**Student’s Name: Rajeev Kumar**

**Roll Number: B20124**

**Mobile No: 9341062431**

**Branch: CSE**

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 81 | 27 |
| 27 | 201 |

Figure 1 KNN Confusion Matrix for K = 1

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 83 | 25 |
| 12 | 216 |

Figure 2 KNN Confusion Matrix for K = 3

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 82 | 26 |
| 9 | 219 |

Figure 3 KNN Confusion Matrix for K = 5

**b.**

Table 1 KNN Classification Accuracy for K = 1, 3 and 5

|  |  |
| --- | --- |
| **K** | **Classification**  **Accuracy (in %)** |
| 1 | **83.928** |
| 3 | **88.988** |
| 5 | **89.583** |

# Inferences:

1. The highest classification accuracy is obtained with K =5.
2. Here increasing the value of K increases the prediction accuracy.
3. Increasing the value of K increases the prediction accuracy Because the areas predicting each class will be more "smoothed", since it's the majority of the k-nearest neighbors which decide the class of any point.
4. As the classification accuracy increases with the increase in value of K then the number of diagonal elements increase
5. During testing, if the predicted class of test sample increases, then the diagonal element of the confusion matrix is incremented by 1
6. if the predicted class is incorrect , then the off-diagonal element is incremented by 1

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 107 | 1 |
| 0 | 228 |

Figure 4 KNN Confusion Matrix for K = 1 post data normalization

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 108 | 0 |
| 0 | 228 |

Figure 5 KNN Confusion Matrix for K = 3 post data normalization

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 108 | 0 |
| 0 | 228 |

Figure 6 KNN Confusion Matrix for K = 5 post data normalization

**b.**

Table 2 KNN Classification Accuracy for K = 1, 3 and 5 post data normalization

|  |  |
| --- | --- |
| **K** | **Classification**  **Accuracy (in %)** |
| 1 | **99.702** |
| 3 | **100** |
| 5 | **100** |

# Inferences:

1. Data normalization increases classification accuracy.
2. Normalization gives equal weights/importance to each variable so that no single variable steers model performance in one direction just because they are bigger numbers .
3. The highest classification accuracy is obtained with K =3 and 5.
4. increasing the value of K increases the prediction accuracy.
5. increasing the value of K increases True positive and hence the prediction accuracy will increase.
6. During testing, if the predicted class of test sample increases, then the diagonal element of the confusion matrix is incremented by 1 .
7. if the predicted class is incorrect , then the off-diagonal element is incremented by 1.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 104 | 4 |
| 15 | 213 |

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 94.34 %.

Table 3 Mean for class 0 and class 1

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Attribute Name** | **Mean** | |
| **Class 0** | **Class 1** |
|  | X\_Minimum | Nan | Nan |
|  | X\_Maximum | 2.863322e+02 | 7.465840e+02 |
|  | Y\_Minimum | Nan | Nan |
|  | Y\_Maximum | 1.711478e+06 | 1.445964e+06 |
|  | Pixels\_Areas | 7.268032e+03 | 5.835120e+02 |
|  | X\_Perimeter | 3.556148e+02 | 5.218400e+01 |
|  | Y\_Perimeter | 2.071555e+02 | 4.311200e+01 |
|  | Sum\_of\_Luminosity | 8.086157e+05 | 6.155241e+04 |
|  | Minimum\_of\_Luminosity | 5.340283e+01 | 9.480400e+01 |
|  | Maximum\_of\_Luminosity | 1.358587e+02 | 1.301840e+02 |
|  | Length\_of\_Conveyer | 1.382516e+03 | 1.486630e+03 |
|  | TypeOfSteel\_A300 | Nan | Nan |
|  | TypeOfSteel\_A400 | 9.964664e-01 | 6.220000e-01 |
|  | Steel\_Plate\_Thickness | 4.024735e+01 | 1.004340e+02 |
|  | Edges\_Index | 1.264470e-01 | 3.888644e-01 |
|  | Empty\_Index | 4.496078e-01 | 4.186428e-01 |
|  | Square\_Index | 5.932530e-01 | 5.103224e-01 |
|  | Outside\_X\_Index | 1.081728e-01 | 1.985380e-02 |
|  | Edges\_X\_Index | 5.658509e-01 | 6.256006e-01 |
|  | Edges\_Y\_Index | 5.246919e-01 | 8.374430e-01 |
|  | Outside\_Global\_Index | 2.685512e-01 | 6.110000e-01 |
|  | LogOfAreas | 3.599567e+00 | 2.264311e+00 |
|  | Log\_X\_Index | 2.048011e+00 | 1.214075e+00 |
|  | Log\_Y\_Index | 1.825003e+00 | 1.299494e+00 |
|  | Orientation\_Index | -3.280714e-01 | 1.319460e-01 |
|  | Luminosity\_Index | -1.090739e-01 | -1.226320e-01 |
|  | SigmoidOfAreas | 9.158700e-01 | 5.270244e-01 |

In Fig. 8 and 9 representing covariance matrices for class 0 and class 1 respectively the column numbers and row numbers correspond to attribute with serial number as in Table 3.

