IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – IV

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

Student's Name: name here
Mobile No.: Mobile No. here
Roll Number: Roll No. here
Branch: Branch here

1

a.

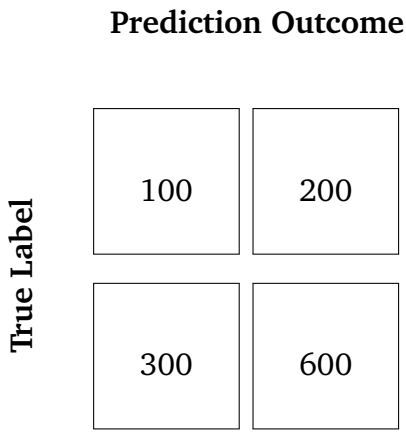

Figure 1: KNN confusion matrix for K = 1

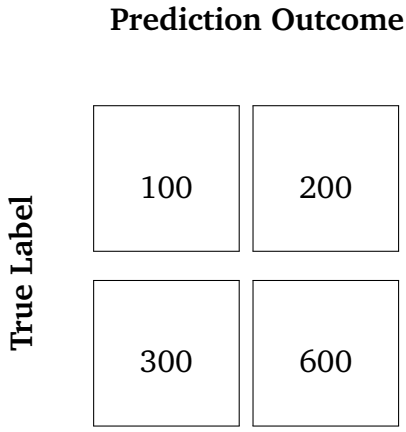

Figure 2: KNN confusion matrix for K = 3

Prediction Outcome

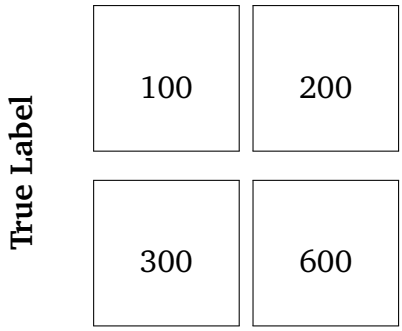

Figure 3: KNN confusion matrix for K = 5

b.

Table 1: KNN classification accuracy for K = 1,3 and 5

K value	**Classification accuracy (in %)**
1	
3	
5	

Inferences:

1. The highest classification accuracy is obtained with K =.
2. Infer whether increasing the value of K increases/decreases the prediction accuracy.
3. State a suitable reason why increasing the value of K increases/decreases the prediction accuracy.
4. As the classification accuracy increases/decreases with the increase in value of K infer does the number of diagonal elements increase/decrease.
5. State the reason for increase/decrease in diagonal elements.
6. As the classification accuracy increases/decreases with the increase in value of K infer does the number of off-diagonal elements increase/decrease.
7. State the reason for increase/decrease in off-diagonal elements.
8. Inference 8 (You may add or delete the number of inferences).
   Note: Dummy values have been filled in the confusion matrix. Replace it with values obtained by you.

2

a.

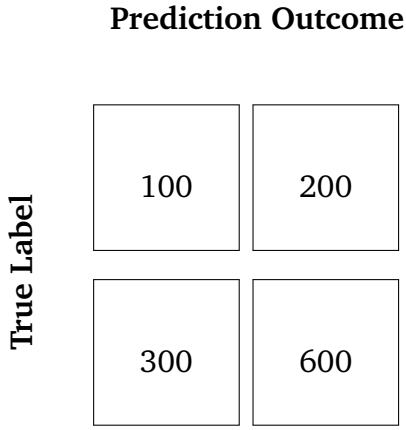

Figure 4: KNN confusion matrix for K = 1 post data normalization

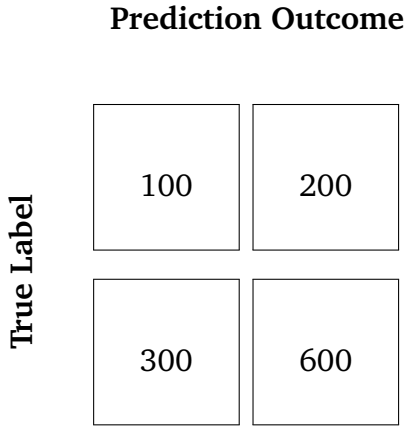

Figure 5: KNN confusion matrix for K = 3 post data normalization

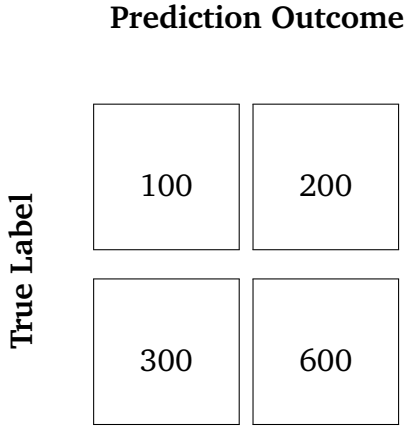

Figure 6: KNN confusion matrix for K = 5 post data normalization

b.

