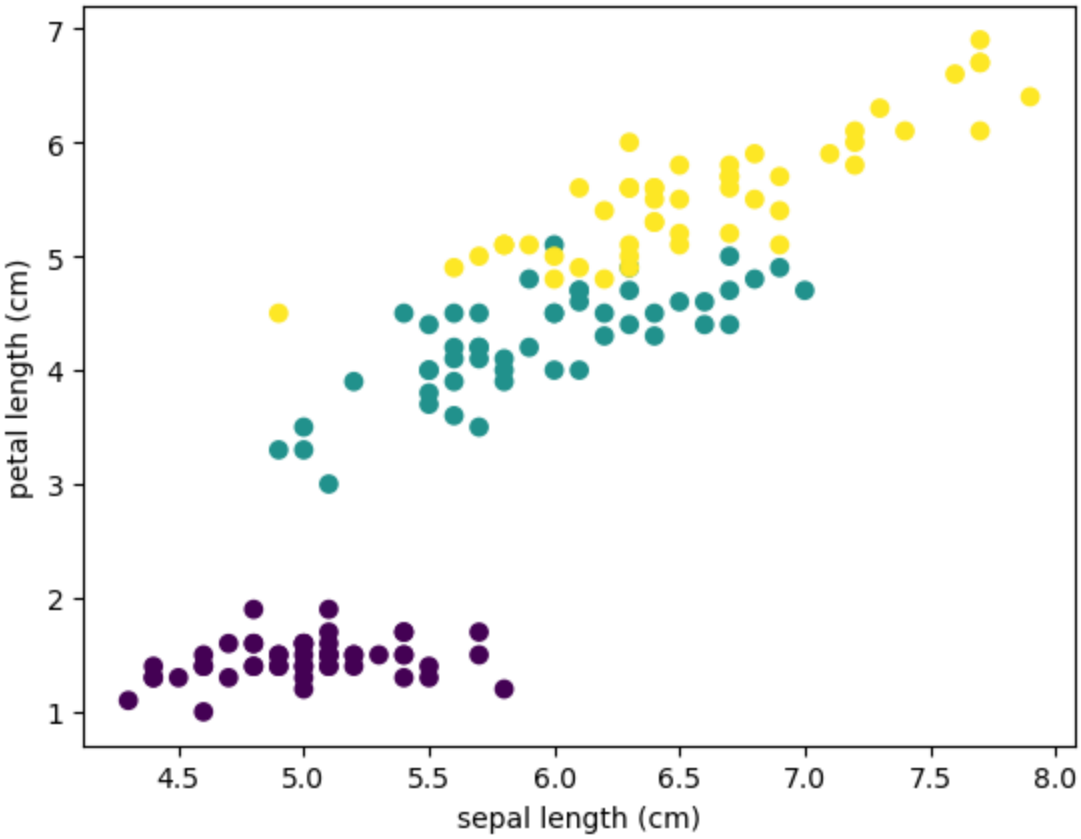


Scatter plots of the different attributes

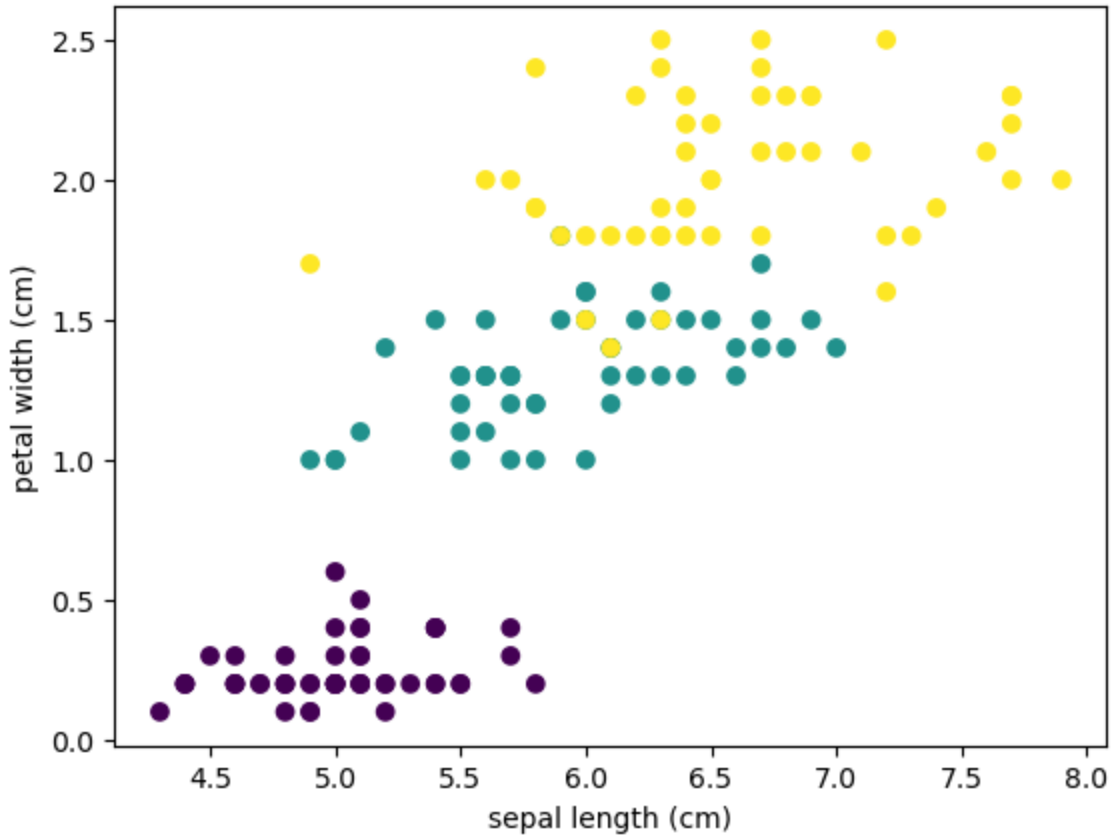
