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: Arpit Singh **Mobile No: 6265104315**

Roll Number: B20084 **Branch: CSE**

1 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**
4	**89.318**

Inferences:

1. The highest classification accuracy is obtained with K = 3.
2. No, increasing the value of K does not increase/decrease the prediction accuracy always. In this case it increases and becomes almost constant.
3. Increasing the value of K means an increase in data we are using to predict accuracy therefore increase in accuracy.
4. As the classification accuracy increases with the increase in value of K the no. of diagonal increase because the diagonal have true negative and true positive values.
5. Increase in diagonal elements is observed when the model predicts the correct class.
6. Number of off-diagonal elements depends upon the incorrect value of model prediction. It increases when it predicts incorrectly.

2 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	112	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. Data normalization increases classification accuracy.
2. Increase in classification accuracy after data normalization is observed as normalizing the data brings the values at the same scale.
3. The highest classification accuracy is obtained with K = 5.
4. Increasing the value of K increases the prediction accuracy.
5. Increasing the value of K increases the data set therefore the prediction accuracy increases.
6. As the classification accuracy increases with the increase in value of K the number of diagonal elements increases.
7. As accuracy increases the number of diagonal elements increases because diagonal elements have true positive and true negative values.
8. With the increase in value of K the number of off-diagonal elements decreases.
9. decrease in off-diagonal elements is observed when the accuracy increases as the value of k increases as it predicts the model correctly.

3

		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
1.	X_Minimum		
2.	X_Maximum	273.418	723.656
3.	Y_Minimum		
4.	Y_Maximum	1583170	1431589
5.	Pixels_Areas	7779.663	585.967
6.	X_Perimeter	393.835	54.491
7.	Y_Perimeter	273.183	45.658
8.	Sum_of_Luminosity	843350.3	62191.13
9.	Minimum_of_Luminosity	53.332	96.236
10.	Maximum_of_Luminosity	135.761	130.451

11.	Length_of_Conveyer	1382.762	1480.018
12.	TypeOfSteel_A300		
13.	TypeOfSteel_A400		
14.	Steel_Plate_Thickness	40.0732	104.214
15.	Edges_Index	0.123	0.385
16.	Empty_Index	0.459	0.426
17.	Square_Index	0.591	0.512
18.	Outside_X_Index	0.107	0.019
19.	Edges_X_Index	0.550	0.608
20.	Edges_Y_Index	0.523	0.831
21.	Outside_Global_Index	0.287	0.608
22.	LogOfAreas	3.622	2.287
23.	Log_X_Index	2.056	1.226
24.	Log_Y_Index	1.848	1.317
25.	Orientation_Index	-0.314	0.136
26.	Luminosity_Index	-0.115	-0.116
27.	SigmoidOfAreas	0.925	0.543

IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – IV

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

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2	**46,733.774**	-60,848,696.530	-320,672.329	-15,750.508	-12,943.764	-32,609,924.843	3,686.073	2,040.905	1,237.644	16.734	25.360	-6.929	4.696	-1.516	16.654	22.505	30.839	-76.320	-47.782	-31.147	27.679	18.083	-30.093
3	-60,848,696.530	**1,821,810,249,729.120**	1,027,980,975.569	83,317,353.381	160,209,448.886	48,997,689,854.156	-5,669,890.139	-6,007,837.239	-7,505,510.376	-114,611.188	-47,711.367	21,948.268	-59,251.278	4,294.736	-19,165.628	-35,306.426	-86,404.069	168,069.821	111,447.699	73,014.357	-82,046.880	-50,711.211	73,811.605
4	-320,672.329	1,027,980,975.569	**104,771,842.621**	6,692,648.900	10,371,695.260	9,008,476,632.104	-154,934.007	6,294.464	10,070.206	547.010	-492.113	585.231	200.195	223.056	-1,121.193	-354.573	556.075	3,456.879	1,427.026	2,840.741	980.333	-300.211	575.040
5	-15,750.508	83,317,353.381	6,692,648.900	**442,770.572**	706,256.501	557,116,030.431	-7,764.045	769.586	771.604	31.924	-24.093	38.161	10.596	10.994	-67.824	-13.284	45.342	183.057	68.412	169.129	72.436	-15.703	28.521
6	-12,943.764	160,209,448.886	10,371,695.260	706,256.501	**1,206,390.510**	807,551,258.148	-6,894.472	1,492.073	-1,364.195	10.207	-17.571	44.182	-16.550	6.496	-65.417	13.411	63.250	176.640	44.055	207.792	105.120	-21.062	19.506
7	-32,609,924.843	48,997,689,854.156	9,008,476,632.104	557,116,030.431	807,551,258.148	**819,345,542,425.229**	-16,498,427.917	777,671.294	2,214,134.327	49,759.906	-53,267.330	58,474.643	44,601.845	25,470.520	-123,180.770	-50,984.933	60,033.134	361,544.755	157,340.839	278,177.342	96,509.492	-22,290.543	62,063.263
8	3,686.073	-5,669,890.139	-154,934.007	-7,764.045	-6,894.472	-16,498,427.917	**1,458.213**	439.236	-153.834	-1.973	3.932	-1.750	1.078	-1.455	3.739	4.623	4.759	-22.187	-12.861	-10.747	3.817	4.448	-6.557
9	2,040.905	-6,007,837.239	6,294.464	769.586	1,492.073	777,671.294	439.236	**333.381**	2.285	-0.791	1.769	-0.222	2.058	-0.353	-0.142	1.575	4.207	-5.859	-4.358	-1.529	4.136	2.716	-2.737
10	1,237.644	-7,505,510.376	10,070.206	771.604	-1,364.195	2,214,134.327	-153.834	2.285	**2,521.557**	-1.821	1.322	0.806	3.926	-0.192	-2.697	-0.534	4.536	2.030	-0.002	2.645	4.370	-0.485	0.211
11	16.734	-114,611.188	547.010	31.924	10.207	49,759.906	-1.973	-0.791	-1.821	**0.730**	-0.009	0.015	-0.015	0.019	0.003	-0.015	-0.021	0.041	0.041	0.019	-0.022	-0.008	0.005
12	25.360	-47,711.367	-492.113	-24.093	-17.571	-53,267.330	3.932	1.769	1.322	-0.009	**0.029**	-0.009	0.007	-0.006	0.015	0.022	0.026	-0.084	-0.054	-0.038	0.024	0.016	-0.028
13	-6.929	21,948.268	585.231	38.161	44.182	58,474.643	-1.750	-0.222	0.806	0.015	-0.009	**0.015**	0.005	0.005	-0.018	-0.012	0.003	0.052	0.030	0.036	0.005	-0.003	0.015
14	4.696	-59,251.278	200.195	10.596	-16.550	44,601.845	1.078	2.058	3.926	-0.015	0.007	0.005	**0.064**	-0.004	-0.036	-0.001	0.070	0.001	-0.020	0.023	0.069	0.016	-0.010
15	-1.516	4,294.736	223.056	10.994	6.496	25,470.520	-1.455	-0.353	-0.192	0.019	-0.006	0.005	-0.004	**0.005**	-0.002	-0.007	-0.010	0.029	0.021	0.014	-0.010	-0.004	0.007
16	16.654	-19,165.628	-1,121.193	-67.824	-65.417	-123,180.770	3.739	-0.142	-2.697	0.003	0.015	-0.018	-0.036	-0.002	**0.057**	0.023	-0.039	-0.098	-0.039	-0.073	-0.045	0.003	-0.026
17	22.505	-35,306.426	-354.573	-13.284	13.411	-50,984.933	4.623	1.575	-0.534	-0.015	0.022	-0.012	-0.001	-0.007	0.023	**0.031**	0.025	-0.099	-0.063	-0.045	0.023	0.014	-0.031
18	30.839	-86,404.069	556.075	45.342	63.250	60,033.134	4.759	4.207	4.536	-0.021	0.026	0.003	0.070	-0.010	-0.039	0.025	**0.203**	-0.058	-0.073	0.019	0.138	0.033	-0.033
19	-76.320	168,069.821	3,456.879	183.057	176.640	361,544.755	-22.187	-5.859	2.030	0.041	-0.084	0.052	0.001	0.029	-0.098	-0.099	-0.058	**0.471**	0.267	0.247	-0.044	-0.067	0.135
20	-47.782	111,447.699	1,427.026	68.412	44.055	157,340.839	-12.861	-4.358	-0.002	0.041	-0.054	0.030	-0.020	0.021	-0.039	-0.063	-0.073	0.267	**0.168**	0.124	-0.066	-0.044	0.082
21	-31.147	73,014.357	2,840.741	169.129	207.792	278,177.342	-10.747	-1.529	2.645	0.019	-0.038	0.036	0.023	0.014	-0.073	-0.045	0.019	0.247	0.124	**0.157**	0.029	-0.025	0.065
22	27.679	-82,046.880	980.333	72.436	105.120	96,509.492	3.817	4.136	4.370	-0.022	0.024	0.005	0.069	-0.010	-0.045	0.023	0.138	-0.044	-0.066	0.029	**0.133**	0.031	-0.028
23	18.083	-50,711.211	-300.211	-15.703	-21.062	-22,290.543	4.448	2.716	-0.485	-0.008	0.016	-0.003	0.016	-0.004	0.003	0.014	0.033	-0.067	-0.044	-0.025	0.031	**0.027**	-0.026
24	-30.093	73,811.605	575.040	28.521	19.506	62,063.263	-6.557	-2.737	0.211	0.005	-0.028	0.015	-0.010	0.007	-0.026	-0.031	-0.033	0.135	0.082	0.065	-0.028	-0.026	**0.049**

