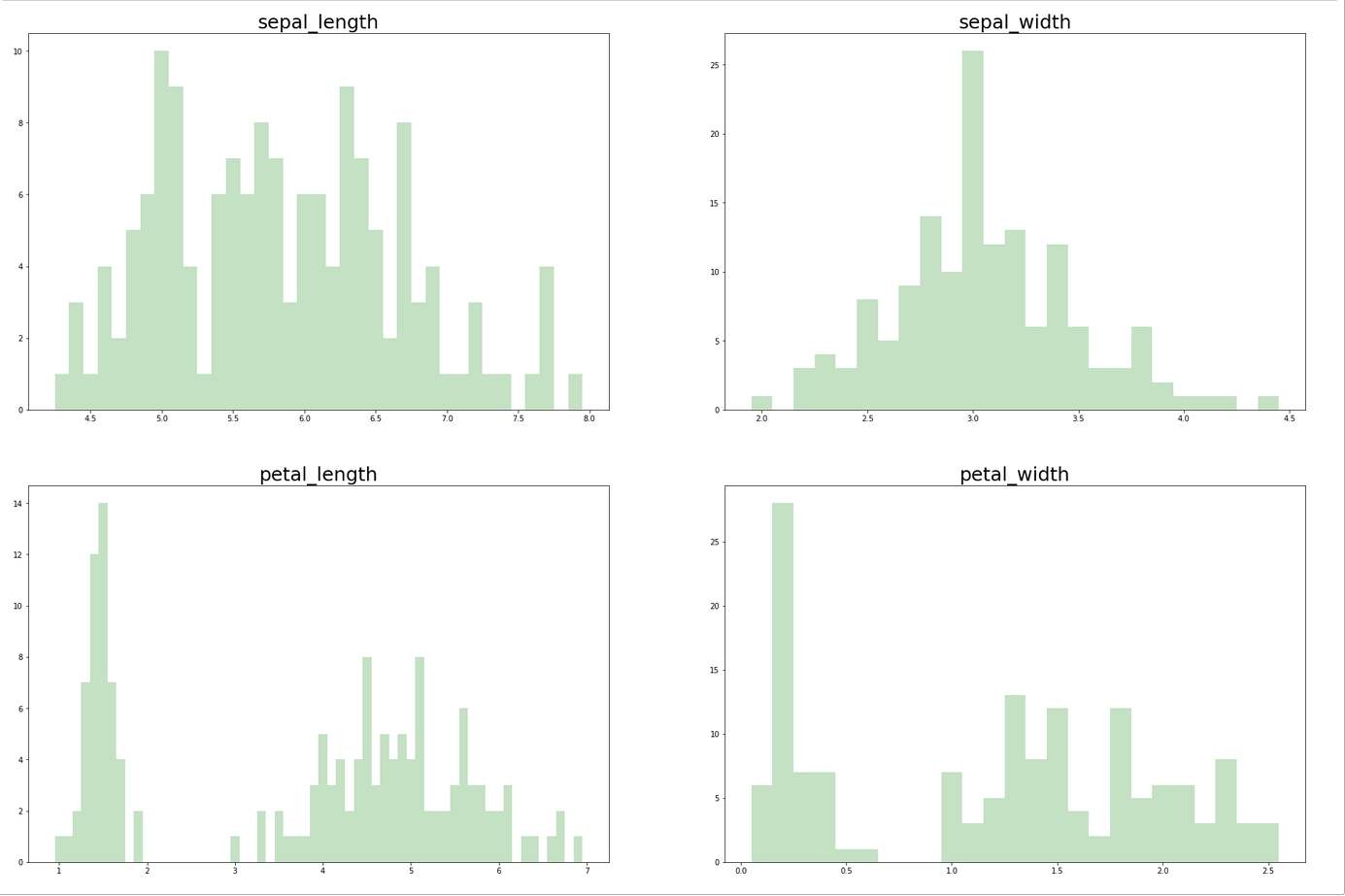
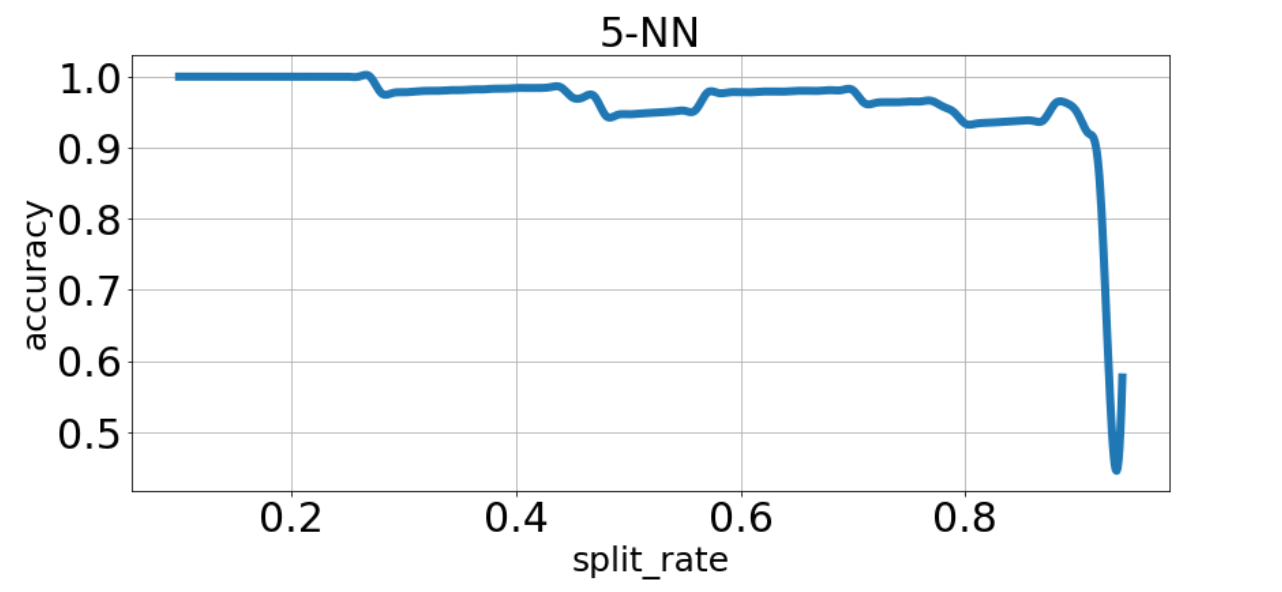
Group work 1