2D:



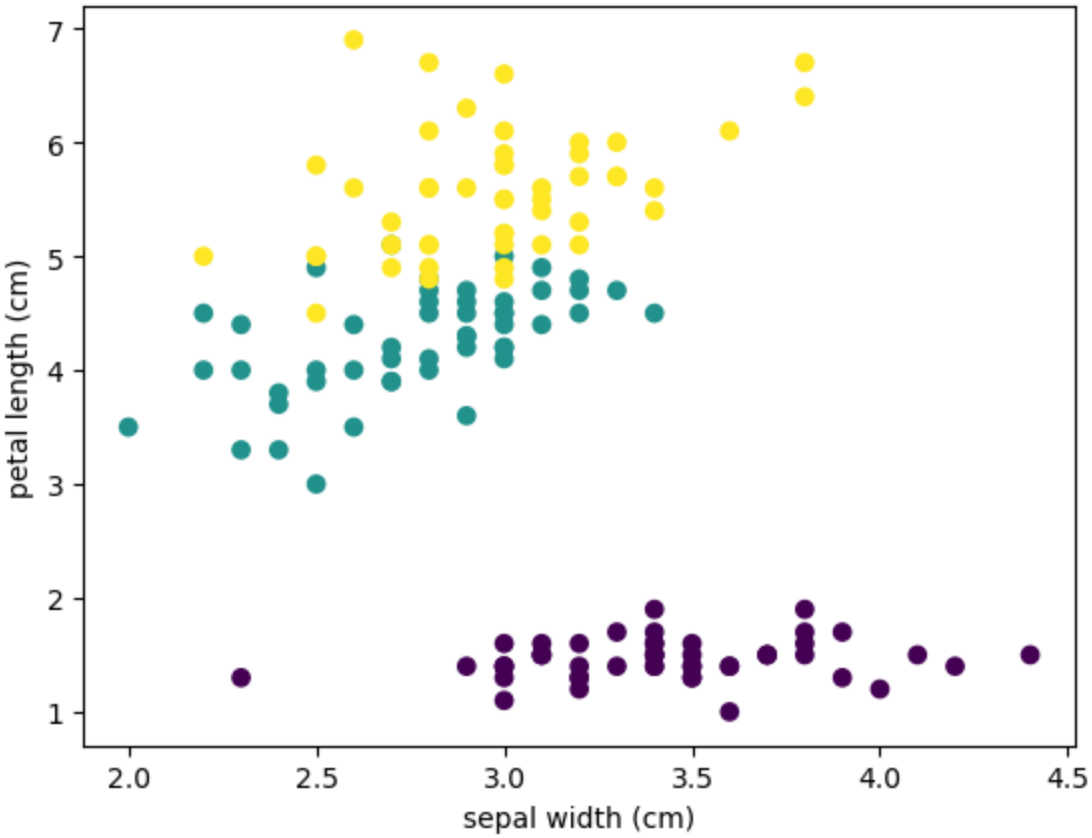
There appears to be little correlation between these 2 attributes.



