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: Aryan Tiwari **Mobile No: 8982562898**

Roll Number: B20187 **Branch: Electrical 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

IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – IV

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

		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	**83.92**
3	**88.98**
5	**89.58**

Inferences:

1. The highest classification accuracy is obtained with K =5
2. The increasing value of K increases the prediction accuracy.
3. More the number of nearest neighbors, better the analysis hence accuracy increases.
4. The diagonal elements (true values) increase with increasing k.
5. As the accuracy increases, so does the number of true predictions .
6. The off-diagonal elements decrease with higher k values.
7. The increase in accuracy with k decreases the number of false predictions.

IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – IV

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

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	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	**95.23**
3	**95.53**
5	**96.72**

Inferences:

1. Normalisation increases accuracy.
2. Normalising data eliminates bias due to uneven range of different attributes.
3. The highest classification accuracy is obtained with K =5.
4. More the number of nearest neighbors, better the analysis hence accuracy increases.
5. The diagonal elements (true values) increase with increasing k.
6. As the accuracy increases, so does the number of true predictions .
7. The off-diagonal elements decrease with higher k values.
8. The increase in accuracy with k decreases the number of false predictions.

3

	Prediction Outcome	
True	96	12

	2	226

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 95.83%.

Table 3 Mean for class 0 and class 1

S. No.	Attribute Name	Mean	
		Class 0	Class 1
1.	X_Maximum	286.3321554770318	746.584
2.	Y_Maximum	1711478.0565371024	1445963.75
3.	Pixels_Areas	7268.031802120141	583.512
4.	X_Perimeter	355.6148409893993	52.184
5.	Y_Perimeter	207.15547703180212	43.112
6.	Sum_of_Luminosity	808615.6925795053	61552.412
7.	Minimum_of_Luminosity	53.40282685512368	94.804
8.	Maximum_of_Luminosity	135.85865724381625	130.184
9.	Length_of_Conveyer	1382.5159010600707	1486.63
10.	Steel_Plate_Thickness	40.24734982332156	100.434
11.	Edges_Index	0.12644699646643126	0.3888644
12.	Empty_Index	0.44960777385159006	0.4186427999999997
13.	Square_Index	0.5932530035335692	0.5103224
14.	Outside_X_Index	0.10817279151943467	0.019853799999999998
15.	Edges_X_Index	0.5658508833922259	0.6256006000000003
16.	Edges_Y_Index	0.5246918727915195	0.8374429999999995
17.	Outside_Global_Index	0.26855123674911663	0.611
18.	LogOfAreas	3.599567137809189	2.2643111999999976
19.	Log_X_Index	2.048011307420494	1.214075400000001
20.	Log_Y_Index	1.825002826855123	1.2994936
21.	Orientation_Index	-0.3280713780918728	0.131946
22.	Luminosity_Index	-0.10907385159010606	-0.12263200000000007
23.	SigmoidOfAreas	0.9158699646643109	0.5270244

IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – IV

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

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.

