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: ISHAAN GUPTA **Mobile No: 9179242114**

Roll Number: B20292 **Branch: MECHANICAL ENGINEERING**

1 a.

		Prediction Outcome
True Label	81	27
	27	201

Figure 1 KNN Confusion Matrix for K = 1

		Prediction Outcome
True Label	83	25
	12	216

Figure 2 KNN Confusion Matrix for K = 3

		Prediction Outcome
True Label	82	26
	9	219

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	**83.928**
3	**88.988**
5	**89.583**

Inferences:

1. The highest classification accuracy is obtained with K=5.
2. Increasing the value of K increases the prediction accuracy.
3. Increasing the value of K increases the prediction accuracy as if there will be more nearest neighbors then there will be more accuracy.
4. As the classification accuracy increases with the increase in value of K, the number of diagonal elements increase.
5. Increase in diagonal elements with increase in k is because more accuracy leads to more number of true values.
6. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements decrease.
7. Decrease in off-diagonal elements with increase in k is because more accuracy leads to less number of false values.

2 a.

		Prediction Outcome
True Label	100	8
	8	220

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

		Prediction Outcome
True Label	100	8
	7	221

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

		Prediction Outcome
True Label	101	7
	4	224

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	**95.238**
3	**95.535**
5	**96.726**

Inferences:

1. Data normalization increases classification accuracy.
2. Data normalization increases classification accuracy because bias is present in data which gets removed by normalization and range of different features or attributes become even.
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 prediction accuracy as if there will be more nearest neighbors then there will be more accuracy.
6. As the classification accuracy increases with the increase in value of K, the number of diagonal elements increase.
7. Increase in diagonal elements with increase in k is because more accuracy leads to more number of true values.
8. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements decrease.
9. Decrease in off-diagonal elements with increase in k is because more accuracy leads to less number of false values.

3

		Prediction Outcome
True Label	96	12
	2	226

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 95.833 %.

Table 3 Mean for class 0 and class 1

S. No.	Attribute Name	Mean	
		Class 0	Class 1
1.	X_Maximum	286.3322	746.584
2.	Y_Maximum	1711478	1445964
3.	Pixels_Areas	7268.032	583.512
4.	X_Perimeter	355.6148	52.184
5.	Y_Perimeter	207.1555	43.112
6.	Sum_of_Luminosity	808615.7	61552.41
7.	Minimum_of_Luminosity	53.40283	94.804
8.	Maximum_of_Luminosity	135.8587	130.184
9.	Length_of_Conveyer	1382.516	1486.63
10.	Steel_Plate_Thickness	40.24735	100.434
11.	Edges_Index	0.126447	0.388864
12.	Empty_Index	0.449608	0.418643
13.	Square_Index	0.593253	0.510322
14.	Outside_X_Index	0.108173	0.019854
15.	Edges_X_Index	0.565851	0.625601
16.	Edges_Y_Index	0.524692	0.837443
17.	Outside_Global_Index	0.268551	0.611
18.	LogOfAreas	3.599567	2.264311
19.	Log_X_Index	2.048011	1.214075
20.	Log_Y_Index	1.825003	1.299494
21.	Orientation_Index	-0.32807	0.131946
22.	Luminosity_Index	-0.10907	-0.12263
23.	SigmoidOfAreas	0.91587	0.527024