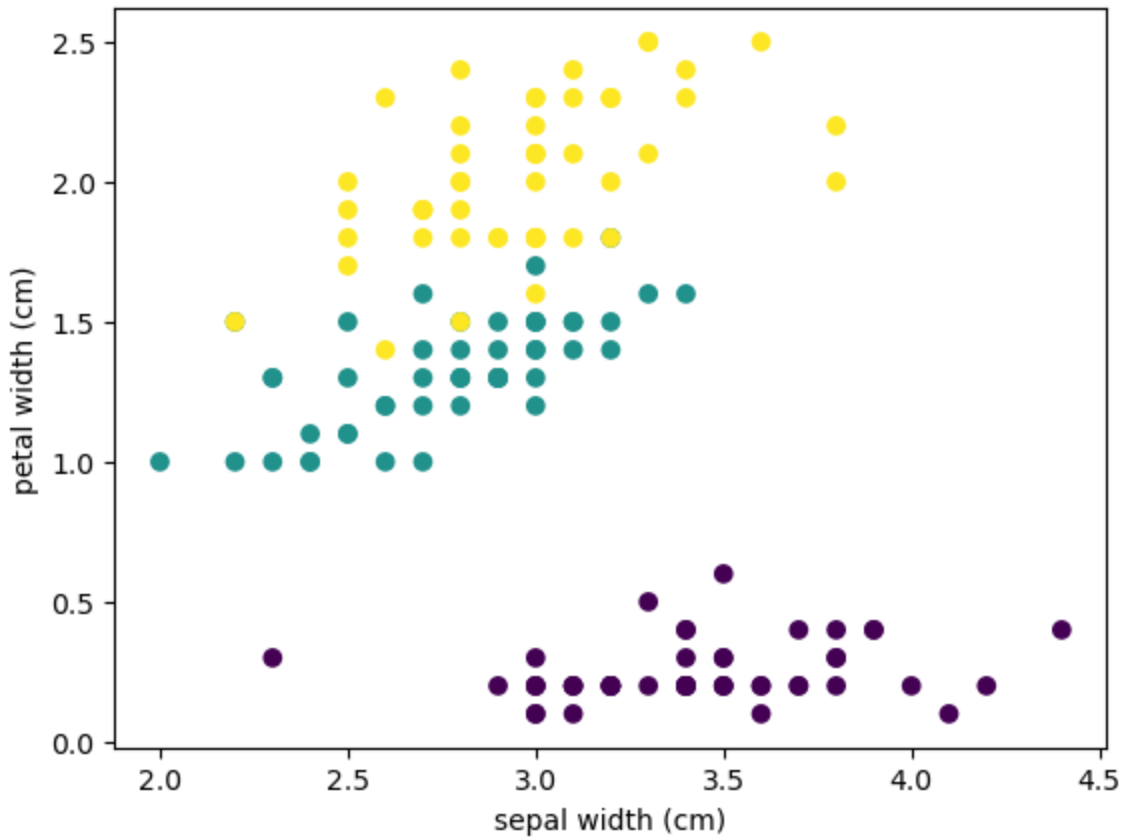
These 2 attributes appear to have a positive correlation.



