HACETTEPE UNIVERSITY

ELECTRICAL AND ELECTRONICS ENGINEERING

ELE 489-Fundamentals of Machine Learning

Homework 1

k-NN Classification

Akanay Boran Özcan

2200357021

I downloaded the wine data, normalized the data then I implemented the k-NN algorithm.

```
The accuracy of my implementation is 0.8888888888888888
The accuracy of sklearn implementation is 0.7777777777777778
```

Figure 1.

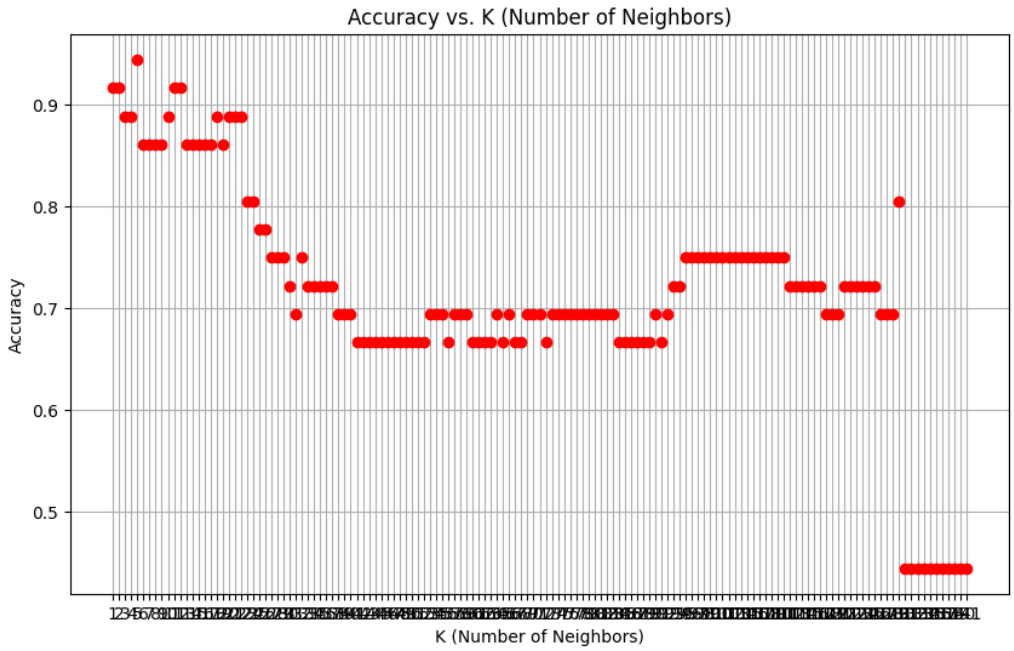

Figure 2.

Figure 1 compares the accuracies between my k-NN function and the sklearn implementation

Figure 2 shows the change in accuracy depending on the value k (k goes up to 142 there are some scaling issues in the image).

	0	1	2	...	10	11	12
count	142.000000	142.000000	142.000000	...	142.000000	142.000000	142.000000
mean	0.019859	0.003740	0.003637	...	0.001467	0.003945	0.986027
std	0.007046	0.002374	0.001380	...	0.000715	0.001940	0.009872
min	0.008941	0.001093	0.001519	...	0.000579	0.001750	0.951812
25%	0.013810	0.001730	0.002423	...	0.000970	0.002479	0.981687
50%	0.019042	0.003259	0.003468	...	0.001232	0.003394	0.988268
75%	0.024675	0.005139	0.004460	...	0.001774	0.004870	0.993172
max	0.040439	0.011652	0.007305	...	0.004105	0.010858	0.997738

Figure 3.

In figure 3 I used .describe function on the futures and mean is close to 0 but it can be closer, and variance is not equal to one as we can see.

So I used another function called StandartScaler to make the mean 0 and variance to 1 to improve the performance .

	0	1	...	11	12
count	1.420000e+02	1.420000e+02	...	1.420000e+02	1.420000e+02
mean	-1.488637e-15	3.221210e-16	...	-2.376815e-16	-8.131211e-17
std	1.003540e+00	1.003540e+00	...	1.003540e+00	1.003540e+00
min	-2.430426e+00	-1.468929e+00	...	-1.827912e+00	-1.494255e+00
25%	-7.924121e-01	-6.928989e-01	...	-9.877434e-01	-7.633656e-01
50%	6.855519e-02	-4.297234e-01	...	2.412637e-01	-2.224249e-01
75%	8.114115e-01	7.646884e-01	...	8.071624e-01	6.290406e-01
max	2.194554e+00	2.948820e+00	...	1.963262e+00	2.658188e+00

Figure 4.

In figure 4 we can see that the mean value is close to zero and the variance is nearly 1.

```
The accuracy of my implementation is 0.9722222222222222
The accuracy of sklearn implementation is 0.9444444444444444
```

Figure 5.

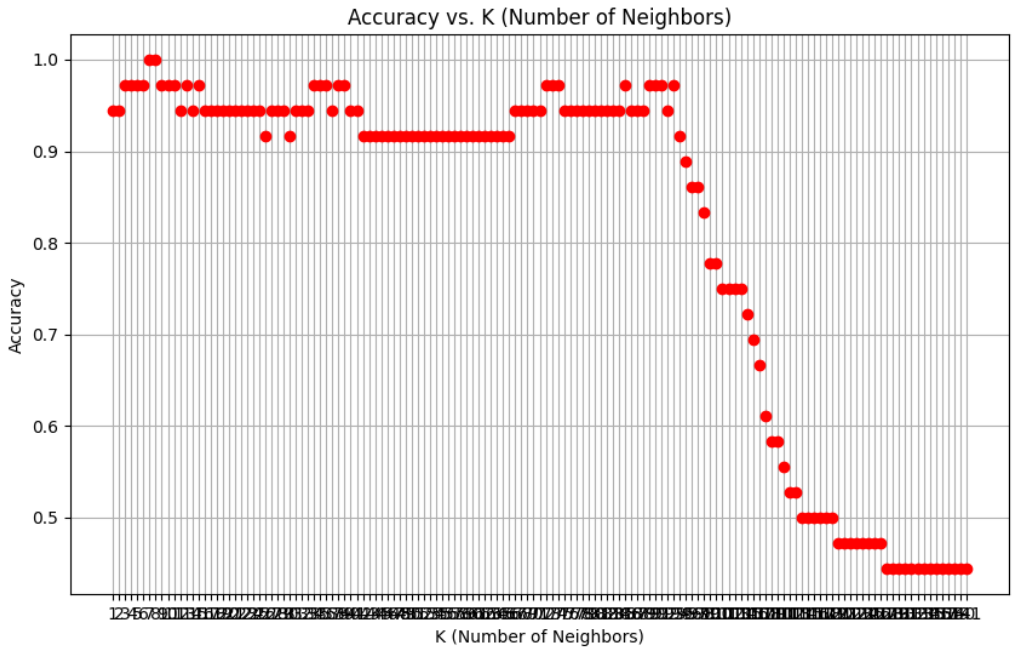

Figure 6.

In figures 5 and 6 we can see that accuracy has improved quite well.

```
Confusion Matrix for my KNN:
[[14  0  0]
 [ 0 15  1]
 [ 0  0  6]]

Confusion Matrix for Sklearn KNN:
[[14  0  0]
 [ 1 14  1]
 [ 0  0  6]]
```

Figure 7.

Figure 7 shows confusion matrix for k=3 for both my k-NN function and sklearn implementation.

It seems that in k=3 my function classified class 2 better.