COMP307 ASSIGNMENT 1
Karu Skipper : 300417869

1. Report the class labels of each instance in the test set predicted by the basic nearest neighbour method (where k=1), and the classification accuracy on the test set of the basic nearest neighbour method.

```
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-virginica | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
```

```
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
```

Accuracy: 92.0%

2. Report the classification accuracy on the test set of the k-nearest neighbour method where k=3, and compare and comment on the performance of the two classifiers (k=1 and k=3)

```
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-setosa | Actual: Iris-setosa
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
```

```
Predicted: Iris-virginica | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-versicolor | Actual: Iris-versicolor
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-versicolor | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
Predicted: Iris-virginica | Actual: Iris-virginica
```

Accuracy: 93.33%

Here we see a 2.33% increase from k = 1 to k = 3. This is because we include more neighbours nearby which are likely to include the same characteristics as the predicted flower. However, if we increase the number of neighbours past a certain threshold e.g k = 9 for this example. We will see a slight accuraccy decrease.

3. Discuss the main advantages and disadvantages of k-Nearest Neighbour method.

Pros :
- Easy algorithm to implement
- Very accurate with large data sets

Cons :
- Can be CPU instensive on large data sets
- Greedy algorithm
- Need to determine best k value

4. Assuming that you are asked to apply the k-fold cross validation method for the above problem with k=5, what would you do? State the major steps.

1. Create 5 subsets of the current set. (Since k=5)
2. Make 1 of the sub-sets the test set.
3. Have the other 4 subsets as training sets.
4. Train the classifier with the training set.
5. Apply to the test set.
6. Repeat for each of the other 4 subsets.
7. Average the results.

5. In the above problem, assuming that the class labels are not available in the training set and the test set, and that there are three clusters, which method would you use to group the examples in the data set? State the major steps.

K-means Clustering – Steps taken from lecture slides

1. Set k initial “means” randomly from the data set.
2. Create k clusters by associating every instance with the nearest mean based on a distance measure.
3. Replace the old means with the centroid of each of the k clusters (as the new means).
4. Repeat the above two steps until convergence (no change in each cluster center).

6. (Optional, bonus 5 marks) Implement the clustering method above.