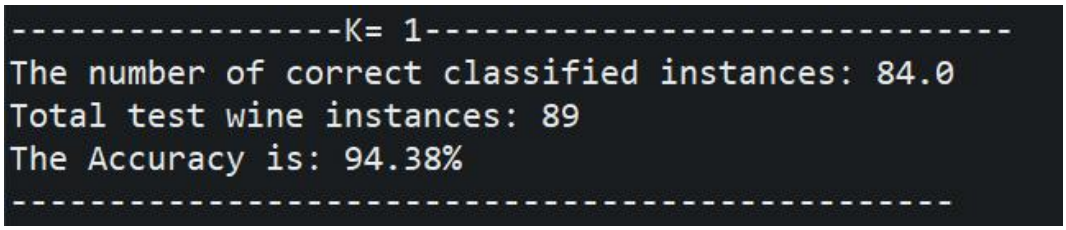
Name: Yun Zhou
ID: 300442776

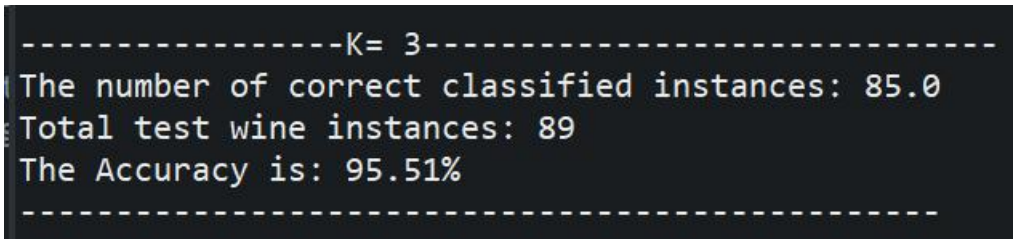
Part 1: KNN

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. Make sure you keep the same order as in the test set file.



The above screenshot is from my application part 1.
As we can see that in the case of the basic nearest neighbour which is when K = 1, there are 84 correct classified instances out of 89 test wine instances, the accuracy is 94.38%.
(p.s. The full log that reports the guess class label of each test Wine instance, whether each test Wine has been labelled correctly or not is at the end of this report. Also, you can see it in **sampleoutput_part1.txt**)

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).



As we can see, when the number of K = 3, the accuracy of my KNN is 95.51%(85/89), the correct classified instances are 85 out of 89.
Compare to the performance from basic nearest neighbour(K =1), we can see that: the correct classified instance is increased from 84 to 85, so that the accuracy is creasing from 94.38% to 95.51% as well.
Although the correct classified number only increase by 1, the accuracy is still increasing not decreasing, we have reason to believe that my KNN can also run pretty well when the k is increasing,Wine test set and the accuracy is around 95%.

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

The KNN is the classification method which depends on the similarity/distance of the features, when do the regression which is classify which class is this test instance belongs to, just sort the train list then find K's number of nearest neighbors, pretty easy to use.

Therefore, the main **advantage** is that: it is pretty *easy to use as well as it is pretty easy to be constructed*, the logic and the principle can be easy to understand. Also, the accuracy of KNN is also pretty high, part 1 result proves that *the KNN achieve good results in many cases,* these advantages means the KNN method can be rely on.

The main **disadvantage** is pretty hurt as well. First, due to the regression need to sort the whole training list, so *the effiency will be very low* if the train list is huge, it can run a pretty long time until it reaches the end of the world. Except KNN is not effiency, also, KNN depends on the particular feature. In our part 1, we divide each feature by their range, I think it is assigning the weight to each feature, if each feature is not be assigned the weight preperly, the accuracy result may be very low.

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. Merge the training Wine set and testing Wine set into one big set, let's call it the whole set.
2. Chop the whole set into 5 equals subsets(k=5), which means the size of 5 subsets are equal and if add them together, it is the Whole set.
3. For each subsets,
    (1) Treat it as the test set
    (2) Treat the rest 4 subsets (i.e.k-1) as the training set
    (3) Train classifier using the training set, apply it to the test set

The step 3 will be repeated 5 times which is the number of the K(the folds), which means the training/test process will be repeated 5 times with each of the 5 subsets used exactly once as the test set. For each fold, the test results can be added together in order to get the average single estimation.

