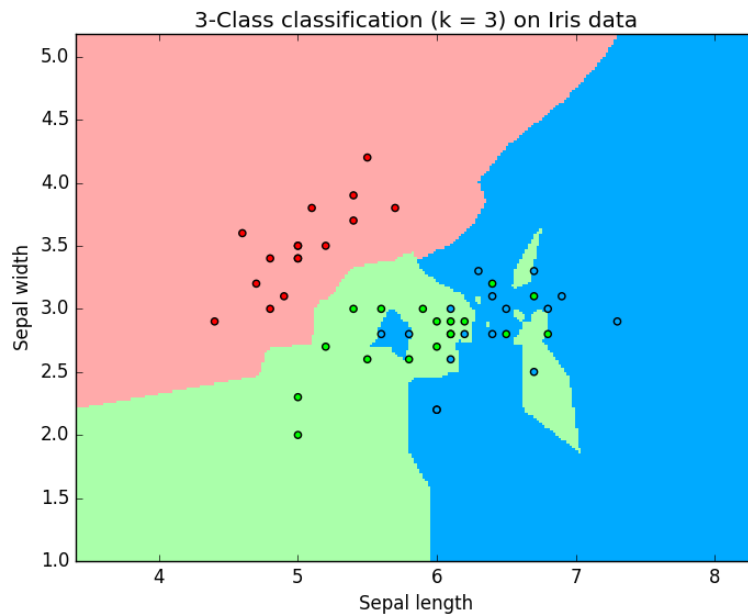
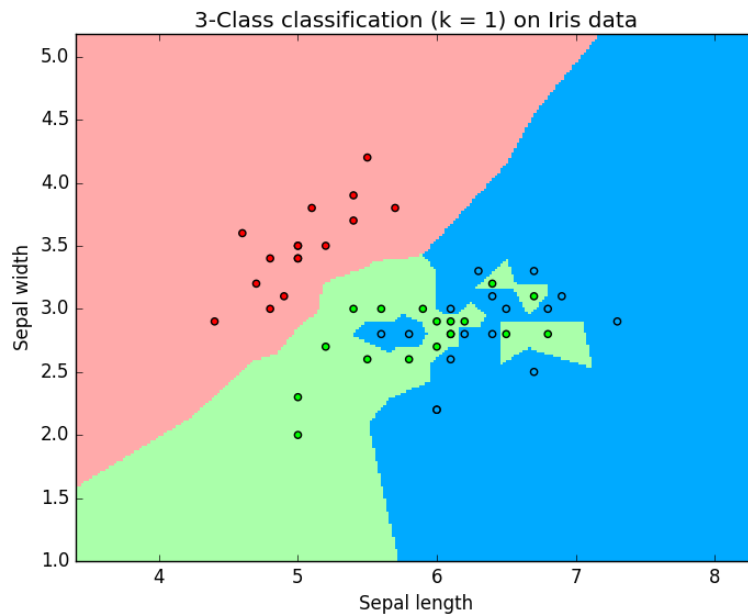
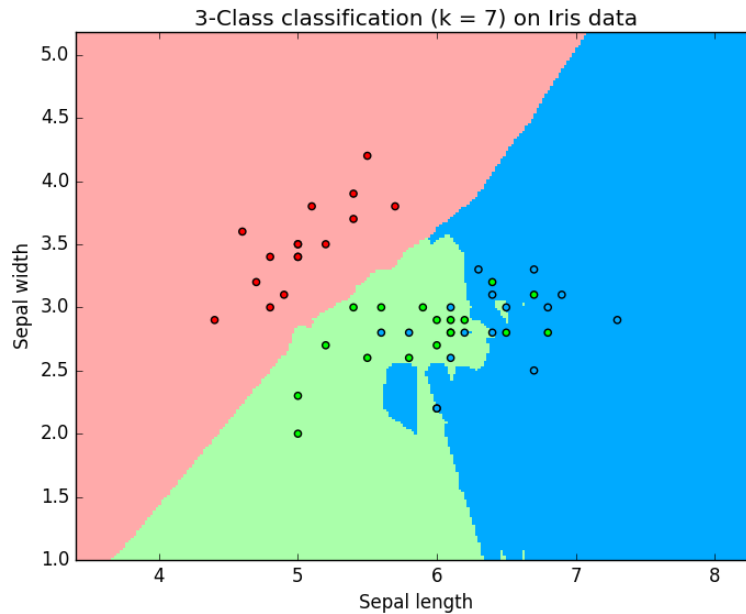


1 Question 1

The decision boundary changes from linear segments to smooth and complex curves as number of k increases.





2 Question 2

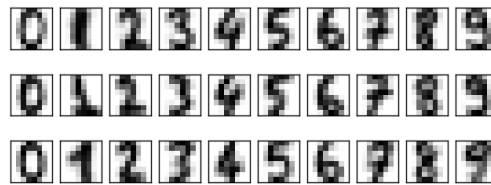


Figure 1: K means for k=3

3 Question 3

For decision tree, for fishiris.csv, both of the gini index and entropy index gives the same result on testing data. This is because fishiris is a smaller dataset and has fewer features. However, for digits.csv dataset, which is larger and more complicated, entropy performs better than gini index.

	fishiris.csv	digits.csv
gini index	0.9607	0.8180
entropy	0.9607	0.8491

For fishiris.csv dataset, we can see the accuracy for both K means and KNN peaks at $k = 5$. Then further increasing k will not help with increasing the accuracy

For digits.csv dataset, the accuracy decreases and k increases for KNN method but the accuracy increases and k increases for Kmeans method. This is because why data within the same set is sparsely distributed, using KNN with large k will incorporate more data from other class. However, using K means with large K will only increase the classification accuracy.