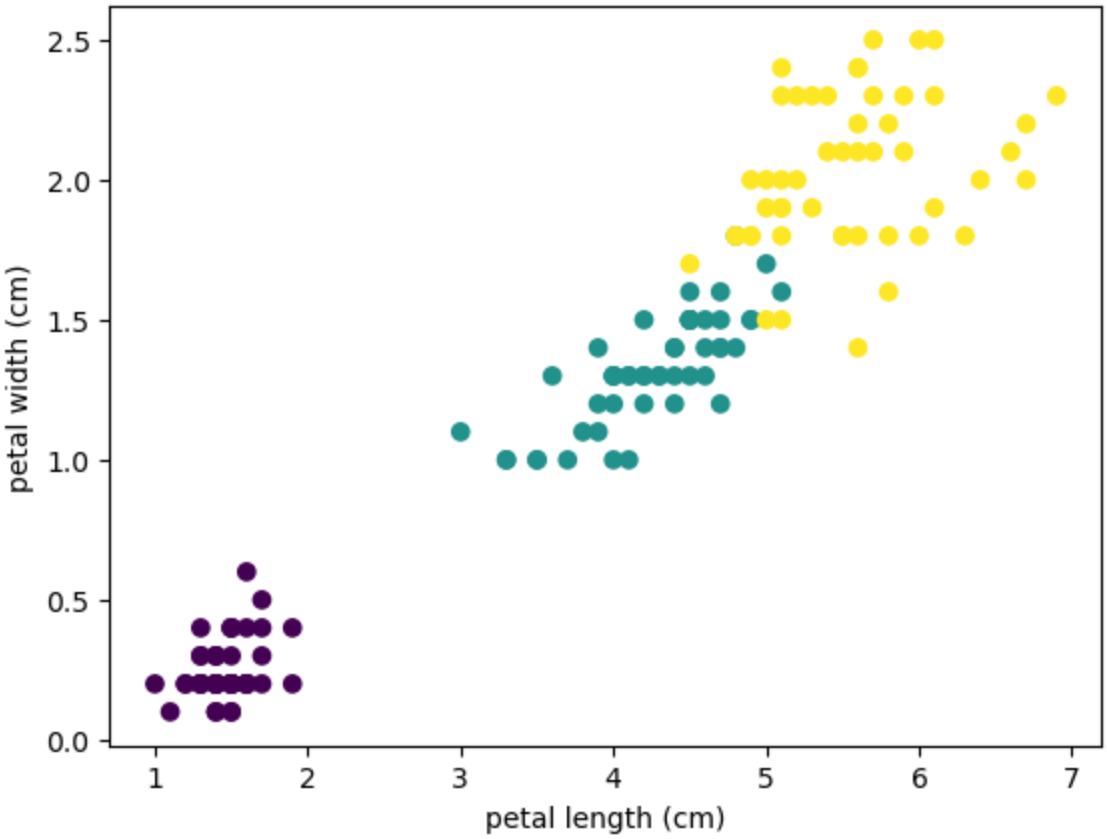
There appears to be little correlation between these 2 attributes.



There appears to be little correlation between these 2 attributes.



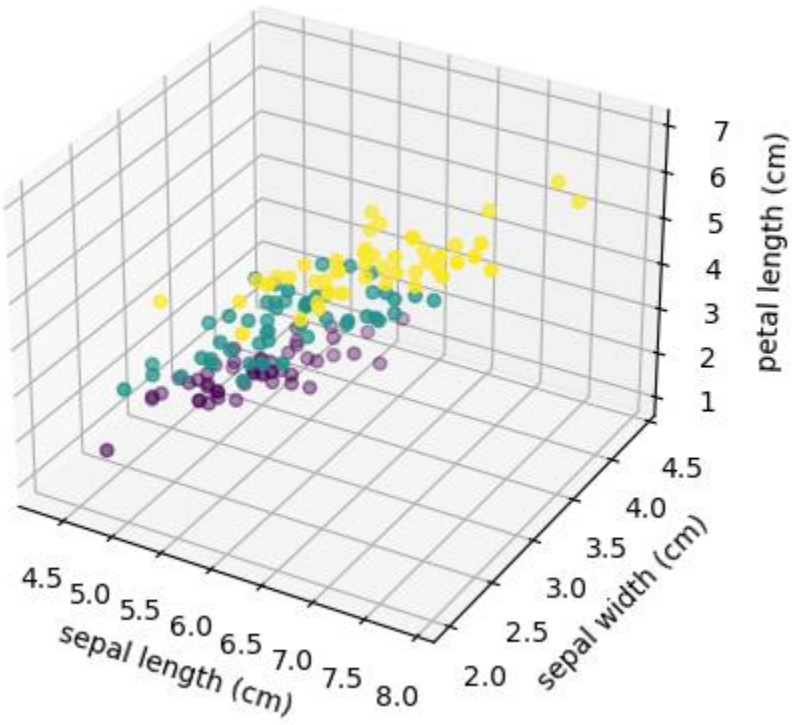
There appears to be little correlation between these 2 attributes.

