

# Data classification using K-nearest neighbor classifier and Bayes classifier with unimodal Gaussian density

Student's Name: Priyanka Kumari Mobile No: 8328354314

Roll Number: B20307 Branch: Electrical Engineering

1 a.

	Prediction Outcome		
Label	93	25	
True Label	19	200	

Figure 1 KNN Confusion Matrix for K = 1

	Prediction Outcome		
Label	92	26	
True	9	210	

Figure 2 KNN Confusion Matrix for K = 3



# Data classification using K-nearest neighbor classifier and Bayes classifier with unimodal Gaussian density

	Prediction Outcome		
Label	92	26	
True Label	10	209	

Figure 3 KNN Confusion Matrix for K = 5

b.

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

К	Classification Accuracy (in %)
1	86.9
3	89.6
5	89.3

#### Inferences:

- 1. The highest classification accuracy is obtained with K = 3
- 2. The k value initially increases from k=1 and k=3 and then decreases for k=5.
- 3. On increasing the value of k the data gets more accurate on increasing the larger values of k it
- 4. As the classification accuracy increases with the increase in value of K the number of diagonal elements increase.
- 5. In general diagonal elements tells how the data is correctly assigned to that particular class such that the accuracy increases it means number of diagonal elements increases.
- 6. As the classification accuracy increases with the increase in value of k the number of off-diagonal elements decreases as the off-diagonal elements indicate the elements which are not correctly predicted.
- 7. The off-diagonal elements indicate which are not correctly predicted so accuracy increases the off-diagonal elements decreases.



# Data classification using K-nearest neighbor classifier and Bayes classifier with unimodal Gaussian density

#### 2 a.

	Prediction Outcome		
Label	111	7	
True Label	6	213	

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

	Prediction Outcome		
Label	113	5	
True Label	4	215	

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

	Prediction Outcome		
Label	109	9	
True Label	4	215	

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



## Data classification using K-nearest neighbor classifier and Bayes classifier with unimodal Gaussian density

b.

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

К	Classification Accuracy (in %)
1	96.1
3	97.3
5	96.1

#### Inferences:

- 1. The data normalization increases classification accuracy.
- 2. Data normalization centers the data and removes the noise, so the accuracy increases. But it is not true for every case. Sometimes , it removes the important features from the data and results in accuracy going down as can we see for K = 5.
- 3. The highest classification accuracy is obtained with K = 3
- 4. We can see an increase in the prediction accuracy with increasing value of K, with some exceptions like K=5
- 5. Increasing the value of K increases the prediction accuracy because the number of nearest neighbors, we can get the more number of classification as we increase the value of K so in Result its improving the accuracy.
- 6. With the increase in value of K the number of diagonal elements also increase.
- 7. As the accuracy is increasing with the increase in K, the True positive and True Negative are also increasing, which are the diagonal elements.
- 8. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements are decrease.
- 9. As the accuracy increase the False positive and False negative decreases, which are the off-diagonal elements.

3

	Prediction Outcome		
Label	109	9	
True Label	9	210	

Figure 7 Confusion Matrix obtained from Bayes Classifier



# Data classification using K-nearest neighbor classifier and Bayes classifier with unimodal Gaussian density

The classification accuracy obtained from Bayes Classifier is  $\,$  94.7  $\,$ %.

