CS 677

Hao Wu

Assignment 3

Problem 1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| class | µ(f1) | σ(f1) | µ(f2) | σ(f2) | µ(f3) | σ(f3) | µ(f4) | σ(f4) |
| 0 | 2.28 | 2.02 | 4.26 | 5.14 | 0.80 | 3.24 | -1.15 | 2.13 |
| 1 | -1.87 | 1.88 | -0.99 | 5.41 | 2.15 | 5.26 | -1.25 | 2.07 |
| all | 0.43 | 2.84 | 1.92 | 5.87 | 1.39 | 4.31 | -1.19 | 2.10 |

Problem 2

蓝色的门

中度可信度描述已自动生成蓝色的门

中度可信度描述已自动生成

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TP | FP | TN | FN | accuracy | TPR | TNR |
| 26 | 0 | 296 | 364 | 46.93 | 0.0667 | 1.0 |

Problem 3

形状

描述已自动生成

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TP | FP | TN | FN | accuracy | TPR | TNR |
| 26 | 0 | 296 | 364 | 57.6 | 0.0667 | 1.0 |

The best k value is 3

My k-NN classifier better than simple classifier for any of the measures from the previous table

When a bill x that contains the last 4 digits of your BUID as feature values Use 1-10 to represent the ID

Problem 4

Omitted “Variance”

图表, 折线图

描述已自动生成

Omitted “Skewness”

图表, 折线图

描述已自动生成

Omitted “Curtosis”

图表, 折线图

描述已自动生成

Omitted “Entropy”

图表, 折线图

描述已自动生成

As the graph, we get the lowest error rate when we omitted variance. And the best number of neighbors always in range 3-7.

When we removed “variance”, we get higher accuracy. When we removed “Curtosis”, contributed the least to loss of accuracy.

Problem 5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| TP | FP | TN | FN | accuracy | TPR | TNR |
| 402 | 4 | 278 | 2 | 99.13 | 0.9950 | 0.9858 |

logistic regression better than simple classifier for any of the measures from the previous table, accuracy rate is 99.13%

If a bill x that contains the last 4 digits of your BUID as feature values, color is also the class label for x by logistic regression and same label as predicted by K-NN.

Problem 6

Omitted “Variance”:80.18

Omitted “Skewness”:89.94

Omitted “Curtosis”:87.61

Omitted “Entropy”:99.13

Accuracy decrease in the 1,2,3 case; however, accuracy remain same in case 4.

Variance feature omitted contributed the most to loss of accuracy.

Entropy feature omitted contributed the least to loss of accuracy.

Same with using K-NN, Variance is an important feature in our model.