Assignment 2

Student Name: Hao Bai

Student ID: 180223545

Q1-3-3:

When K = 4, there are 745 correctly classified instances among 1000 instances. After setting the K equals to 999, the amount of the correctly classified instances decreases to 700 among 1000 instances. The changing of the K value means using K nearest neighbor(s) for classification. After changing the K value to 999, the classifier will classify all of the instances into one class, the type of the class is the most frequency class. According to the Weka summary, the relative absolute error up to 99.9578 %. Using too many nearest neighbors for classification lead the classification result under-fitting.

=== Confusion Matrix ===

a	b	<-- classified as
700	0	a = good
300	0	b = bad

Q2-4-4:

After changed the value of the weight 0,1,2, the best weight was founded after 126 times loop executed. The train accuracy reached 0.79 and the test accuracy reached 0,82.

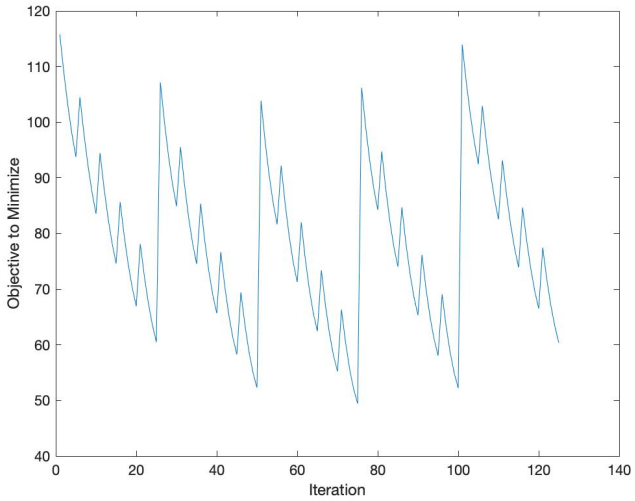
The best weight found is

- Weight 0 = 0.000
- Weight 1 = 1.000
- Weight 2 = 1.000

The code is changed to

```
1. lw0 = -1:0.5:1;
2. lw1 = -1:0.5:1;
3. lw2 = -1:0.5:1;
```

The iteration graph:



Q2-5-1-2

The original likelihood is 81.6513. After update the weights to move in the direction of the gradient, the new likelihood is 45.3826. The new train accuracy is 0.76.

The weight is

- Weight 0 = 0.066543178776321
- Weight 1 = 1.222607395813489
- Weight 2 = 1.677419405413738

Updating the weights to move in the direction of the gradient in MATLAB:

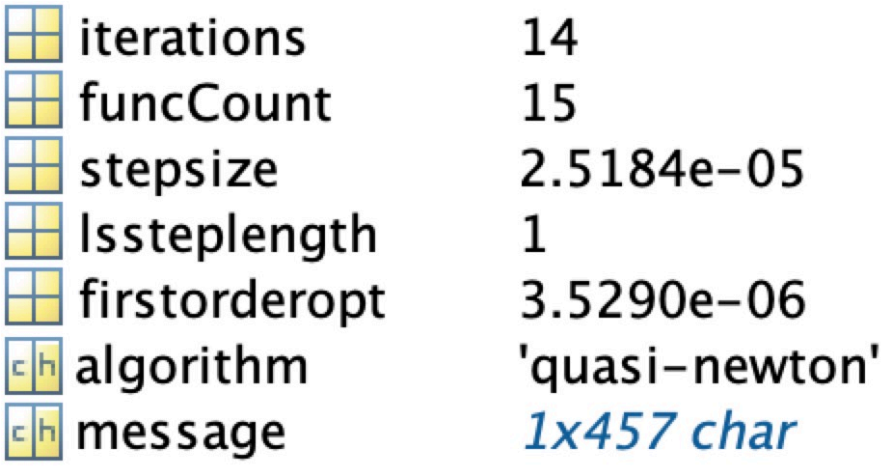
```
step = 0.1;
w = w-step*dllh;
```

Q2-6-4

The function evaluated the weights for 15 times. The train accuracy reached 0.80 and the test accuracy reached 0.81.

The best weights were founded:

- Weight 0 = -0.033142883698109
- Weight 1 = 4.666626351710899
- Weight 2 = 2.886043083158085



In Q2.4 the designed weight step size is 0.5 and iterations times is 126. The train accuracy reached 0.76. However, this function reached 0.80 train accuracy using only 15 times iteration. The step size of this function is 2.5184e-05.

3-3-c

Because of the 'knnclassify' function in MATLAB has been removed. This 'fitcknn' function in MATLAB was taken to be used to classify data.

Original code:

```
predAll = knnclassify(LX,Xtr,Ytr,K,'euclidean');
predTr = knnclassify(Xtr,Xtr,Ytr,K);
predTe = knnclassify(Xte,Xtr,Ytr,K);
```

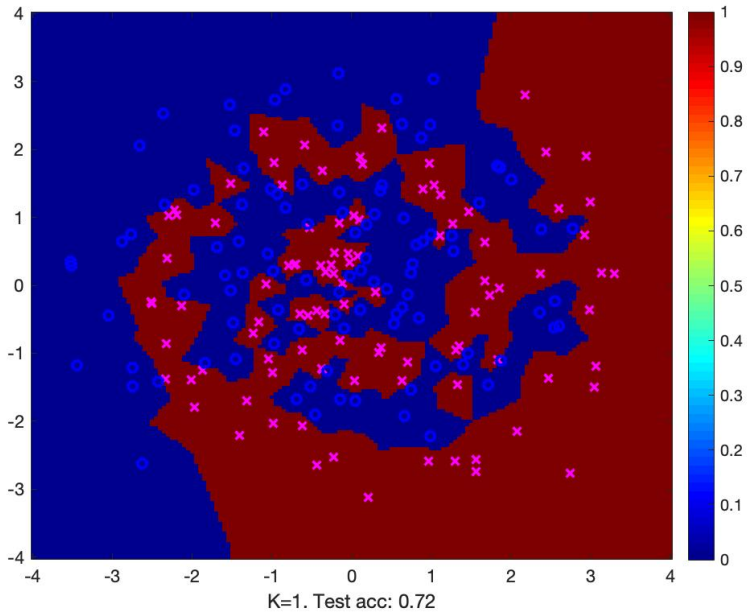
Changed to:

```
model = fitcknn(Xtr,Ytr,'NumNeighbors',K,'Distance','euclidean');
predAll = predict(model,LX);

predTr = predict(model,Xtr);

predTe = predict(model,Xte);
```

When k = 1, the best train accuracy reached 100%. (over-fitting)



However, the best test accuracy is 75%, when k = 6.