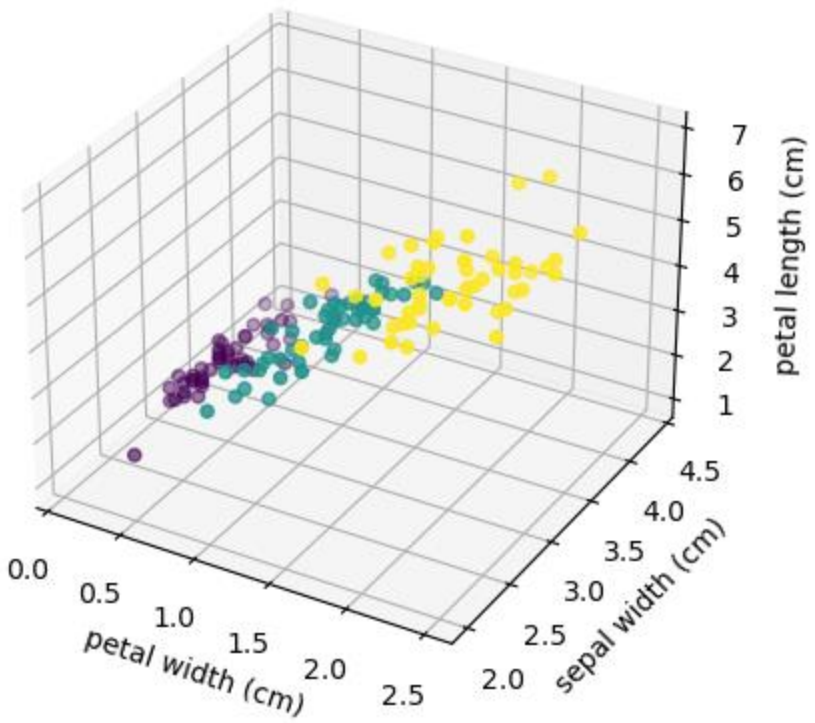


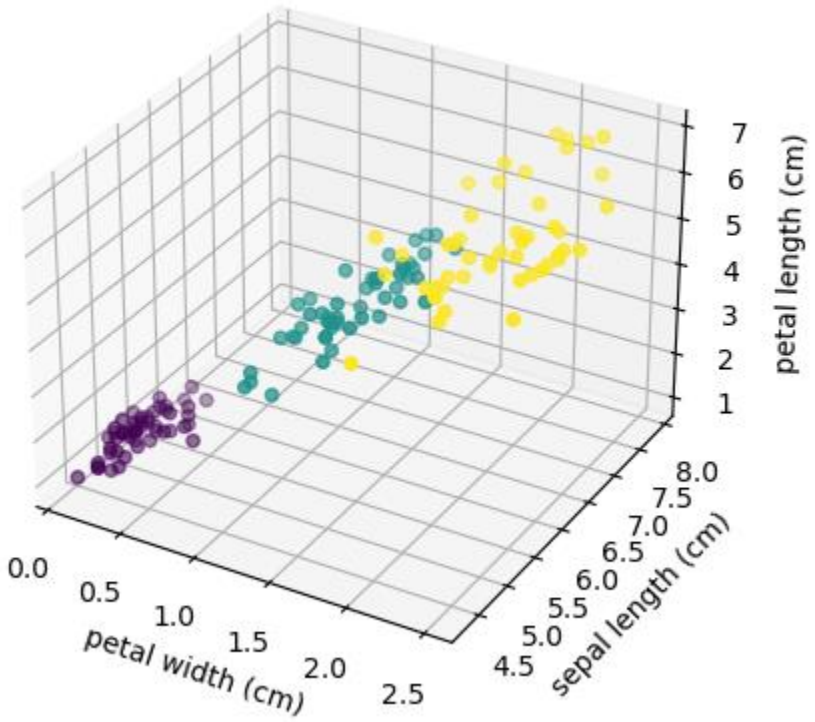
These 2 attributes appear to have a positive correlation.

