

## Machine Learning Assignment 5

Name: Girish Chandar G

Roll No: 16110057

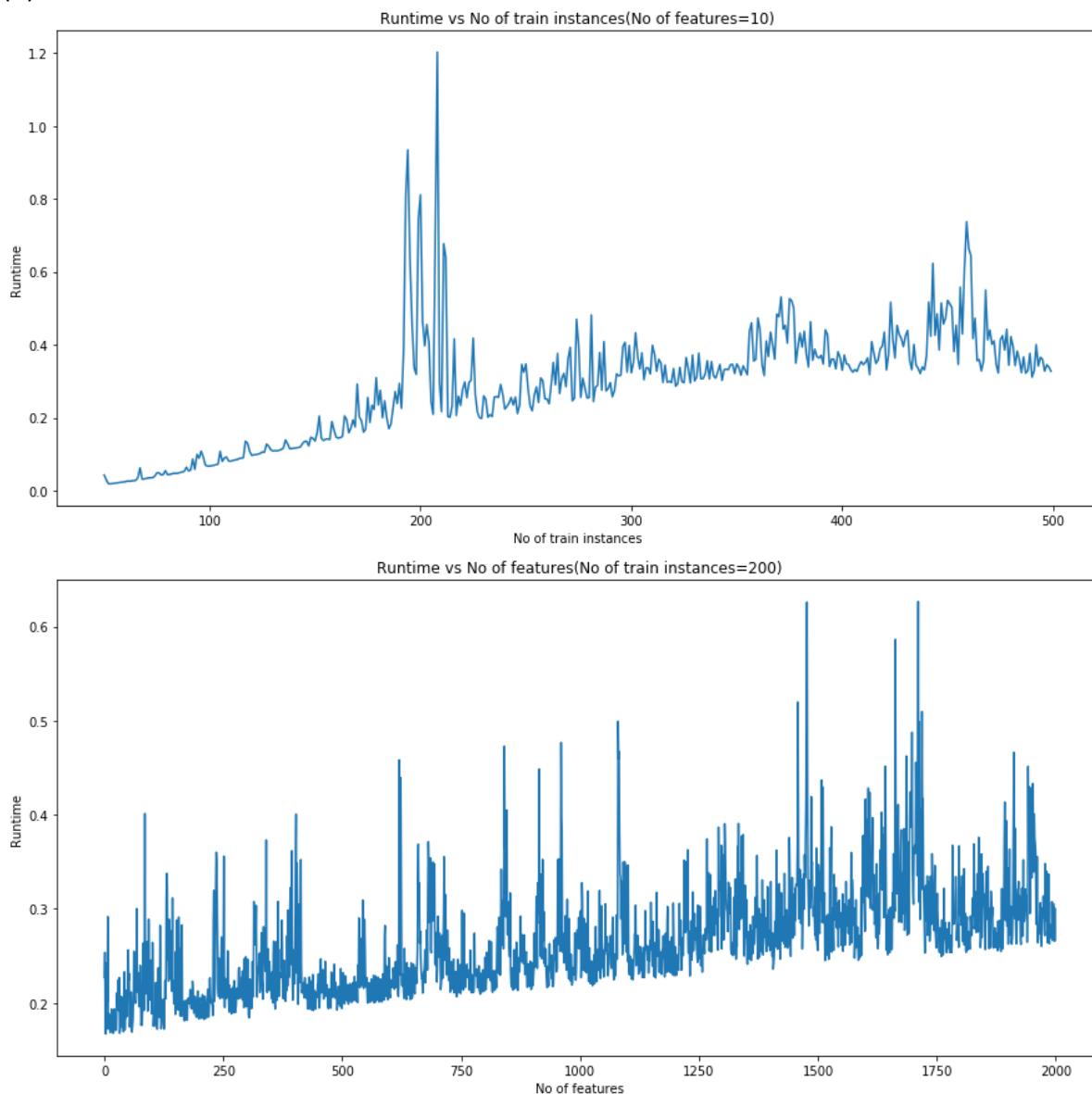
Department: Electrical Engineering

Codes for all questions has been coded in single jupyter notebook and headings given accordingly

<https://gist.github.com/girish1511/9e5fe749a17fce26ecad9201ed5e210a#file-assignment5-ipynb>

Q1.

