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: Madhur Jajoo **Mobile No: 7597389137**

Roll Number: B20211 **Branch: DSE**

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**
5	**89.318**

Inferences:

1. The highest classification accuracy is obtained with K = 3.
2. The best or most accurate value of k is determined by comparing the accuracy of classifier at different values of k we cannot say that accuracy will increase or decrease as increase in value of k. But in our case Increasing the value of K increases the accuracy at first but after that it becomes almost constant.
3. Increase in K implies that we are predicting on basis more data therefore accuracy increases.
4. As accuracy increases the no. of diagonal elements of confusion matrix increases because diagonal elements have value of true negative and true positive.
5. When the model predicts the class correctly the value of diagonal elements increases.
6. When the model predicts the class incorrectly the value of the off-diagonal elements increases.

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. Normalization of data increases the accuracy of the classification.
2. Different attributes have different ranges of values but when we normalize the data we can see the data in a better way we can see the relations better we bring them to a common scale which results in increased accuracy.
3. The highest classification accuracy is obtained with K = 5.
4. Increasing the value of k increases the accuracy.
5. Higher the value of K larger the data on which classifier predict therefore the accuracy increases.
6. With increase in value of K the value of diagonal element increases
7. As accuracy increases the no. of diagonal elements of confusion matrix increases because diagonal elements have value of true negative and true positive
8. When the model predicts the class incorrectly (Low accuracy/ low value of k) the value of the off-diagonal elements increases.