Table 2: KNN classification accuracy for K = 1, 3 and 5 post data normalization

K value	Classification accuracy (in %)
1	
3	
5	

Inferences:

1. Infer whether data normalization increases/decreases classification accuracy.
2. The highest classification accuracy is obtained with K =.
3. Infer whether increasing the value of K increases/decreases the prediction accuracy.
4. State a suitable reason why increasing the value of K increases/decreases the prediction accuracy.
5. As the classification accuracy increases/decreases with the increase in value of K infer does the number of diagonal elements increase/decrease.
6. State the reason for increase/decrease in diagonal elements.
7. As the classification accuracy increases/decreases with the increase in value of K infer does the number of off-diagonal elements increase/decrease.
8. State the reason for increase/decrease in off-diagonal elements.
9. Inference 8 (You may add or delete the number of inferences).
   Note: Dummy values have been filled in the confusion matrix. Replace it with values obtained by you.

3

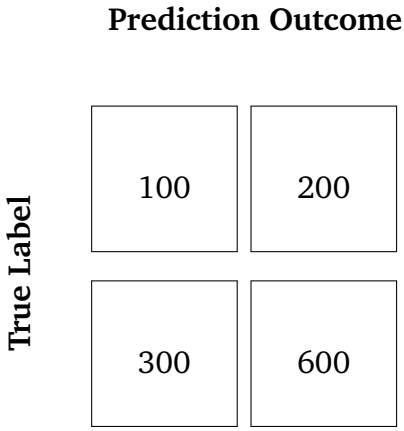

Figure 7: Confusion matrix obtained from Bayes classifier

The classification accuracy obtained from Bayes classifier is %.

Table 3: Mean for class 0 and class 1

S. No.	Attributes	Mean	
		Class 0	Class 1
1	X_Minimum		
2	X_Maximum		
3	Y_Minimum		
4	Y_Maximum		
5	Pixels_Areas		
6	X_Perimeter		
7	Y_Perimeter		
8	Sum_of_Luminosity		
9	Minimum_of_Luminosity		
10	Maximum_of_Luminosity		
11	Length_of_Conveyer		
12	TypeOfSteel_A300		
13	TypeOfSteel_A400		
14	Steel_Plate_Thickness		
15	Edges_Index		
16	Empty_Index		
17	Square_Index		
18	Outside_X_Index		
19	Edges_X_Index		
20	Edges_Y_Index		
21	Outside_Global_Index		
22	LogOfAreas		
23	Log_X_Index		
24	Log_Y_Index		
25	Orientation_Index		
26	Luminosity_Index		
27	SigmoidOfAreas		

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.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	1	2	3	4	5	6	7	8	9	10	10	11	12	13	14	15	16	18	19	18	19	20	21	22	23	24	24
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
8	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
9	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
10	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
12	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
13	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
14	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
16	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
19	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
21	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
24	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
25	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
27	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

Figure 8: Covariance Matrix for Class 0

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	1	2	3	4	5	6	7	8	9	10	10	11	12	13	14	15	16	18	19	18	19	20	21	22	23	24	24
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
8	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
9	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
10	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
12	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
13	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
14	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
16	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
19	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
21	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
24	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
25	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
27	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

Figure 9: Covariance Matrix for Class 1

Inferences:

1. Write the accuracy of Bayes classifier and state reason why it is lesser / greater than previous classification approaches.
2. Infer from covariance matrix the nature of values along the diagonal. State the reason.
3. Infer from off-diagonal elements the covariance between attributes. Write 2 pair of attributes with maximum and 2 pair of attributes with minimum covariance
   Note: Please write diagonal values of covariance matrices in boldface. On moodle, the template for covariance matrix is uploaded as .docx and .xlsx format. Fill in the values and change the covariance matrices into images. Insert the covariance matrices as images to the document.

4

Table 4: Comparison between classifiers based upon classification accuracy

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

Inferences:

1. Mention the classifiers with highest and lowest accuracy.
2. Arrange the classifiers in ascending order of classification accuracy. Classifier a $<$ Classifier b $<$ Classifier c $<$ Classifier d $<$ Classifier e.
3. State the reasons behind Inference 1 and 2. Any other inference (You may add or delete the number of inferences).

Guidelines for Report (Delete this while you submit the report):

- **The plot/graph/figure/table should be centre justified with sequence number and caption.**
- **Inferences should be written as a numbered list.**
- **Use specific and technical terms to write inferences**
- **Values observed/calculated should be rounded off to three decimal places.**
- **The quantities which have units should be written with units.**
- **Please fit a confusion matrix in one page only.**