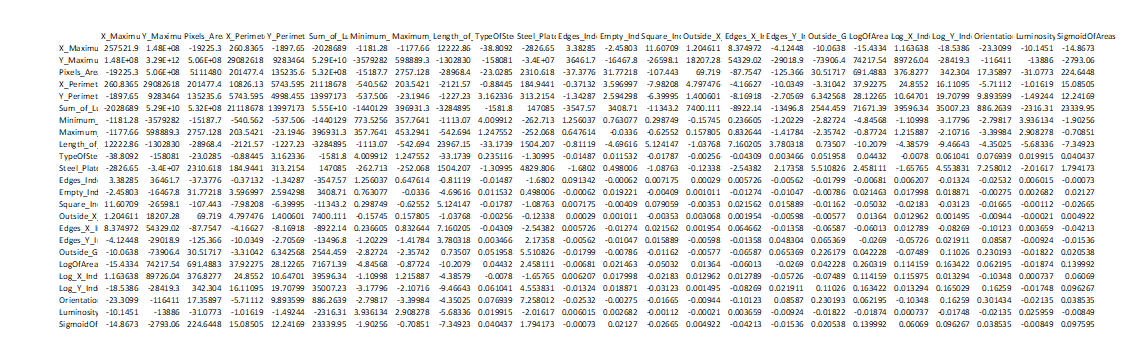
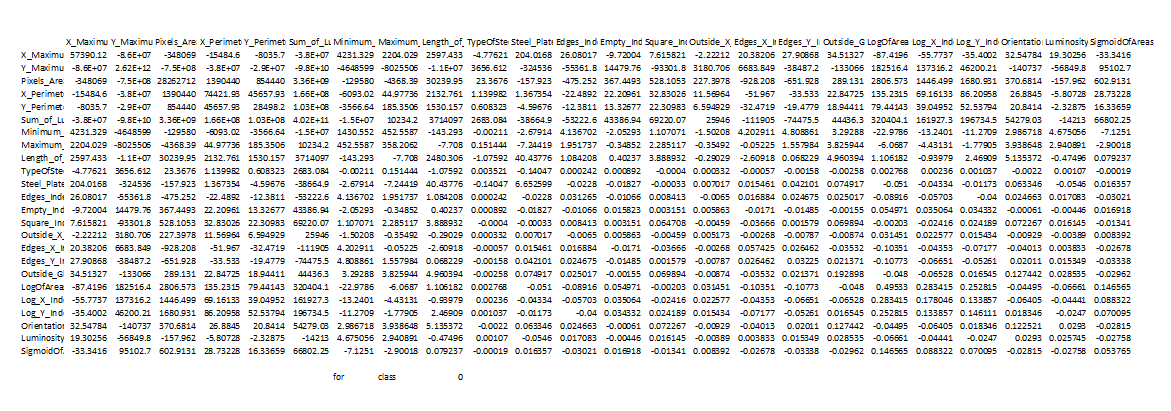


Figure 8: Covariance matrix for class 0

Figure 9: Covariance matrix for class 1

# Inferences:

1. Accuracy of Bayes Classifier is 94.34 percent .it is lesser than previous classification approaches because it doesn't assume independence between attributes/variables .
2. The diagonal elements of the matrix contain the variances of the variables and the off-diagonal elements contain the co variances between all possible pairs of variables . Because row column matrix multiplication of mean subtracted data and it’s transpose has expression same as variance and covariance.

Table 4 Comparison between classifiers based upon classification accuracy

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Classifier** | **Accuracy (in %)** |
|  | KNN | 83.9 |
|  | KNN on normalized data | 99.7 |
|  | Bayes | 94.34 |

# Inferences:

1. KNN on normalized data and Knn has highest and lowest accuracy respectively.
2. classifiers in ascending order of classification accuracy- KNN ,Bayes,KNN on normalized data .