X_Maximum	Y_Maximum	Pixels_Areas	X_Perimeter	Y_Perimeter	Sum_of_Lum	Minimum_of	Maximum_o	Length_of_C	Steel_Plate_	Edges_Index	Empty_Index	Square_Inde	Outside_X_I	Edges_X_Ind	Edges_Y_Ind	Outside_Glo	LogOfAreas	Log_X_Index	Log_Y_Index	Orientation_	Luminosity_I	SigmoidOfAr
57593.634	-86700603	-349303.67	-15539.489	-8064.1972	-38068098	4246.33381	2211.84499	2606.64364	204.740245	26.1726577	-9.7545065	7.64282701	-2.2299973	20.4543405	28.0076481	34.6356602	-87.729562	-55.971493	-35.525749	32.6632543	19.3710044	-33.459881
-86700603	2.6343E+12	-754411187	-38584596	-29519236	-9.799E+10	-4665083.7	-8053964.8	-10797723	-325686.4	-55558.111	14531.1097	-93632.707	3191.98492	6707.55061	-38623.707	-133538.2	183163.591	137803.088	46364.0404	-141235.62	-57051.381	95439.9397
-349303.67	-754411187	28362934.2	1395370.66	857469.928	3373269717	-130039.3	-4383.8856	30347.1821	-158.48307	-476.93692	368.752327	529.978051	228.204148	-931.49974	-654.23987	290.156323	2816.52507	1451.62832	1686.8919	371.995881	-158.52226	605.051105
-15539.489	-38584596	1395370.66	74685.8334	45819.8402	166734113	-6114.6244	45.1368569	2140.32352	1.37220259	-22.568996	22.2883679	32.9466804	11.6106668	-52.151251	-33.651946	22.9282698	135.711025	69.4065827	86.5152844	26.9798362	-5.8278697	28.8341696
-8064.1972	-29519236	857469.928	45819.8402	28599.2594	103157362	-3579.2863	186.007869	1535.58263	-4.6130617	-12.424968	13.3740311	22.3889453	6.61831524	-32.587	-19.54701	19.0112899	79.7231402	39.1879972	52.7242425	20.915304	-2.3370094	16.3945224
-38068098	-9.799E+10	3373269717	166734113	103157362	4.0348E+11	-14678207	10270.4919	3727267.82	-38801.98	-53411.29	43540.7935	69465.529	26038.0089	-112302.15	-74739.591	44593.8719	321540.321	162501.462	197432.138	54471.5075	-14263.373	67039.1397
4246.33381	-4665083.7	-130039.3	-6114.6244	-3579.2863	-14678207	1435.62439	454.163522	-143.80075	-2.688645	4.15137072	-2.0602085	1.11099701	-1.5074099	4.21781454	4.82591392	3.3045573	-23.060082	-13.287033	-11.310914	2.99730935	4.69163446	-7.1503634
2211.84499	-8053964.8	-4383.8856	45.1368569	186.007869	10270.4919	454.163522	359.476405	-7.7353332	-7.2698795	1.95865773	-0.3497528	2.29322064	-0.3561766	-0.052436	1.56350878	3.83951081	-6.0902196	-4.447025	-1.7853542	3.9526147	2.95131967	-2.9104642
2606.64364	-10797723	30347.1821	2140.32352	1535.58263	3727267.82	-143.80075	-7.7353332	2489.1017	40.5811593	1.08805262	0.40379668	3.90272256	-0.2913207	-2.6184306	0.06847087	4.97798411	1.11010496	-0.9431204	2.47784605	5.15358234	-0.4766416	0.07951839
204.740245	-325686.4	-158.48307	1.37220259	-4.6130617	-38801.98	-2.688645	-7.2698795	40.5811593	6.67618976	-0.022884	-0.0183317	-0.0003323	0.00704221	0.01551574	0.04225024	0.07518232	-0.0511798	-0.0434922	-0.0117702	0.06357091	-0.0547973	0.01641526
26.1726577	-55558.111	-476.93692	-22.568996	-12.424968	-53411.29	4.15137072	1.95865773	1.08805262	-0.022884	0.03137594	-0.0106964	0.00844312	-0.0065201	0.01694349	0.02476217	0.02510613	-0.0894748	-0.0572308	-0.0401421	0.02475006	0.01714353	-0.0303127
-9.7545065	14531.1097	368.752327	22.2883679	13.3740311	43540.7935	-2.0602085	-0.3497528	0.40379668	-0.0183317	-0.0106964	0.01587908	0.0031618	0.00588363	-0.0171605	-0.0149049	-0.0015512	0.05516606	0.03518784	0.03445395	-0.000615	-0.0044721	0.01697751
7.64282701	-93632.707	529.978051	32.9466804	22.3889453	69465.529	1.11099701	2.29322064	3.90272256	-0.0003323	0.00844312	0.0031618	0.06493792	-0.0046098	-0.0367916	0.00158466	0.0701421	-0.0020323	-0.0242445	0.02427526	0.07252356	0.01620267	-0.0134611
-2.2299973	3191.98492	228.204148	11.6106668	6.61831524	26038.0089	-1.5074099	-0.3561766	-0.2913207	0.00704221	-0.0065201	0.00588363	-0.0046098	0.0051917	-0.0026865	-0.0078948	-0.0087671	0.03156268	0.02265726	0.01548856	-0.0093191	-0.0039082	0.00842183
20.4543405	6707.55061	-931.49974	-52.151251	-32.587	-112302.15	4.21781454	-0.052436	-2.6184306	0.01551574	0.01694349	-0.0171605	-0.0367916	-0.0026865	0.05762837	0.02655619	-0.035449	-0.1038782	-0.0436831	-0.072027	-0.0402731	0.00384707	-0.0268723
28.0076481	-38623.707	-654.23987	-33.651946	-19.54701	-74739.591	4.82591392	1.56350878	0.06847087	0.04225024	0.02476217	-0.0149049	0.00158466	-0.0078948	0.02655619	0.03236387	0.02144634	-0.108107	-0.0667494	-0.0527923	0.02018127	0.01540362	-0.0334963
34.6356602	-133538.2	290.156323	22.9282698	19.0112899	44593.8719	3.3045573	3.83951081	4.97798411	0.07518232	0.02510613	-0.0015512	0.0701421	-0.0087671	-0.035449	0.02144634	0.19358194	-0.0481752	-0.0655075	0.01660367	0.12789406	0.02863604	-0.0297291
-87.729562	183163.591	2816.52507	135.711025	79.7231402	321540.321	-23.060082	-6.0902196	1.11010496	-0.0511798	-0.0894748	0.05516606	-0.0020323	0.03156268	-0.1038782	-0.108107	-0.0481752	0.49708655	0.2844205	0.25371164	-0.0451129	-0.0668461	0.14708501
-55.971493	137803.088	1451.62832	69.4065827	39.1879972	162501.462	-13.287033	-4.447025	-0.9431204	-0.0434922	-0.0572308	0.03518784	-0.0242445	0.02265726	-0.0436831	-0.0667494	-0.0655075	0.2844205	0.17867695	0.1343316	-0.0642812	-0.0445669	0.08863525
-35.525749	46364.0404	1686.8919	86.5152844	52.7242425	197432.138	-11.310914	-1.7853542	2.47784605	-0.0117702	-0.0401421	0.03445395	0.02427526	0.01548856	-0.072027	-0.0527923	0.01660367	0.25371164	0.1343316	0.14662905	0.01841134	-0.0247914	0.07034317
32.6632543	-141235.62	371.995881	26.9798362	20.915304	54471.5075	2.99730935	3.9526147	5.15358234	0.06357091	0.02475006	-0.000615	0.07252356	-0.0093191	-0.0402731	0.02018127	0.12789406	-0.0451129	-0.0642812	0.01841134	0.12295571	0.02940416	-0.0282501
19.3710044	-57051.381	-158.52226	-5.8278697	-2.3370094	-14263.373	4.69163446	2.95131967	-0.4766416	-0.0547973	0.01714353	-0.0044721	0.01620267	-0.0039082	0.00384707	0.01540362	0.02863604	-0.0668461	-0.0445669	-0.0247914	0.02940416	0.02583603	-0.0276808
-33.459881	95439.9397	605.051105	28.8341696	16.3945224	67039.1397	-7.1503634	-2.9104642	0.07951839	0.01641526	-0.0303127	0.01697751	-0.0134611	0.00842183	-0.0268723	-0.0334963	-0.0297291	0.14708501	0.08863525	0.07034317	-0.0282501	-0.0276808	0.05395551

