**Student’s Name:** Sachin Mahawar

**Roll Number:** B20129

**Mobile No:** 9166843951

**Branch:** CSE

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 93 | 25 |
| 19 | 200 |

Figure 1 KNN Confusion Matrix for K = 1

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 92 | 26 |
| 9 | 210 |

Figure 2 KNN Confusion Matrix for K = 3

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 92 | 26 |
| 10 | 209 |

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 | **86.944 %** |
| 3 | **89.614 %** |
| 5 | **89.318 %** |

# Inferences:

1. The highest classification accuracy is obtained with K = 3.
2. Here no monotonic change is seen between k and accuracy as it increases for k=1 to k=3 then decreases for k=3 to k=5.
3. Increasing the value of k increases the accuracy because when we increase k we take into account more samples which reduces effect of noise thus decreasing error.
4. As the value of k increases, the number of diagonal elements increase.
5. Number of diagonal elements increases because they represent the number of test samples predicted corrected.

As the error decreases, accuracy increases, diagonal elements also increases.

1. With increase in value of k, accuracy increases thus number of diagonal elements.
2. With increase in value of k, we can see that the number of off-diagonal values decreases.
3. Off diagonal elements represents samples which are wrongly predicted so with increase in k, the error reduced and so the count of off diagonal elements also decreased.

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 111 | 7 |
| 6 | 213 |

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

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 112 | 6 |
| 4 | 215 |

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

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 11 | 6 |
| 3 | 216 |

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 | **96.142 %** |
| 3 | **97.033 %** |
| 5 | **97.329 %** |

# Inferences:

1. Clearly, we can see that by normalizing the data, classification accuracy increases.
2. Accuracy increases because KNN is based on Euclidean distance in which error increases when any attribute have very high range.
3. The highest classification accuracy is obtained with K = **97.329 %**.
4. Increasing the value of k increases the accuracy because when we increase k we take into account more samples which reduces effect of noise thus decreasing error.
5. Number of diagonal elements increases because they represent the number of test samples predicted corrected.

As the error decreases, accuracy increases, diagonal elements also increases.

1. With increase in value of k, accuracy increases thus number of diagonal elements.
2. With increase in value of k, we can see that the number of off-diagonal values decreases.
3. Off diagonal elements represents samples which are wrongly predicted so with increase in k, the error reduced and so the count of off diagonal elements also decreased.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 102 | 16 |
| 3 | 216 |

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is **94.362** %.

Table 3 Mean for class 0 and class 1

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Attribute Name** | **Mean** | |
| **Class 0** | **Class 1** |
|  | X\_Maximum | 273.4176 | 723.6562 |
|  | Y\_Maximum | 1583170 | 1431589 |
|  | Pixels\_Areas | 7779.663 | 585.9666 |
|  | X\_Perimeter | 393.8352 | 54.49116 |
|  | Y\_Perimeter | 273.1832 | 45.65815 |
|  | Sum\_of\_Luminosity | 843350.3 | 62191.13 |
|  | Minimum\_of\_Luminosity | 53.32601 | 96.23576 |
|  | Maximum\_of\_Luminosity | 135.7619 | 130.4519 |
|  | Length\_of\_Conveyer | 1382.762 | 1480.018 |
|  | Steel\_Plate\_Thickness | 40.07326 | 104.2141 |
|  | Edges\_Index | 0.122571 | 0.385421 |
|  | Empty\_Index | 0.459035 | 0.42658 |
|  | Square\_Index | 0.591617 | 0.51286 |
|  | Outside\_X\_Index | 0.107915 | 0.019595 |
|  | Edges\_X\_Index | 0.549996 | 0.608297 |
|  | Edges\_Y\_Index | 0.522965 | 0.830538 |
|  | Outside\_Global\_Index | 0.287546 | 0.608055 |
|  | LogOfAreas | 3.622767 | 2.28743 |
|  | Log\_X\_Index | 2.056679 | 1.226597 |
|  | Log\_Y\_Index | 1.848037 | 1.317861 |
|  | Orientation\_Index | -0.31361 | 0.136379 |
|  | Luminosity\_Index | -0.11465 | -0.11567 |
|  | SigmoidOfAreas | 0.925434 | 0.54347 |
|  | X\_Maximum | 273.4176 | 723.6562 |
|  | Y\_Maximum | 1583170 | 1431589 |
|  | Pixels\_Areas | 7779.663 | 585.9666 |
|  | X\_Perimeter | 393.8352 | 54.49116 |

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.

Table

Description automatically generated

Figure 8: Covariance matrix for class 0

Figure 8: Covariance matrix for class 0

Table

Description automatically generated

# Inferences:

1. Accuracy of Bayes classifier is **94.362 %.** It is less accurate compared to the other classifications because Bayes model with unimodal gaussian distribution is simplest classifier which cannot be used on real life data as they are complex. But if we would have used multimodal Gaussian distribution accuracy would be higher.
2. The data along the diagonal element represent the variance of that particular attribute.
3. Pairs with maximum covariance : (Y\_Maximum and SUM\_Of\_Luminosity ); (SUM\_Of\_Luminosity and Pixel\_Areas) Pairs with minimum covariance : (Square\_Index and Edges\_Y\_Index); (Square\_Index and LogOfAreas)

Table 4 Comparison between classifiers based upon classification accuracy

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Classifier** | **Accuracy (in %)** |
|  | KNN | **89.614 %** |
|  | KNN on normalized data | **97.329 %** |
|  | Bayes | **94.362 %** |

# Inferences:

1. The classifiers with highest accuracy is KNN on normalized data and lowest accuracy is KNN with nonnormalized data.
2. Bayes classifier < KNN on nonnormalized data < KNN on normalized data.
3. As bayes is very simple classifier but given data is real world complex its accuracy is less than KNN.
4. As KNN is based on Euclidean distance so by normalizing data we are less prone to error that’s why normalized data has more accuracy than nonnormalized data.