Fig - 8 Covariance matrices for class 0

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
2	256,526.309	111,783,525.149	-22,254.624	1,101.079	-1,973.565	-2,334,975.575	-1,224.809	-744.043	13,220.079	-1,932.619	8.914	-3.806	10.893	1.504	6.695	-5.018	-16.564	-13.781	5.306	-21.204	-25.896	-8.452	-14.221
3	111,783,525.149	115,833,894,697.76	322,720,784.181	20,351,188.007	4,659,661.772	32,954,294,851.122	-3,631,824.683	-43,295.897	3,999,505.635	96,154,262.57	23,556.302	-19,250.999	-38,009.673	13,457.300	64,532.972	-22,198.760	-74,705.160	15,298.090	64,300.311	-63,426.815	-119,869.735	-14,717.928	-37,674.924
4	-22,254.624	322,720,784.181	4,714,217.261	178,492.148	129,451.109	488,874,179.455	-15,631.976	-300.304	-23,834.665	4,262.208	-47.646	35.619	-90.634	52.909	-101.643	-96.057	55.178	653.051	330.779	355.115	65.419	-32.384	218.948
5	1,101.079	20,351,188.007	178,492.148	9,807.203	5,546.899	18,662,200.102	-570.116	30.150	-1,446.877	282.113	-1.332	4.156	-7.318	3.972	-4.850	-9.176	-2.152	36.620	23.557	16.864	-3.758	-1.119	15.508
6	-1,973.565	4,659,661.772	129,451.109	5,546.899	5,000.647	13,453,352.777	-557.423	-79.146	-1,139.311	438.560	-2.244	2.952	-6.496	1.204	-8.612	-2.367	7.110	29.028	10.681	21.025	11.045	-1.556	13.014
7	-2,334,975.575	32,954,294,851.122	488,874,179.455	18,662,200.102	13,453,352.777	50,945,346,301.268	-1,463,160.736	84,723.028	-2,735,155.116	343,512.396	-4,688.897	3,985.075	-9,652.577	5,577.969	-10,534.585	-10,271.865	5,462.295	67,782.655	34,740.286	36,734.778	6,364.119	-2,282.381	22,864.848
8	-1,224.809	-3,631,824.683	-15,631.976	-570.116	-557.423	-1,463,160.736	733.909	348.045	-993.311	-204.836	1.066	0.591	0.775	-0.151	0.427	-0.833	-2.224	-5.043	-1.299	-3.287	-2.503	3.684	-1.984
9	-744.043	-43,295.897	-300.304	30.150	-79.146	84,723.028	348.045	406.461	-381.093	-205.394	0.429	-0.025	-0.267	0.044	0.878	-1.090	-2.018	-1.504	0.678	-2.165	-2.874	2.786	-0.960
10	13,220.079	3,999,505.635	-23,834.665	-1,446.877	-1,139.311	-2,735,155.116	-993.311	-381.093	23,100.769	1,243.443	-0.090	-5.160	2.468	-0.698	6.591	1.971	-3.138	-7.953	-1.440	-10.567	-7.431	-4.547	-5.967
11	-1,932.619	-36,154,262.573	4,262.208	282.113	438.560	343,512.396	-204.836	-205.394	1,243.443	5,645.306	-1.331	0.699	-1.134	-0.165	-3.443	2.058	6.623	3.627	-1.376	5.403	7.846	-1.662	2.390
12	8.914	23,556.302	-47.646	-1.332	-2.244	-4,688.897	1.066	0.429	-0.090	-1.331	0.090	-0.001	0.011	0.000	0.008	-0.003	-0.017	-0.012	0.005	-0.017	-0.024	0.005	-0.004
13	-3.806	-19,250.999	35.619	4.156	2.952	3,985.075	0.591	-0.025	-5.160	0.699	-0.001	0.020	-0.002	0.001	-0.012	-0.011	-0.008	0.026	0.022	0.022	-0.004	0.002	0.024
14	10.893	-38,009.673	-90.634	-7.318	-6.496	-9,652.577	0.775	-0.267	2.468	-1.134	0.011	-0.002	0.082	-0.003	0.020	0.015	-0.016	-0.053	-0.021	-0.033	-0.021	0.001	-0.028
15	1.504	13,457.300	52.909	3.972	1.204	5,577.969	-0.151	0.044	-0.698	-0.165	0.000	0.001	-0.003	0.002	0.002	-0.005	-0.005	0.012	0.012	0.001	-0.008	0.000	0.005
16	6.695	64,532.972	-101.643	-4.850	-8.612	-10,534.585	0.427	0.878	6.591	-3.443	0.008	-0.012	0.020	0.002	0.065	-0.014	-0.068	-0.066	0.011	-0.086	-0.103	0.004	-0.045
17	-5.018	-22,198.760	-96.057	-9.176	-2.367	-10,271.865	-0.833	-1.090	1.971	2.058	-0.003	-0.011	0.015	-0.005	-0.014	0.049	0.064	-0.025	-0.058	0.024	0.086	-0.007	-0.017
18	-16.564	-74,705.160	55.178	-2.152	7.110	5,462.295	-2.224	-2.018	-3.138	6.623	-0.017	-0.008	-0.016	-0.005	-0.068	0.064	0.227	0.048	-0.073	0.113	0.229	-0.015	0.022
19	-13.781	15,298.090	653.051	36.620	29.028	67,782.655	-5.043	-1.504	-7.953	3.627	-0.012	0.026	-0.053	0.012	-0.066	-0.025	0.048	0.271	0.116	0.177	0.073	-0.019	0.147
20	5.306	64,300.311	330.779	23.557	10.681	34,740.286	-1.299	0.678	-1.440	-1.376	0.005	0.022	-0.021	0.012	0.011	-0.058	-0.073	0.116	0.119	0.017	-0.101	0.000	0.065
21	-21.204	-63,426.815	355.115	16.864	21.025	36,734.778	-3.287	-2.165	-10.567	5.403	-0.017	0.022	-0.033	0.001	-0.086	0.024	0.113	0.177	0.017	0.178	0.169	-0.017	0.103
22	-25.896	-119,869.735	65.419	-3.758	11.045	6,364.119	-2.503	-2.874	-7.431	7.846	-0.024	-0.004	-0.021	-0.008	-0.103	0.086	0.229	0.073	-0.101	0.169	0.302	-0.019	0.041
23	-8.452	-14,717.928	-32.384	-1.119	-1.556	-2,282.381	3.684	2.786	-4.547	-1.662	0.005	0.002	0.001	0.000	0.004	-0.007	-0.015	-0.019	0.000	-0.017	-0.019	0.025	-0.009
24	-14.221	-37,674.924	218.948	15.508	13.014	22,864.848	-1.984	-0.960	-5.967	2.390	-0.004	0.024	-0.028	0.005	-0.045	-0.017	0.022	0.147	0.065	0.103	0.041	-0.009	0.102