Scatter plots of the different attributes

3D:




petal length (cm)
1
2
3
4
5
6
7
sepal width (cm)
2.0
2.5
3.0
3.5
4.0
4.5
petal width (cm)
0.0
0.5
1.0
1.5
2.0
2.5

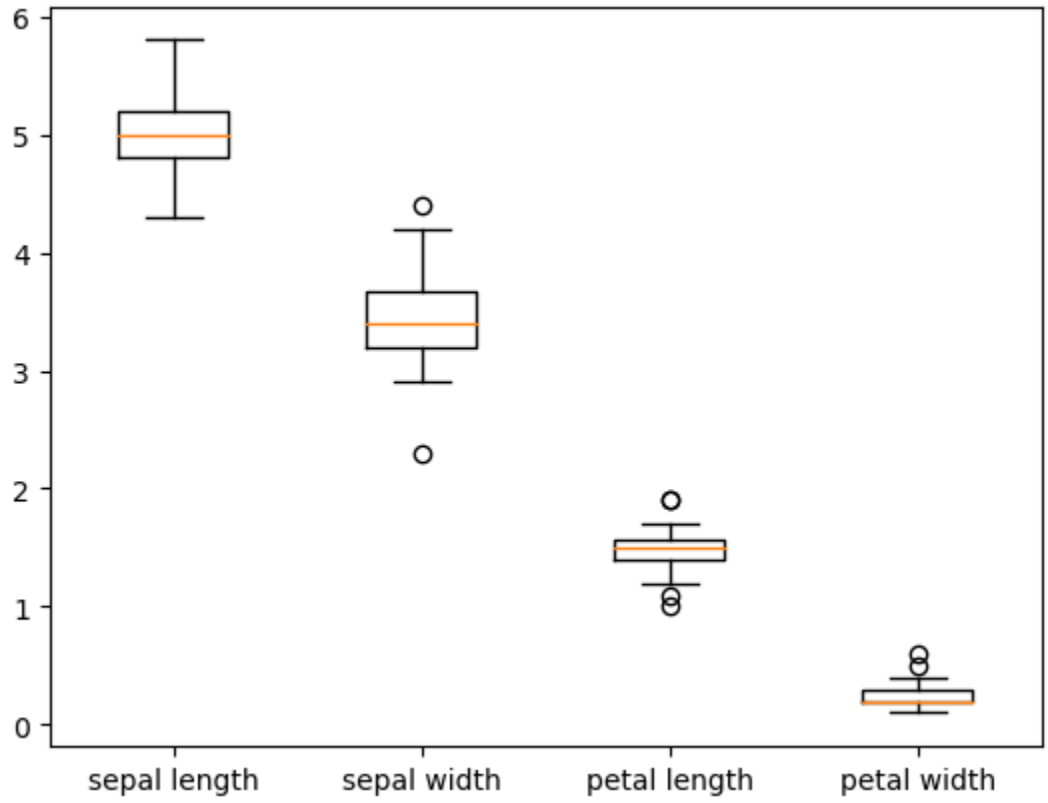

petal length (cm)
1
2
3
4
5
6
7
sepal length (cm)
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
petal width (cm)
0.0
0.5
1.0
1.5
2.0
2.5

Correlation matrix:

