

**Dmitriy Nikitin**  
**dnikitin@ucsc.edu**

**Part 1**

A.)

Test Accuracy For Test: 1  
68.15

Test Accuracy For Test: 2  
64.60000000000001

Test Accuracy For Test: 3  
64.35

Test Accuracy For Test: 4  
65.2

Test Accuracy For Test: 5  
65.10000000000001

Test Accuracy For Test: 6  
65.75

Test Accuracy For Test: 7  
66.75

Test Accuracy For Test: 8  
64.64999999999999

Test Accuracy For Test: 9  
66.8

Test Accuracy For Test: 10  
65.75

Test Accuracy For Test: 11  
66.55

Test Accuracy For Test: 12  
65.60000000000001

Test Accuracy For Test: 13  
66.9

Test Accuracy For Test: 14  
68.0

Test Accuracy For Test: 15  
67.0

Test Accuracy For Test: 16  
64.95

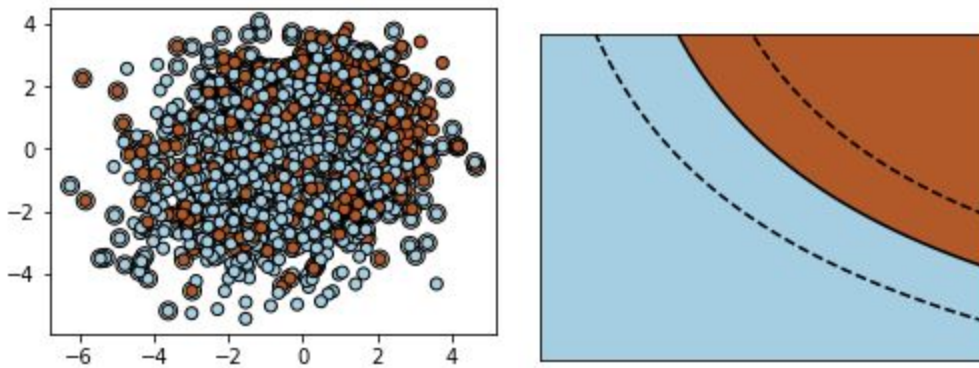
Test Accuracy For Test: 17  
65.85

Test Accuracy For Test: 18  
64.85

Test Accuracy For Test: 19  
65.55

Test Accuracy For Test: 20  
65.45

Average Test Score: 65.88999999999999



B.)

Test Accuracy For Test: 1

71.89999999999999

Test Accuracy For Test: 2

73.1

Test Accuracy For Test: 3

71.8

Test Accuracy For Test: 4

72.05

Test Accuracy For Test: 5

72.8

Test Accuracy For Test: 6

75.7

Test Accuracy For Test: 7

76.7

Test Accuracy For Test: 8

74.9

Test Accuracy For Test: 9

73.55000000000001

Test Accuracy For Test: 10

72.6

Test Accuracy For Test: 11

73.8

Test Accuracy For Test: 12

74.45

Test Accuracy For Test: 13

74.0

Test Accuracy For Test: 14

73.05

Test Accuracy For Test: 15

72.55

Test Accuracy For Test: 16

73.1

Test Accuracy For Test: 17

70.39999999999999

Test Accuracy For Test: 18

73.7

Test Accuracy For Test: 19

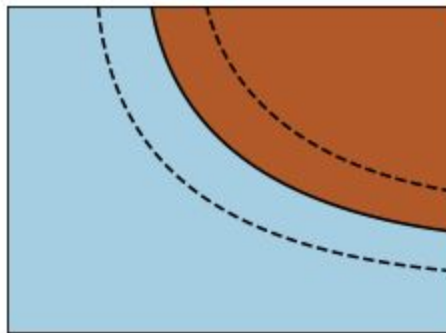
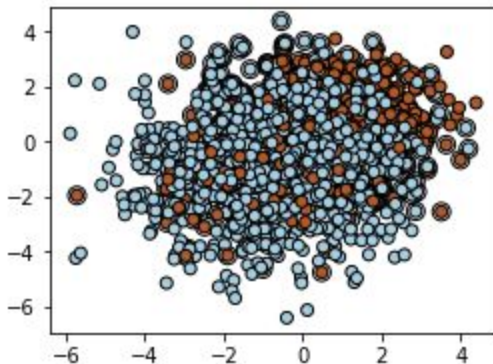
74.45

Test Accuracy For Test: 20

72.7

Average Test Score: 73.36500000000001

The average accuracy using the K-nearest neighbor algorithm improved the accuracy about 7.48 points. Therefore the K-nearest neighbor algorithm improves test results in a noisy environment.



C.)

Test Accuracy For Test using 2 clusters.

99.05000000000001

Test Accuracy For Test using 3 clusters.

98.6

Test Accuracy For Test using 4 clusters.

97.1

Test Accuracy For Test using 5 clusters.

97.6

Test Accuracy For Test using 6 clusters.

98.45

Test Accuracy For Test using 7 clusters.

97.6

Test Accuracy For Test using 8 clusters.

94.3

Test Accuracy For Test using 9 clusters.