(b)



The theoretical time complexity is  $O(n(d+f))$  where  $d$  is the number of instances and  $f$  is the number of features.

We can see that the runtime increases linearly as number of instances is increased keeping number of features fixed and runtime increases linearly as number of features is increased keeping number of instances fixed.

Thus we can observe that it follows the theoretical time complexity.

Q2.

(a)

```
Fold No  1
Optimal K:  11
Train Error:  8.153174052330494
Validation Error:  8.670054195335254
Test Error:  8.28499571521974
Fold No  2
Optimal K:  57
Train Error:  9.375371810430476
Validation Error:  7.374646350415736
Test Error:  9.28469510000683
Fold No  3
Optimal K:  2
Train Error:  5.0212950990302
Validation Error:  6.730217171141782
Test Error:  7.654063961187626
Fold No  4
Optimal K:  3
Train Error:  5.732683362665808
Validation Error:  9.6457736086581
Test Error:  8.417327109240595
Fold No  5
Optimal K:  4
Train Error:  6.447514607491878
Validation Error:  7.707828812318032
Test Error:  10.142209854450227
```

The test error is less compared to the test errors found in the previous assignments.

(b)

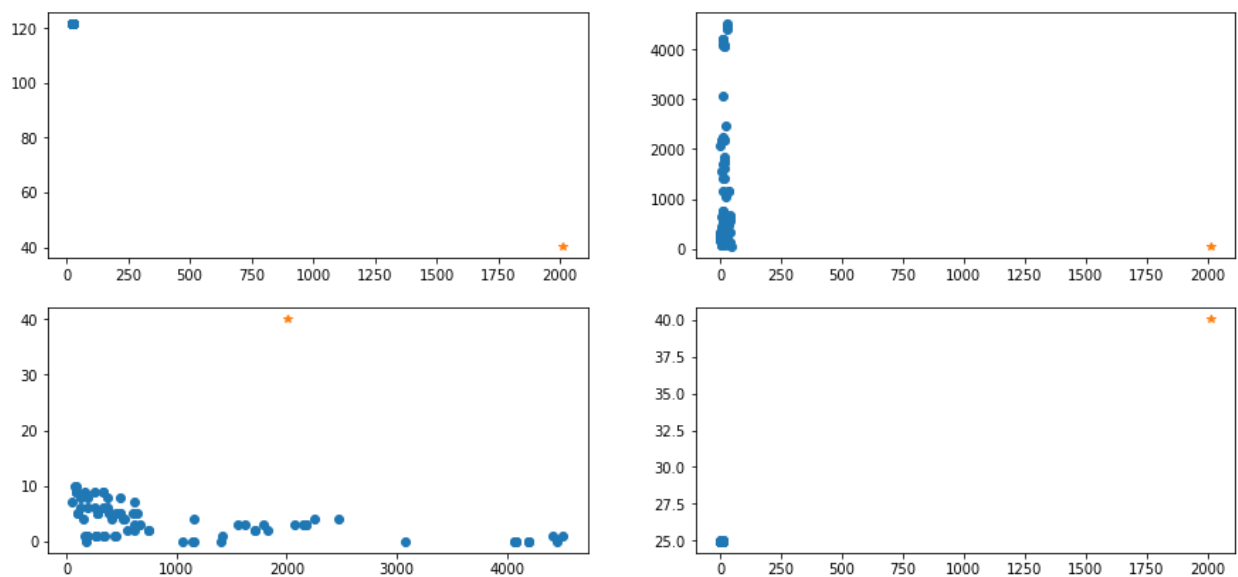
No all the features are not in the same scale. Yes, it does impact how the KNN works as it depends on the magnitudes of the distances and a feature with greater magnitude would assert more dominance.

```
Fold No  1
Optimal K:  9
Train Error:  8.483510435689702
Validation Error:  7.170728665449276
Test Error:  7.567780839538799
Fold No  2
```