In this case, this k-fold cross validation can be applied by the above steps, and can be used to measure the performance of my KNN.

5. In the above problem, assuming that there were actually no class labels available. Which method would you use to group the examples in the data set? State the major steps.

No class label provided means it is the Unsupervised learning, I will use the **K-means clustering method**, since it is used for unlabelled data and can partition m instances into k clusters.

To be able to implement K-means, I will:

1. Initialize k initial "means" randomly from the data set.

2. Create k clusters by assigning every instance to the nearest cluster: based on the nearest mean according to the distance measure.

3. Update the centroid, which will replace the old means with the centroid (mean) of each cluster

Repeat the above two steps until convergence (no change in each cluster centroid).(p.s. Centroid is not an instance)

Full Log, for Question 1:

```
============      Part 1      ===========
-----------------K= 1------------------------------
-------------------------------------------------
For the 1th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true
-------------------------------------------------
For the 2th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true
-------------------------------------------------
For the 3th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true
-------------------------------------------------
For the 4th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true
```

For the 5th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 6th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 7th test wine:
The real class label is 2
The guess class label by my knn is 1
which is: false

For the 8th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 9th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 10th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 11th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 12th test wine:
The real class label is 2
The guess class label by my knn is 3
which is: false

For the 13th test wine:
The real class label is 3
The guess class label by my knn is 3

which is: true

For the 14th test wine:

The real class label is 3

The guess class label by my knn is 3

which is: true

For the 15th test wine:

The real class label is 1

The guess class label by my knn is 1

which is: true

For the 16th test wine:

The real class label is 2

The guess class label by my knn is 2

which is: true

For the 17th test wine:

The real class label is 3

The guess class label by my knn is 3

which is: true

For the 18th test wine:

The real class label is 3

The guess class label by my knn is 3

which is: true

For the 19th test wine:

The real class label is 1

The guess class label by my knn is 1

which is: true

For the 20th test wine:

The real class label is 1

The guess class label by my knn is 1

which is: true

For the 21th test wine:

The real class label is 3

The guess class label by my knn is 3

which is: true

For the 22th test wine:

The real class label is 2

The guess class label by my knn is 2
which is: true

For the 23th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 24th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 25th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 26th test wine:
The real class label is 2
The guess class label by my knn is 3
which is: false

For the 27th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 28th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 29th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 30th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 31th test wine:

The real class label is 2
The guess class label by my knn is 2
which is: true

For the 32th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 33th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 34th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 35th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 36th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 37th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 38th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 39th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 40th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 41th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 42th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 43th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 44th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 45th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 46th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 47th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 48th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 49th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 50th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 51th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 52th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 53th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 54th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 55th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 56th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 57th test wine:
The real class label is 1
The guess class label by my knn is 1

which is: true

For the 58th test wine:

The real class label is 3

The guess class label by my knn is 3

which is: true

For the 59th test wine:

The real class label is 1

The guess class label by my knn is 1

which is: true

For the 60th test wine:

The real class label is 3

The guess class label by my knn is 3

which is: true

For the 61th test wine:

The real class label is 3

The guess class label by my knn is 3

which is: true

For the 62th test wine:

The real class label is 2

The guess class label by my knn is 1

which is: false

For the 63th test wine:

The real class label is 2

The guess class label by my knn is 2

which is: true

For the 64th test wine:

The real class label is 3

The guess class label by my knn is 3

which is: true

For the 65th test wine:

The real class label is 2

The guess class label by my knn is 2

which is: true

For the 66th test wine:

The real class label is 3

The guess class label by my knn is 3
which is: true

For the 67th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 68th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 69th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 70th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 71th test wine:
The real class label is 2
The guess class label by my knn is 1
which is: false

For the 72th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 73th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 74th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 75th test wine:

The real class label is 1
The guess class label by my knn is 1
which is: true

For the 76th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 77th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 78th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 79th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 80th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 81th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 82th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 83th test wine:
The real class label is 3
The guess class label by my knn is 3
which is: true

For the 84th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 85th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 86th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 87th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

For the 88th test wine:
The real class label is 2
The guess class label by my knn is 2
which is: true

For the 89th test wine:
The real class label is 1
The guess class label by my knn is 1
which is: true

----------K= 1----------
The number of correct classified instances: 84.0
Total test wine instances: 89
The Accuracy is: 94.38%