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_Maximum	273.418	723.656
2.	Y_Maximum	1583170	1431589
3.	Pixels_Areas	7779.663	585.967
4.	X_Perimeter	393.835	54.491
5.	Y_Perimeter	273.183	45.658
6.	Sum_of_Luminosity	843350.3	62191.13
7.	Minimum_of_Luminosity	53.326	96.236
8.	Maximum_of_Luminosity	135.762	130.452
9.	Length_of_Conveyer	1382.762	1480.018
10.	Steel_Plate_Thickness	40.073	104.214
11.	Edges_Index	0.123	0.385
12.	Empty_Index	0.459	0.427
13.	Square_Index	0.592	0.513
14.	Outside_X_Index	0.108	0.02
15.	Edges_X_Index	0.55	0.608
16.	Edges_Y_Index	0.523	0.831
17.	Outside_Global_Index	0.288	0.608
18.	LogOfAreas	3.623	2.287
19.	Log_X_Index	2.057	1.227
20.	Log_Y_Index	1.848	1.318
21.	Orientation_Index	-0.314	0.136
22.	Luminosity_Index	-0.115	-0.116
23.	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.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	**46733.774**	-60848696.53	-320672.329	-15750.508	-12943.764	-32609924.84	3686.073	2040.905	1237.644	16.734	25.36	-6.929	4.696	-1.516	16.654	22.505	30.839	-76.32	-47.78	-31.147	27.679	18.083	-30.093
2	-60848696.5	**1.82181E+12**	1027360976	83317353.38	160209448.9	48997689854	-5669890.14	-6007837.2	-7505510.4	-114611.188	-47711.37	21948.268	-59251.28	4294.736	-19165.63	-35306.43	-86404.07	168070	111448	73014.357	-82046.88	-50711.21	73811.605
3	-320672.329	1027360976	**104771843**	6632648.9	10371695.26	9008476632	-154934.007	6294.464	10070.206	547.01	-492.113	585.231	200.195	223.056	-1121.193	-354.573	556.075	3456.9	1427	2840.741	980.333	-300.211	575.04
4	-15750.508	83317353.38	6632648.9	**442770.57**	706256.501	557116030.4	-7764.045	769.586	771.604	31.924	-24.093	38.161	10.596	10.994	-67.824	-13.284	45.342	183.06	68.412	169.129	72.436	-15.703	28.521
5	-12943.764	160209448.9	10371695.26	706256.501	**1206390.5**	807551258.1	-6894.472	1492.073	-1364.195	10.207	-17.571	44.182	-16.55	6.496	-65.417	13.411	63.25	176.64	44.055	207.732	105.12	-21.062	19.506
6	-32609924.8	48997689854	9008476632	557116030.4	807551258.1	**8.19346E+11**	-16498427.9	777671.294	2214134.33	49759.906	-53267.33	58474.643	44601.845	25470.52	-123180.8	-50984.93	60033.134	361545	157341	278177.34	96509.492	-22290.54	62063.263
7	3686.073	-5669890.139	-154934.007	-7764.045	-6894.472	-16498427.92	**1458.213**	439.236	-153.834	-1.973	3.932	-1.75	1.078	-1.455	3.739	4.623	4.759	-22.19	-12.86	-10.747	3.817	4.448	-6.557
8	2040.905	-6007837.239	6294.464	769.586	1492.073	777671.294	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
9	1237.644	-7505510.376	10070.206	771.604	-1364.195	2214134.327	-153.834	2.285	**2521.557**	-1.821	1.322	0.806	3.926	-0.192	-2.697	-0.534	4.536	2.03	-0.002	2.645	4.37	-0.485	0.211
10	16.734	-114611.188	547.01	31.924	10.207	49759.906	-1.973	-0.791	-1.821	**0.73**	-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
11	25.36	-47711.367	-492.113	-24.093	-17.571	-53267.33	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
12	-6.929	21948.268	585.231	38.161	44.182	58474.643	-1.75	-0.222	0.806	0.015	-0.009	**0.015**	0.005	0.005	-0.018	-0.012	0.003	0.052	0.03	0.036	0.005	-0.003	0.015
13	4.696	-59251.278	200.195	10.596	-16.55	44601.845	1.078	2.058	3.926	-0.015	0.007	0.005	**0.064**	-0.004	-0.036	-0.001	0.07	0.001	-0.02	0.023	0.069	0.016	-0.01
14	-1.516	4294.736	223.056	10.994	6.496	25470.52	-1.455	-0.353	-0.192	0.019	-0.006	0.005	-0.004	**0.005**	-0.002	-0.007	-0.01	0.029	0.021	0.014	-0.01	-0.004	0.007
15	16.654	-19165.628	-1121.193	-67.824	-65.417	-123180.77	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
16	22.505	-35306.426	-354.573	-13.284	13.411	-50984.933	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
17	30.839	-86404.069	556.075	45.342	63.25	60033.134	4.759	4.207	4.536	-0.021	0.026	0.003	0.07	-0.01	-0.039	0.025	**0.203**	-0.058	-0.073	0.019	0.138	0.033	-0.033
18	-76.32	168069.821	3456.879	183.057	176.64	361544.755	-22.187	-5.859	2.03	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
19	-47.782	111447.639	1427.026	68.412	44.055	157340.839	-12.861	-4.358	-0.002	0.041	-0.054	0.03	-0.02	0.021	-0.039	-0.063	-0.073	0.267	**0.168**	0.124	-0.066	-0.044	0.082
20	-31.147	73014.357	2840.741	169.129	207.732	278177.342	-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
21	27.679	-82046.88	980.333	72.436	105.12	96509.492	3.817	4.136	4.37	-0.022	0.024	0.005	0.069	-0.01	-0.045	0.023	0.138	-0.044	-0.066	0.029	**0.133**	0.031	-0.028
22	18.083	-50711.211	-300.211	-15.703	-21.062	-22290.543	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
23	-30.093	73811.605	575.04	28.521	19.506	62063.263	-6.557	-2.737	0.211	0.005	-0.028	0.015	-0.01	0.007	-0.026	-0.031	-0.033	0.135	0.082	0.065	-0.028	-0.026	**0.049**