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	57593.63	-8.7E+07	-349304	-15539.5	-8064.2	-3.8E+07	4246.334	2211.845	2606.644	204.7402	26.17266	-9.75451	7.642827	-2.23	20.45434	28.00765	34.63566	-87.7296	-55.9715	-35.5257	32.66325	19.371	-33.4599
2	-8.7E+07	2.63E+12	-7.5E+08	-3.9E+07	-3E+07	-9.8E+10	-4665084	-8053965	-1.1E+07	-325686	-55558.1	14531.11	-93632.7	3191.985	6707.551	-38623.7	-133538	183163.6	137803.1	46364.04	-141236	-57051.4	95439.94
3	-349304	-7.5E+08	28362934	1395371	857469.9	3.37E+09	-130039	-4383.89	30347.18	-158.483	-476.937	368.7523	529.9781	228.2041	-931.5	-654.24	290.1563	2816.525	1451.628	1686.892	371.9959	-158.522	605.0511
4	-15539.5	-3.9E+07	1395371	74685.83	45819.84	1.67E+08	-6114.62	45.13686	2140.324	1.372203	-22.569	22.28837	32.94668	11.61067	-52.1513	-33.6519	22.92827	135.711	69.40658	86.51528	26.97984	-5.82787	28.83417
5	-8064.2	-3E+07	857469.9	45819.84	28599.26	1.03E+08	-3579.29	186.0079	1535.583	-4.61306	-12.425	13.37403	22.38895	6.618315	-32.587	-19.547	19.01129	79.72314	39.188	52.72424	20.9153	-2.33701	16.39452
6	-3.8E+07	-9.8E+10	3.37E+09	1.67E+08	1.03E+08	4.03E+11	-1.5E+07	10270.49	3727268	-38802	-53411.3	43540.79	69465.53	26038.01	-112302	-74739.6	44593.87	321540.3	162501.5	197432.1	54471.51	-14263.4	67039.14
7	4246.334	-4665084	-130039	-6114.62	-3579.29	-1.5E+07	1435.624	454.1635	-143.801	-2.68864	4.151371	-2.06021	1.110997	-1.50741	4.217815	4.825914	3.304557	-23.0601	-13.287	-11.3109	2.997309	4.691634	-7.15036
8	2211.845	-8053965	-4383.89	45.13686	186.0079	10270.49	454.1635	359.4764	-7.73533	-7.26988	1.958658	-0.34975	2.293221	-0.35618	-0.05244	1.563509	3.839511	-6.09022	-4.44702	-1.78535	3.952615	2.95132	-2.91046
9	2606.644	-1.1E+07	30347.18	2140.324	1535.583	3727268	-143.801	-7.73533	2489.102	40.58116	1.088053	0.403797	3.902723	-0.29132	-2.61843	0.068471	4.977984	1.110105	-0.94312	2.477846	5.153582	-0.47664	0.079518
10	204.7402	-325686	-158.483	1.372203	-4.61306	-38802	-2.68864	-7.26988	40.58116	6.67619	-0.02288	-0.01833	-0.00033	0.007042	0.015516	0.04225	0.075182	-0.05118	-0.04349	-0.01177	0.063571	-0.0548	0.016415
11	26.17266	-55558.1	-476.937	-22.569	-12.425	-53411.3	4.151371	1.958658	1.088053	-0.02288	0.031376	-0.0107	0.008443	-0.00652	0.016943	0.024762	0.025106	-0.08947	-0.05723	-0.04014	0.02475	0.017144	-0.03031
12	-9.75451	14531.11	368.7523	22.28837	13.37403	43540.79	-2.06021	-0.34975	0.403797	-0.01833	-0.0107	0.015879	0.003162	0.005884	-0.01716	-0.0149	-0.00155	0.055166	0.035188	0.034454	-0.00062	-0.00447	0.016978
13	7.642827	-93632.7	529.9781	32.94668	22.38895	69465.53	1.110997	2.293221	3.902723	-0.00033	0.008443	0.003162	0.064938	-0.00461	-0.03679	0.001585	0.070142	-0.00203	-0.02424	0.024275	0.072524	0.016203	-0.01346
14	-2.23	3191.985	228.2041	11.61067	6.618315	26038.01	-1.50741	-0.35618	-0.29132	0.007042	-0.00652	0.005884	-0.00461	0.005192	-0.00269	-0.00789	-0.00877	0.031563	0.022657	0.015489	-0.00932	-0.00391	0.008422
15	20.45434	6707.551	-931.5	-52.1513	-32.587	-112302	4.217815	-0.05244	-2.61843	0.015516	0.016943	-0.01716	-0.03679	-0.00269	0.057628	0.026556	-0.03545	-0.10388	-0.04368	-0.07203	-0.04027	0.003847	-0.02687
16	28.00765	-38623.7	-654.24	-33.6519	-19.547	-74739.6	4.825914	1.563509	0.068471	0.04225	0.024762	-0.0149	0.001585	-0.00789	0.026556	0.032364	0.021446	-0.10811	-0.06675	-0.05279	0.020181	0.015404	-0.0335
17	34.63566	-133538	290.1563	22.92827	19.01129	44593.87	3.304557	3.839511	4.977984	0.075182	0.025106	-0.00155	0.070142	-0.00877	-0.03545	0.021446	0.193582	-0.04818	-0.06551	0.016604	0.127894	0.028636	-0.02973
18	-87.7296	183163.6	2816.525	135.711	79.72314	321540.3	-23.0601	-6.09022	1.110105	-0.05118	-0.08947	0.055166	-0.00203	0.031563	-0.10388	-0.10811	-0.04818	0.497087	0.28442	0.253712	-0.04511	-0.06685	0.147085
19	-55.9715	137803.1	1451.628	69.40658	39.188	162501.5	-13.287	-4.44702	-0.94312	-0.04349	-0.05723	0.035188	-0.02424	0.022657	-0.04368	-0.06675	-0.06551	0.28442	0.178677	0.134332	-0.06428	-0.04457	0.088635
20	-35.5257	46364.04	1686.892	86.51528	52.72424	197432.1	-11.3109	-1.78535	2.477846	-0.01177	-0.04014	0.034454	0.024275	0.015489	-0.07203	-0.05279	0.016604	0.253712	0.134332	0.146629	0.018411	-0.02479	0.070343
21	32.66325	-141236	371.9959	26.97984	20.9153	54471.51	2.997309	3.952615	5.153582	0.063571	0.02475	-0.00062	0.072524	-0.00932	-0.04027	0.020181	0.127894	-0.04511	-0.06428	0.018411	0.122956	0.029404	-0.02825
22	19.371	-57051.4	-158.522	-5.82787	-2.33701	-14263.4	4.691634	2.95132	-0.47664	-0.0548	0.017144	-0.00447	0.016203	-0.00391	0.003847	0.015404	0.028636	-0.06685	-0.04457	-0.02479	0.029404	0.025836	-0.02768
23	-33.4599	95439.94	605.0511	28.83417	16.39452	67039.14	-7.15036	-2.91046	0.079518	0.016415	-0.03031	0.016978	-0.01346	0.008422	-0.02687	-0.0335	-0.02973	0.147085	0.088635	0.070343	-0.02825	-0.02768	0.053956