## One

1. Pre-analyze the data

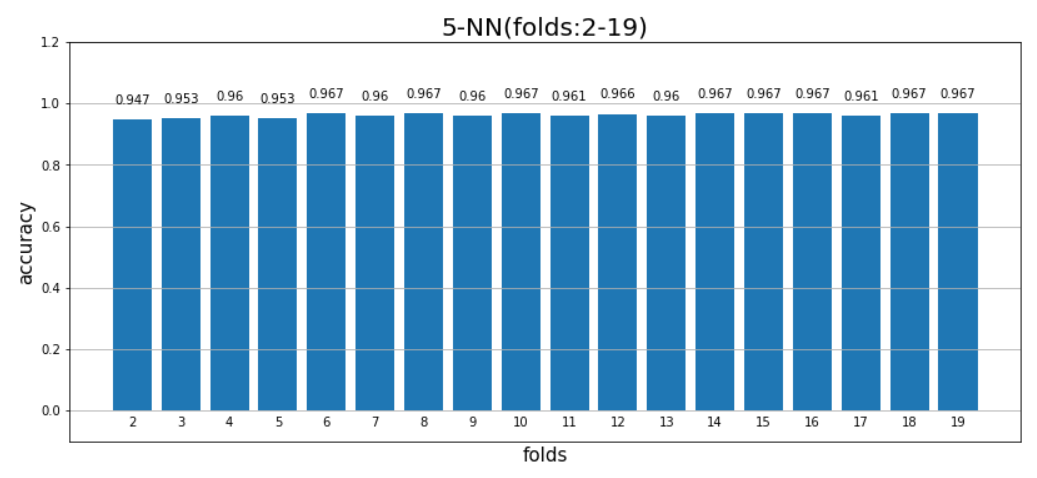


(2)The relevance between split-rate and accuracy are shown as follows。



When the prediction-accuracy goes down when the rate of test data is getting higher, and when the rate is about 90%, the prediction-accuracy goes down very fast, where there are only about 15 data to train the model.