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
1	**256526.31**	111783525.1	-22254.624	1101.079	-1973.565	-2334975.58	-1224.809	-744.043	13220.08	-1932.619	8.914	-3.806	10.893	1.504	6.695	-5.018	-16.564	-13.78	5.306	-21.204	-25.896	-8.452	-14.221
2	111783525	**3.11583E+12**	322720784	20351188	4659661.8	32954294851	-3631824.7	-43295.9	3999506	-36154262.6	23556.3	-19250.999	-38009.7	13457.3	64533	-22198.8	-74705.2	15298	64300	-63426.8	-119869.7	-14717.9	-37674.9
3	-22254.624	322720784.2	**4714217.26**	178492.15	129451.11	488874179.5	-15631.976	-300.304	-23834.67	4262.208	-47.646	35.619	-90.634	52.909	-101.643	-96.057	55.178	653.1	330.8	355.115	65.419	-32.384	218.948
4	1101.079	20351188.01	178492.148	**9807.203**	5546.899	18662200.1	-570.116	30.15	-1446.877	282.113	-1.332	4.156	-7.318	3.972	-4.85	-9.176	-2.152	36.62	23.56	16.864	-3.758	-1.119	15.508
5	-1973.565	4659661.772	129451.109	5546.899	**5000.647**	13453352.78	-557.423	-79.146	-1139.311	438.56	-2.244	2.952	-6.496	1.204	-8.612	-2.367	7.11	29.03	10.68	21.025	11.045	-1.556	13.014
6	-2334975.6	32954294851	488874179	18662200	13453353	**50945346301**	-1463160.7	84723.03	-2735155	343512.396	-4688.9	3985.075	-9652.58	5577.969	-10534.6	-10271.9	5462.3	67783	34740	36734.8	6364.119	-2282.38	22864.8
7	-1224.809	-3631824.683	-15631.976	-570.116	-557.423	-1463160.74	**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
8	-744.043	-43295.897	-300.304	30.15	-79.146	84723.028	348.045	**406.461**	-381.093	-205.394	0.429	-0.025	-0.267	0.044	0.878	-1.09	-2.018	-1.504	0.678	-2.165	-2.874	2.786	-0.96
9	13220.079	3999505.635	-23834.665	-1446.877	-1139.311	-2735155.12	-993.311	-381.093	**23100.77**	1243.443	-0.09	-5.16	2.468	-0.698	6.591	1.971	-3.138	-7.953	-1.44	-10.567	-7.431	-4.547	-5.967
10	-1932.619	-36154262.57	4262.208	282.113	438.56	343512.396	-204.836	-205.394	1243.443	**5645.306**	-1.331	0.699	-1.134	-0.165	-3.443	2.058	6.623	3.627	-1.376	5.403	7.846	-1.662	2.39
11	8.914	23556.302	-47.646	-1.332	-2.244	-4688.897	1.066	0.429	-0.09	-1.331	**0.09**	-0.001	0.011	0	0.008	-0.003	-0.017	-0.012	0.005	-0.017	-0.024	0.005	-0.004
12	-3.806	-19250.999	35.619	4.156	2.952	3985.075	0.591	-0.025	-5.16	0.699	-0.001	**0.02**	-0.002	0.001	-0.012	-0.011	-0.008	0.026	0.022	0.022	-0.004	0.002	0.024
13	10.893	-38009.673	-90.634	-7.318	-6.496	-9652.577	0.775	-0.267	2.468	-1.134	0.011	-0.002	**0.082**	-0.003	0.02	0.015	-0.016	-0.053	-0.021	-0.033	-0.021	0.001	-0.028
14	1.504	13457.3	52.909	3.972	1.204	5577.969	-0.151	0.044	-0.698	-0.165	0	0.001	-0.003	**0.002**	0.002	-0.005	-0.005	0.012	0.012	0.001	-0.008	0	0.005
15	6.695	64532.972	-101.643	-4.85	-8.612	-10534.585	0.427	0.878	6.591	-3.443	0.008	-0.012	0.02	0.002	**0.065**	-0.014	-0.068	-0.066	0.011	-0.086	-0.103	0.004	-0.045
16	-5.018	-22198.76	-96.057	-9.176	-2.367	-10271.865	-0.833	-1.09	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
17	-16.564	-74705.16	55.178	-2.152	7.11	5462.295	-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
18	-13.781	15298.09	653.051	36.62	29.028	67782.655	-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
19	5.306	64300.311	330.779	23.557	10.681	34740.286	-1.299	0.678	-1.44	-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	0.065
20	-21.204	-63426.815	355.115	16.864	21.025	36734.778	-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
21	-25.896	-119869.735	65.419	-3.758	11.045	6364.119	-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
22	-8.452	-14717.928	-32.384	-1.119	-1.556	-2282.381	3.684	2.786	-4.547	-1.662	0.005	0.002	0.001	0	0.004	-0.007	-0.015	-0.019	0	-0.017	-0.019	**0.025**	-0.009
23	-14.221	-37674.924	218.948	15.508	13.014	22864.848	-1.984	-0.96	-5.967	2.39	-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**

Figure 9: Covariance matrix for class 1

Inferences:

1. The accuracy of bayes classifier is 94.362 which is more than the accuracy of KNN but less than the accuracy of the normalized KNN this is because Bayes directly finds the similarity between observations whereas KNN-normalized does batter because of its nature to optimize locally.
2. The data along the diagonal element represent the variance of that particular attribute.
3. 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 on normalized data have the highest accuracy whereas KNN on non-normalized data is the least accurate.
2. KNN < Bayes < KNN on normalized data.
3. Normal KNN has lower accuracy as it works on data in which the attributes have different ranges of values. Bayes Classifier calculates the probability of a test data belonging to a particular class therefor it is more accurate. Normalized KNN has data in which attributes are scaled to a common range therefor it gives more accuracy.