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	258038	1.48E+08	-19263.8	261.3593	-1901.45	-2032754	-1183.65	-1180.02	12247.36	-2832.32	3.389629	-2.46295	11.63035	1.207025	8.391756	-4.13275	-10.084	-15.4643	1.16597	-18.5757	-23.3566	-10.1654	-14.8971
2	1.48E+08	3.3E+12	5.07E+08	29140900	9302068	5.3E+10	3586455	600089.5	1305441	-3.4E+07	36534.77	-16500.8	-26651.4	18243.76	54437.89	-29077.1	-74054.5	74366.27	89905.85	-28476.2	-116645	-13913.8	-2798.65
3	-19263.8	5.07E+08	5121724	201881.2	135506.6	5.33E+08	-15218.1	2762.653	-29026.5	2315.248	-37.4526	31.83585	-107.658	69.85872	-87.9306	-125.617	30.57832	692.874	377.5828	342.99	17.39376	-31.1396	225.095
4	261.3593	29140900	201881.2	10847.83	5755.106	21161000	-541.645	203.95	-2125.82	185.3148	-0.37207	3.604206	-7.99808	4.80709	-4.17462	-10.055	-3.31706	37.99874	24.90501	16.14324	-5.72256	-1.01823	15.11528
5	-1901.45	9302068	135506.6	5755.106	5008.472	14025224	-538.583	-23.2411	-1229.69	313.8431	-1.34556	2.599497	-6.41278	1.403408	-8.18555	-2.71112	6.355279	28.17901	10.66834	19.74748	9.913426	-1.49543	12.26622
6	-2032754	5.3E+10	5.33E+08	21161000	14025224	5.56E+10	1443015	397726.8	3291478	147379.8	-3554.68	3415.541	-11365.9	7414.94	-8940.02	-13523.9	2549.558	71815.02	39675.69	35077.39	888.0399	-2320.96	23386.72
7	-1183.65	-3586455	-15218.1	-541.645	-538.583	-1443015	775.0757	358.481	-1115.3	-263.239	1.258554	0.764607	0.299348	-0.15777	0.237079	-1.2047	-2.83291	-4.85539	-1.1122	-3.18433	-2.80378	3.944022	-1.90637
8	-1180.02	600089.5	2762.653	203.95	-23.2411	397726.8	358.481	454.2025	-543.781	-252.573	0.648911	-0.03367	-0.62677	0.158121	0.834313	-1.42068	-2.36215	-0.87899	1.218323	-2.11138	-3.40665	2.914106	-0.70993
9	12247.36	-1305441	-29026.5	-2125.82	-1229.69	-3291478	-1115.3	-543.781	24015.18	1507.221	-0.81282	-4.70557	5.134416	-1.03976	7.174554	3.787893	0.736543	-10.2283	-4.39458	-9.4854	-4.35897	-5.69474	-7.36396
10	-2832.32	-3.4E+07	2315.248	185.3148	313.8431	147379.8	-263.239	-252.573	1507.221	4839.485	-1.68357	0.499004	-1.08981	-0.12363	-2.54892	2.177936	5.52187	2.463037	-1.66097	4.562957	7.272558	-2.02021	1.797769
11	3.389629	36534.77	-37.4526	-0.37207	-1.34556	-3554.68	1.258554	0.648911	-0.81282	-1.68357	0.091525	-0.00062	0.00719	0.000291	0.005738	-0.00563	-0.01802	-0.00683	0.00622	-0.01327	-0.02537	0.006027	-0.00073
12	-2.46295	-16500.8	31.83585	3.604206	2.599497	3415.541	0.764607	-0.03367	-4.70557	0.499004	-0.00062	0.019259	-0.0041	0.001013	-0.01277	-0.01049	-0.00788	0.021506	0.018034	0.018909	-0.00275	0.002688	0.021312
13	11.63035	-26651.4	-107.658	-7.99808	-6.41278	-11365.9	0.299348	-0.62677	5.134416	-1.08981	0.00719	-0.0041	0.079217	-0.00354	0.021606	0.015921	-0.01164	-0.05042	-0.02188	-0.0313	-0.01668	-0.00112	-0.0267
