ML_HW#2

106033212 曾靖桐

1.

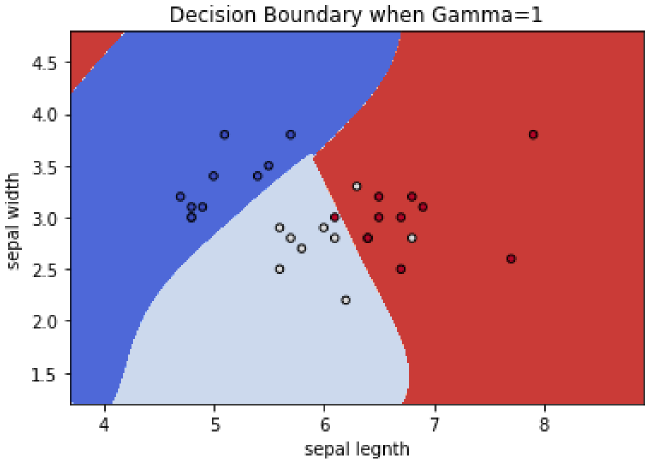
Depth	Accuracy	Confusion Matrix
2	0.967	10 0 0 0 8 1 0 0 11
3	1.0	10 0 0 0 9 0 0 0 11
4	1.0	10 0 0 0 9 0 0 0 11

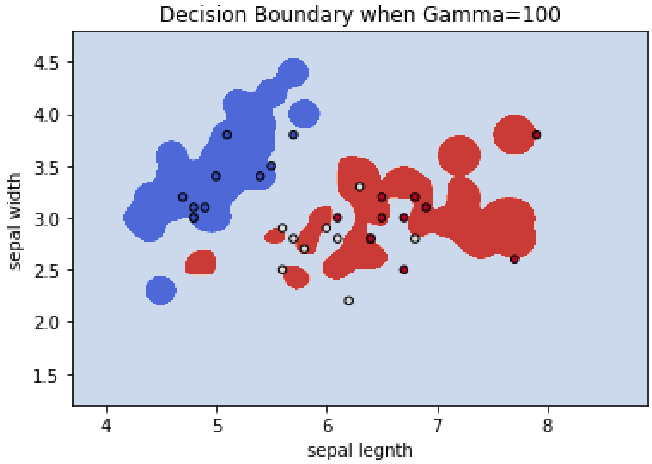
As the depth increases, the accuracy increases. By observing the confusion matrix, there's a wrong prediction when depth = 2. However, when depth=3 and depth=4, there's not any wrong prediction. Depth=3 is a better case. It has the same accuracy with depth=4, and it needs less depth.

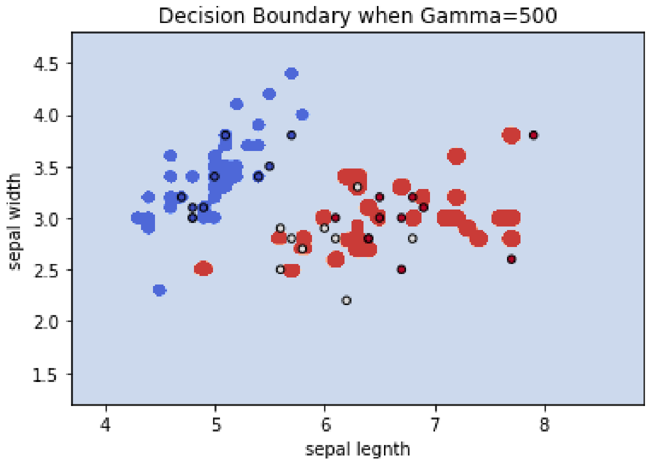
2.

Gamma (with C=1)	Accuracy	Confusion Matrix
1	1.0	10 0 0 0 9 0 0 0 11
100	0.467	4 6 0 0 9 0 0 10 1
500	0.333	1 9 0 0 9 0 0 11 0

3.





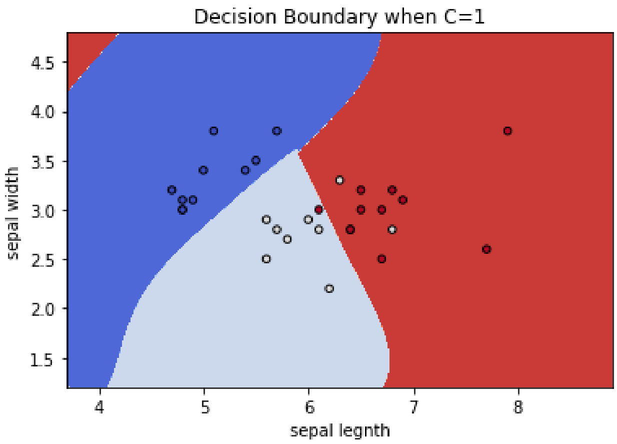


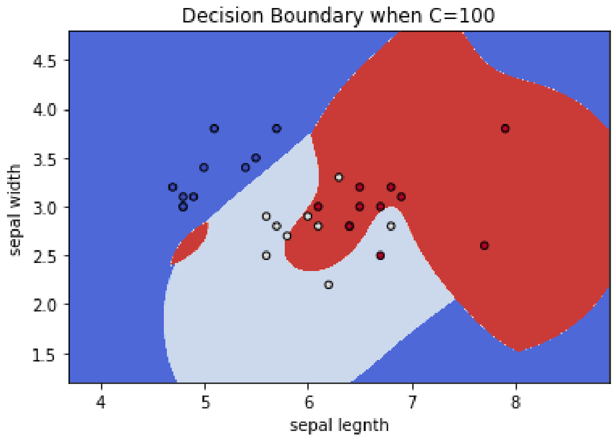
As Gamma increases, the accuracy drops dramatically. Gamma=1 is a better case. By observing the decision boundary and the confusion matrix, most of the data points are wrongly predicted to class 2(light blue region), and the model is overfitting.

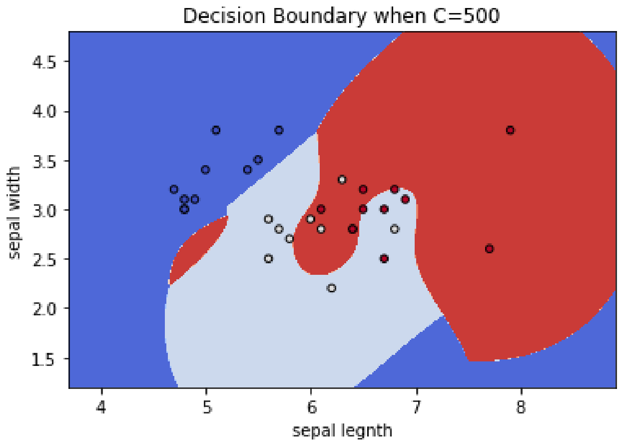
4.

C (with Gamma=1)	Accuracy	Confusion Matrix
1	1.0	10 0 0 0 9 0 0 0 11
100	0.967	10 0 0 0 9 0 0 1 10
500	0.933	10 0 0 0 8 1 0 1 10

5.







As C increases, the accuracy decreases. C=1 is a better case. Since the shape of the red region deforms dramatically, the model seems to be overfitting. And the accuracy shows that the accuracy does decrease though it has a complex shape of the boundary.

6.

In this experiment, both algorithms can get the 100% accuracy. However, I would prefer to use decision tree algorithm in this dataset because it is easier to get high accuracy by just increase the value of depth. Moreover, decision tree algorithm is better for categorical data, and it deals collinearity better than SVM. But the SVM can provide a visual diagram about the boundary which is easier to determine whether it's overfitting or not.