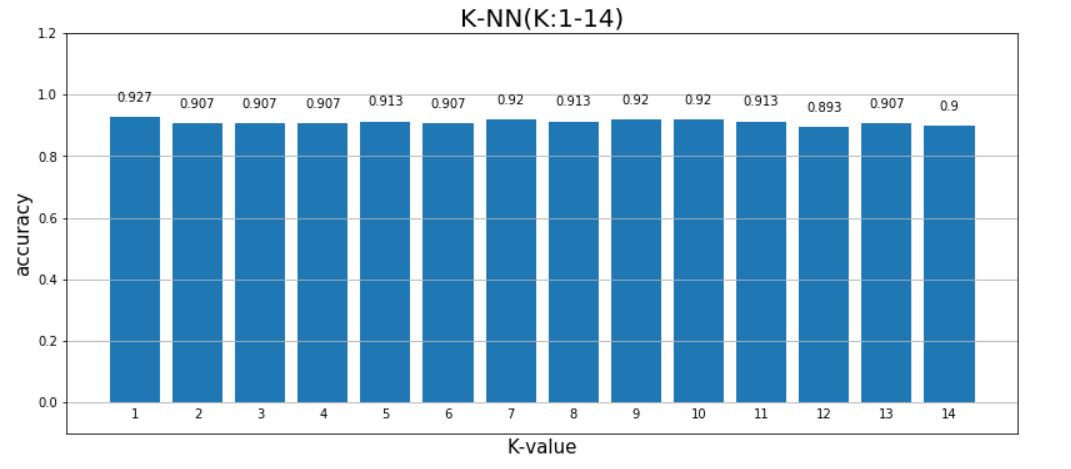
1. Change cross validation folds, the prediction-accuracy is shown as follows:



When we train the model with different folds number, the prediction-accuracy doesn’t change a lot. The range varies during 94.67% and 96.71%.

|  |  |
| --- | --- |
| K | accuracy |
| 2(min) | 94.67% |
| 10(max) | 96.71% |

1. Change the value of K, the prediction-accuracy are(folds = 5):

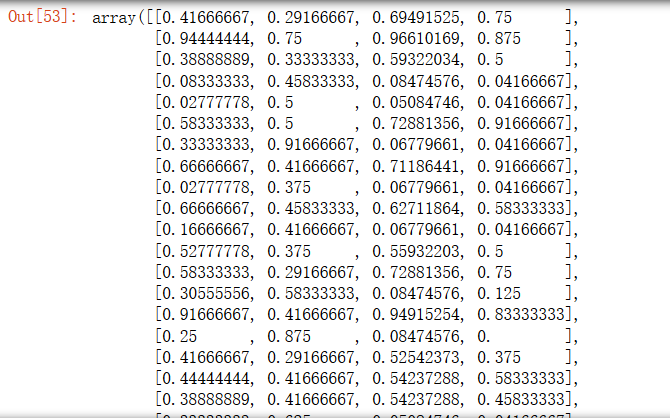


When we train the model with different value of K, the prediction-accuracy doesn’t change a lot. The range varies during 94%(K=2) and 97.33%(K=10).

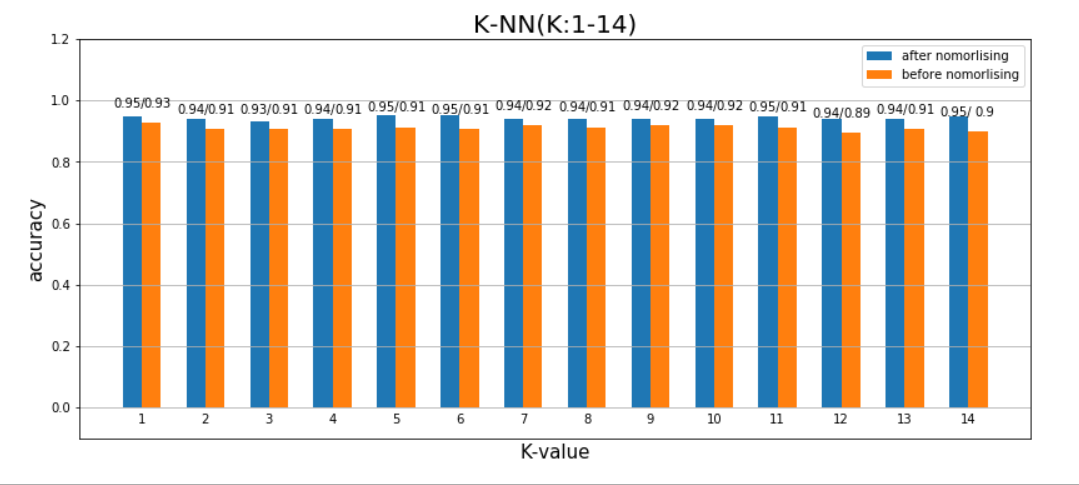
|  |  |
| --- | --- |
| K | accuracy |
| 12(min) | 89.33% |
| 1(max) | 92.66% |

(4)

Normalize the data:



We can see that their ranges are between (0,1),then we train a model with the data, using K-NN,10 folds.



We get the accuracy’s maximum and minimum are:

|  |  |
| --- | --- |
| K | accuracy |
| 5(max) | 95.33% |
| 3(min) | 93.33% |

The conclusion is that normalizing the data may get higher accuracy, but there is little influence.