Fig - 9 Covariance matrices for class 1

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

Inferences:

1. The accuracy of Bayes Classifier is 94.362 and bayes focuses on similarities in observations whereas KNN optimises locally therefore normalized KNN classification did better.
2. The data along the diagonal of the covariance matrix represent the variance of the attribute as covariance of itself will give variance of the data.
3. The data along the off diagonal of the covariance matrix represent the covariance of two attributes, showing how attributes vary with respect to each other. Pairs having maximum covariance Y_Maximum and SUM_Of_Luminosity, SUM_Of_Luminosity and Pixel_Areas. Pairs having minimum covariance Square_Index and Edges_Y_Index, Square_Index and LogOfAreas.

4

Table 4 Comparison between classifiers based upon classification accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN	**89.614**
2.	KNN on normalized data	**97.329**
3.	Bayes	**94.362**

Inferences:

1. KNN and normalized KNN classifiers have highest and lowest accuracy respectively.
2. The classifiers in ascending order of classification accuracy.KNN Classifier < Bayes Classifier < Normalized KNN Classifier.
3. KNN has data of different ranges therefore its accuracy is least whereas bayes classification calculates the probability of test data belonging to the class therefore it is more accurate. Normalized KNN has data in the same range and KNN classification is better than Bayes therefore it has the highest accuracy.