Table 3 Mean for class 0 and class 1

S. No.	Attribute Name	e Name Mean	
		Class 0	Class 1
1.	X_Minimum	-	
2.	X_Maximum	273.418	723.656
3.	Y_Minimum	-	
4.	Y_Maximum	1583169.659	1431588.690
5.	Pixels_Areas	273.183	585.967
6.	X_Perimeter	843350.275	54.491
7.	Y_Perimeter	53.326	45.658
8.	Sum_of_Luminosity	135.762	62191.126
9.	Minimum_of_Luminosity	1382.762	96.236
10.	Maximum_of_Luminosity	40.073	130.452
11.	Length_of_Conveyer	0.123	1480.018
12.	TypeOfSteel_A300	ı	-
13.	TypeOfSteel_A400	ı	-
14.	Steel_Plate_Thickness	40.073	104.214
15.	Edges_Index	0.123	0.385
16.	Empty_Index	0.459	0.427
17.	Square_Index	0.592	0.513
18.	Outside_X_Index	0.108	0.020
19.	Edges_X_Index	0.550	0.608
20.	Edges_Y_Index	0.523	0.831
21.	Outside_Global_Index	0.288	0.608
22.	LogOfAreas	3.623	2.287
23.	Log_X_Index	2.057	1.227
24.	Log_Y_Index	1.848	1.318
25.	Orientation_Index	-0.314	0.136
26.	Luminosity_Index	-0.115	-0.116
27.	SigmoidOfAreas	0.925	0.543

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.



## Data classification using K-nearest neighbor classifier and Bayes classifier with unimodal Gaussian density

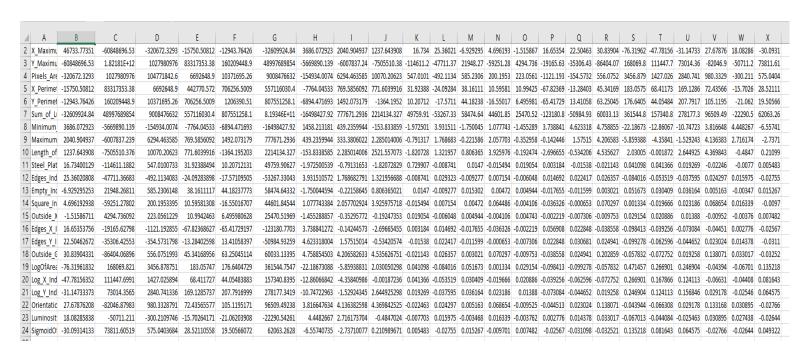
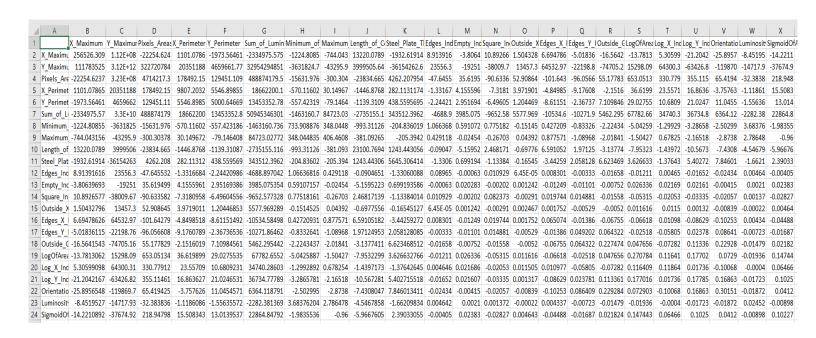


Figure 8: Covariance matrix for class 0





## Data classification using K-nearest neighbor classifier and Bayes classifier with unimodal Gaussian density

Figure 9: Covariance matrix for class 1

#### Inferences:

- 1. The accuracy of Bayes Classifier is 94.7. it is lesser than previous classification approaches because it considers normal distribution.
- 2. Nature of value along the diagonal is high for some and low for others because some particular attributes follow Standard normal distribution, hence the prediction is better for them.
- 3. Off-diagonal elements have varying values. Two pairs of attributes with maximum cov is (Sum\_of\_Luminosity, Y\_Maximum) and (Pixels\_Areas, Sum\_of\_Luminosity).

4

Table 4 Comparison between classifiers based upon classification accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN	94.7
2.	KNN on normalized data	96.100
3.	Bayes	94.659

#### Inferences:

- 1. Highest accuracy is :- KNN on normalized data : 96.100, lowest accuracy:- Bayes: 94.659
- 2. KNN on normalized data > KNN > Bayes.
- 3. Usually Bayes classifier has higher accuracy but in this case the KNN classifier used is using Normalised data points that is why its accuracy is high.