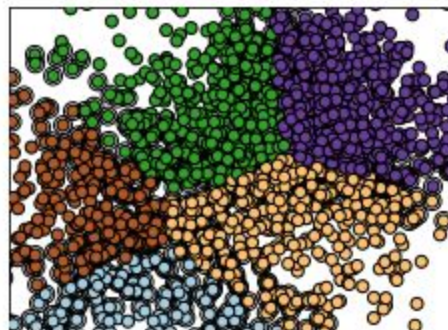
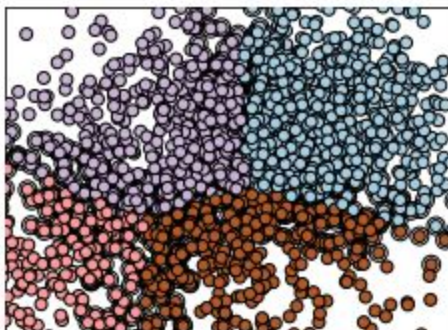
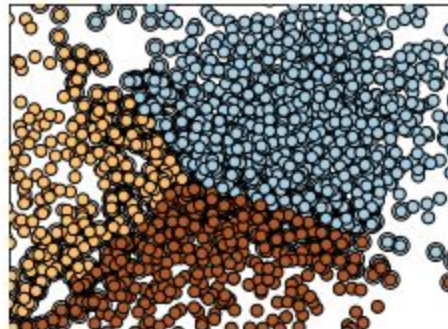
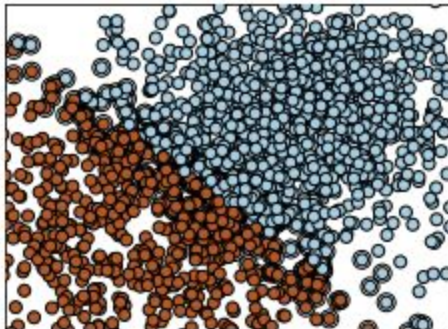
96.39999999999999

Test Accuracy For Test using 10 clusters.

93.89999999999999

Test Accuracy For Test using 11 clusters.  
94.6  
Test Accuracy For Test using 12 clusters.  
92.60000000000001  
Test Accuracy For Test using 13 clusters.  
93.8  
Test Accuracy For Test using 14 clusters.  
92.05  
Test Accuracy For Test using 15 clusters.  
92.75  
Test Accuracy For Test using 16 clusters.  
89.85  
Test Accuracy For Test using 17 clusters.  
89.05  
Test Accuracy For Test using 18 clusters.  
88.8  
Test Accuracy For Test using 19 clusters.  
89.7  
Average Test Score: 94.233333333333

Plots for C  
(3-5 cluster)



The test accuracy was a huge improvement for 2 clusters with an accuracy of 99% but, as the number of clusters increased the accuracy suffered. I believe this is because there are only 2

possible labels for the data to be classified and that is why 2 clusters was optimal for this data set.

## PART 2

a.)

Optimization terminated successfully.

Current function value: 0.365167

Iterations 7

0 0.033935

1 0.033935

2 0.033935

3 0.075014

4 0.075014

37573 0.075014

37574 0.075014

37575 0.045402

37576 0.045402

37577 0.045402

Length: 37578, dtype: float64

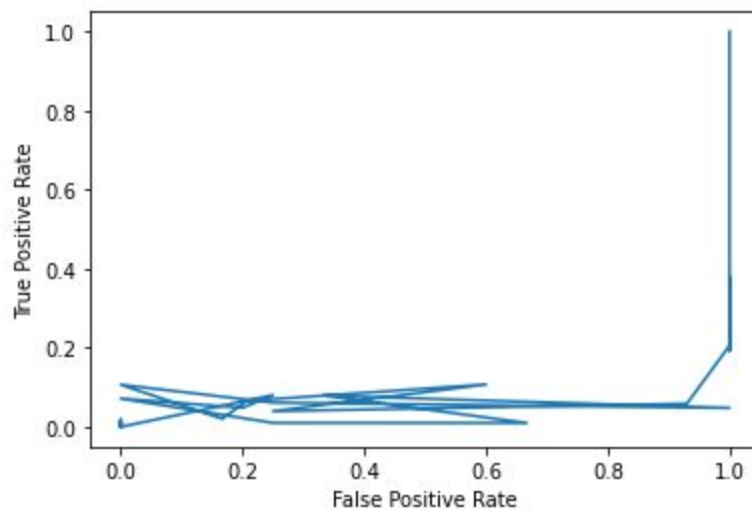
Median = 0.16361914811925177

$P(R=1 | A=w)/P(R=1 | A=b) = 0.02751894483160633$

Since the ratio is extremely less than 0.8 the model does indeed violate the disparate law.

b.)

i.



Plot for the fprs and tprs with respect to the threshold

ii.)

The minimum separation error at the given thresholds that was found is

Separation Error:

0.00048076923076922906

Pair of thresholds:

(0.12999999999999998, 0.31000000000000001)

Although i found it i do not necessarily know what they mean maybe this confirms that the data is very biased toward afircan americans hence their threshold is higher by more than 2 times the amount of the caucasian threshold. Meaning it will inaccurately classify them as being more prone to violence compared to the caucasian's.