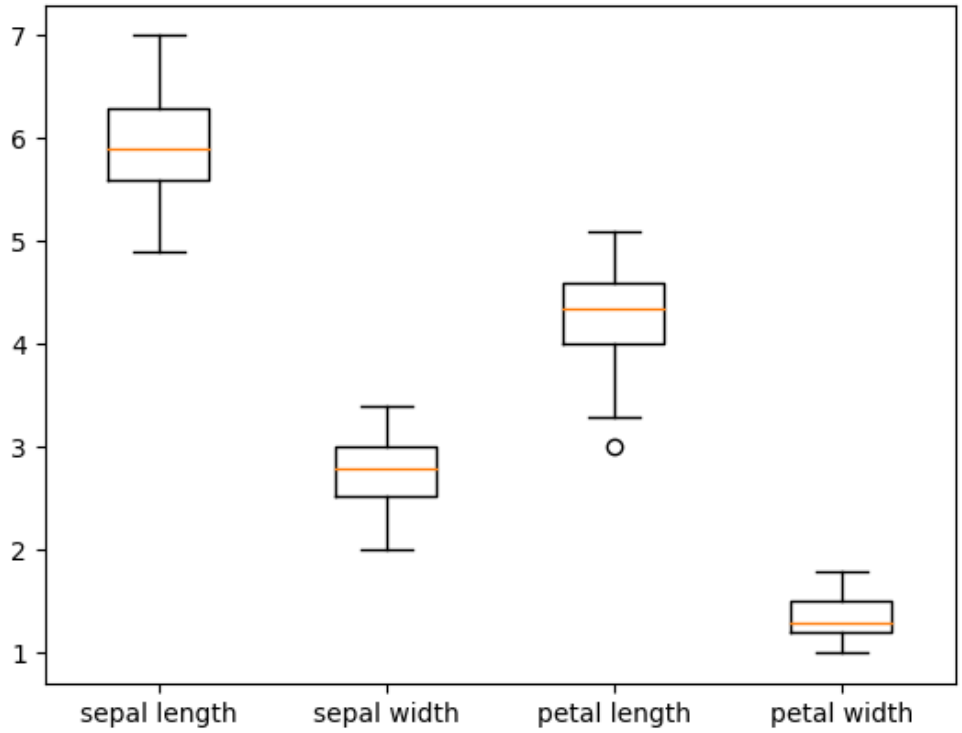
	sepal length	sepal width	petal length	petal width
sepal length	1	-0.11756978	0.87175378	0.81794113
sepal width	-0.11756978	1	-0.4284401	-0.36612593
petal length	0.87175378	-0.4284401	1	0.96286543
petal width	0.81794113	-0.36612593	0.96286543	1

Distribution of different attributes according to class:

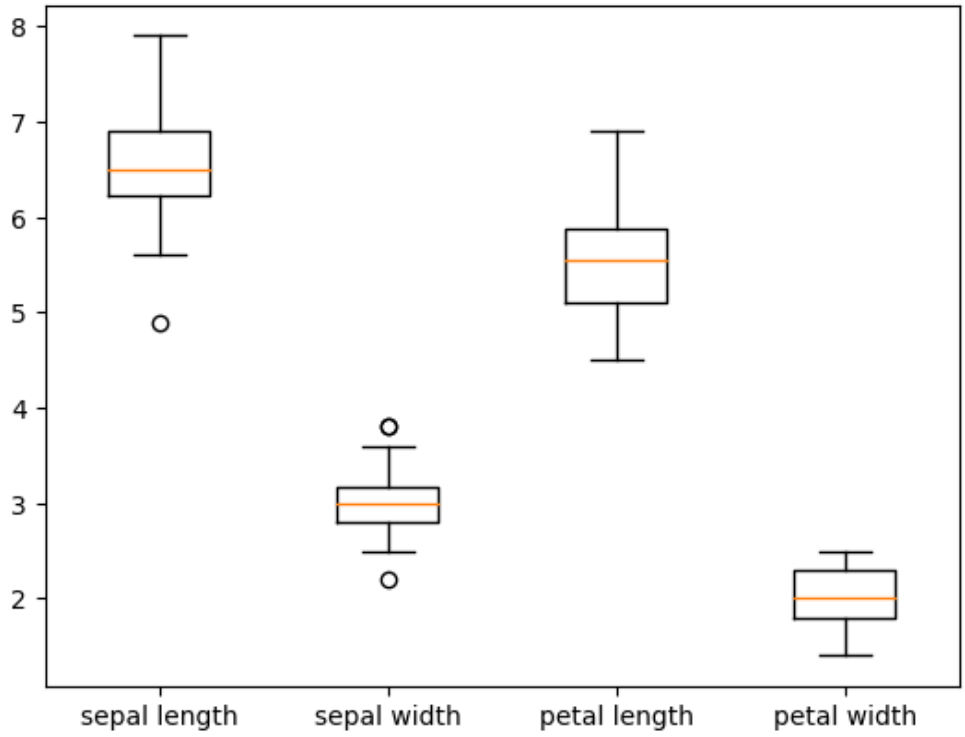
setosa



Versicolor



Virginica



Observations

1. In the different graphs we can observe that each class is clustered together
2. The two attributes petal length and petal width have a very high degree of positive correlation (0.96286543)

Conclusion

Both support vector machines and Trees appear to be promising algorithms in terms of machine learning.