14	1.207025	18243.76	69.85872	4.80709	1.403408	7414.94	-0.15777	0.158121	-1.03976	-0.12363	0.000291	0.001013	-0.00354	0.003074	0.001958	-0.00599	-0.00579	0.013667	0.012988	0.001498	-0.00946	-0.00021	0.004932
15	8.391756	54437.89	-87.9306	-4.17462	-8.18555	-8940.02	0.237079	0.834313	7.174554	-2.54892	0.005738	-0.01277	0.021606	0.001958	0.064792	-0.01361	-0.06601	-0.06026	0.012815	-0.08285	-0.10144	0.003667	-0.04222
16	-4.13275	-29077.1	-125.617	-10.055	-2.71112	-13523.9	-1.2047	-1.42068	3.787893	2.177936	-0.00563	-0.01049	0.015921	-0.00599	-0.01361	0.048401	0.0655	-0.02696	-0.05737	0.021954	0.086042	-0.00926	-0.01539
17	-10.084	-74054.5	30.57832	-3.31706	6.355279	2549.558	-2.83291	-2.36215	0.736543	5.52187	-0.01802	-0.00788	-0.01164	-0.00579	-0.06601	0.0655	0.226632	0.042312	-0.07504	0.110481	0.230654	-0.01825	0.020579
18	-15.4643	74366.27	692.874	37.99874	28.17901	71815.02	-4.85539	-0.87899	-10.2283	2.463037	-0.00683	0.021506	-0.05042	0.013667	-0.06026	-0.02696	0.042312	0.26084	0.114388	0.16375	0.06232	-0.01878	0.140272
19	1.16597	89905.85	377.5828	24.90501	10.66834	39675.69	-1.1122	1.218323	-4.39458	-1.66097	0.00622	0.018034	-0.02188	0.012988	0.012815	-0.05737	-0.07504	0.114388	0.116207	0.013321	-0.10368	0.000739	0.060811
20	-18.5757	-28476.2	342.99	16.14324	19.74748	35077.39	-3.18433	-2.11138	-9.4854	4.562957	-0.01327	0.018909	-0.0313	0.001498	-0.08285	0.021954	0.110481	0.16375	0.013321	0.16536	0.162916	-0.01752	0.09646
21	-23.3566	-116645	17.39376	-5.72256	9.913426	888.0399	-2.80378	-3.40665	-4.35897	7.272558	-0.02537	-0.00275	-0.01668	-0.00946	-0.10144	0.086042	0.230654	0.06232	-0.10368	0.162916	0.302038	-0.0214	0.038612
22	-10.1654	-13913.8	-31.1396	-1.01823	-1.49543	-2320.96	3.944022	2.914106	-5.69474	-2.02021	0.006027	0.002688	-0.00112	-0.00021	0.003667	-0.00926	-0.01825	-0.01878	0.000739	-0.01752	-0.0214	0.026011	-0.0085
23	-14.8971	-2798.65	225.095	15.11528	12.26622	23386.72	-1.90637	-0.70993	-7.36396	1.797769	-0.00073	0.021312	-0.0267	0.004932	-0.04222	-0.01539	0.020579	0.140272	0.060811	0.09646	0.038612	-0.0085	0.09779

Figure 9: Covariance matrix for class 1

Inferences:

1. The accuracy of Bayes Classifier is 95.833 %. 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.

4

Table 4 Comparison between classifiers based upon classification accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN	89.583
2.	KNN on normalized data	96.726
3.	Bayes	95.833

Inferences:

1. The classifier with highest accuracy is KNN on normalized data and lowest accuracy is KNN.
2. The classifiers in ascending order of classification accuracy= KNN < Bayes < KNN on normalized data .