Optimal K: 38  
 Train Error: 8.659505345571938  
 Validation Error: 9.940913150881055  
 Test Error: 7.641982208046778  
 Fold No 3  
 Optimal K: 4  
 Train Error: 6.685827330517588  
 Validation Error: 5.86525008303869  
 Test Error: 9.549679154360083  
 Fold No 4  
 Optimal K: 9  
 Train Error: 7.556026607347043  
 Validation Error: 10.665318060636439  
 Test Error: 6.1992480190104065  
 Fold No 5  
 Optimal K: 7  
 Train Error: 7.10789171315517  
 Validation Error: 7.636356120128864  
 Test Error: 10.776534617278031

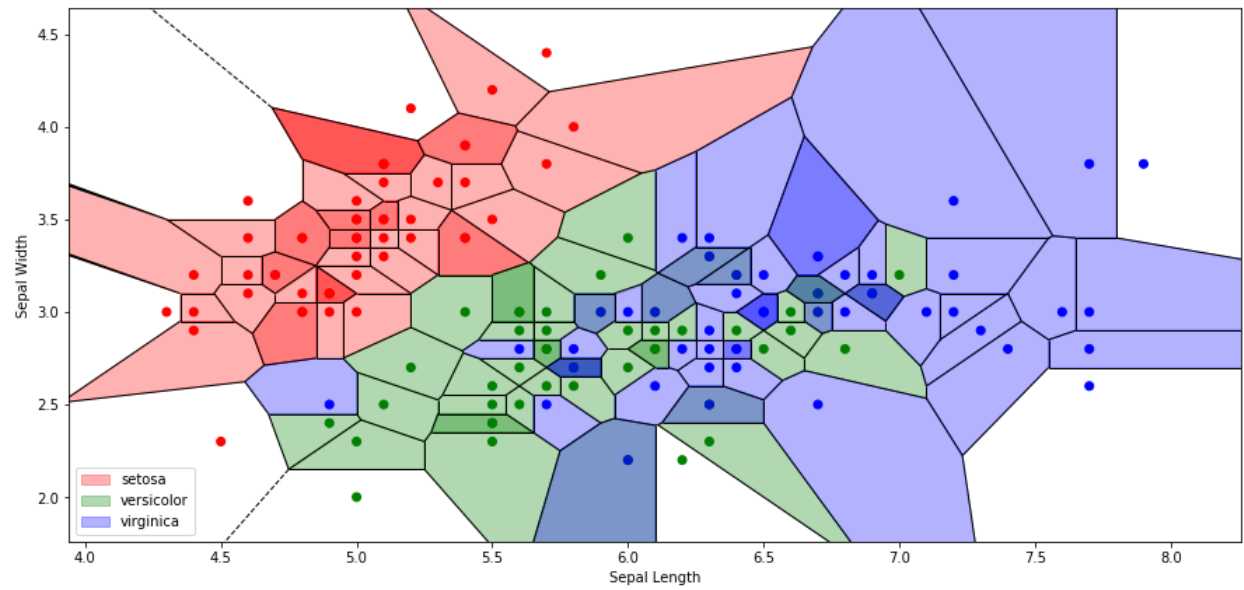
(c)

It can be inferred from the graph below that the home with the maximum RMSE is far away from the rest of the sample points. Either it is an anomaly and needs to be removed or similar sample points need to be added to the training dataset.



Q3

It can be inferred from the Voronoi diagram that 1NN over-fits data. The outliers form a cluster within a bigger cluster of a different class. While 1NN works perfectly for train data it doesn't work well for test data due to overfitting.



**References:**

<https://stackoverflow.com/questions/41244322/how-to-color-voronoi-according-to-a-color-scale-and-the-area-of-each-cell>