For better view open 'c0.csv' in zip

Figure 8: Covariance matrix for class 0

IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – IV

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

X_Maximum	Y_Maximum	Pixels_Areas	X_Perimeter	Y_Perimeter	Sum_of_Lum	Minimum_of	Maximum_o	Length_of_O	Steel_Plate_	Edges_Index	Empty_Index	Square_Inde	Outside_X_In	Edges_X_Ind	Edges_Y_Ind	Outside_Glo	LogOfAreas	Log_X_Index	Log_Y_Index	Orientation_	Luminosity_I	SigmoidOfAr
258038.019	148247097	-19263.781	261.359263	-1901.4483	-2032754.1	-1183.6488	-1180.0175	12247.3588	-2832.3161	3.38962885	-2.4629513	11.6303512	1.20702483	8.39175576	-4.1327484	-10.083992	-15.464289	1.16597031	-18.575736	-23.356581	-10.16544	-14.897115
148247097	3.2968E+12	506836443	29140899.8	9302067.87	5.296E+10	-3586455.2	600089.513	-1305440.9	-34372354	36534.7689	-16500.843	-26651.402	18243.7641	54437.8942	-29077.053	-74054.472	74366.2716	89905.8487	-28476.239	-116644.65	-13913.783	-2798.6542
-19263.781	506836443	5121723.53	201881.166	135506.594	532831804	-15218.092	2762.6531	-29026.474	2315.24829	-37.452553	31.8358468	-107.65827	69.8587161	-87.930551	-125.61698	30.5783246	692.874031	377.582817	342.989992	17.3937558	-31.139554	225.094957
261.359263	29140899.8	201881.166	10847.8258	5755.1056	21160999.6	-541.64523	203.950044	-2125.8236	185.314774	-0.3720684	3.60420554	-7.9980781	4.80708968	-4.1746181	-10.055025	-3.3170581	37.998743	24.9050097	16.1432399	-5.7225648	-1.0182258	15.1152837
-1901.4483	9302067.87	135506.594	5755.1056	5008.4724	14025223.9	-538.58321	-23.24109	-1229.6879	313.843078	-1.3455593	2.599497	-6.4127763	1.40340759	-8.1855518	-2.7111152	6.35527856	28.17901	10.6683422	19.7474841	9.9134259	-1.4954339	12.2662191
-2032754.1	5.296E+10	532831804	21160999.6	14025223.9	5.5619E+10	-1443015.2	397726.778	-3291478.1	147379.793	-3554.681	3415.54063	-11365.915	7414.94048	-8940.0169	-13523.878	2549.55838	71815.0233	39675.6919	35077.3895	888.039931	-2320.9565	23386.7242
-1183.6488	-3586455.2	-15218.092	-541.64523	-538.58321	-1443015.2	775.075735	358.481026	-1115.2991	-263.23941	1.25855413	0.7646066	0.29934789	-0.1577692	0.23707928	-1.2046988	-2.8329098	-4.855393	-1.1122046	-3.1843305	-2.8037773	3.94402157	-1.9063732
-1180.0175	600089.513	2762.6531	203.950044	-23.24109	397726.778	358.481026	454.202549	-543.78148	-252.573	0.64891137	-0.0336688	-0.6267745	0.15812134	0.83431312	-1.4206789	-2.3621483	-0.8789942	1.21832317	-2.1113846	-3.4066518	2.9141063	-0.7099344
12247.3588	-1305440.9	-29026.474	-2125.8236	-1229.6879	-3291478.1	-1115.2991	-543.78148	24015.1835	1507.22102	-0.812817	-4.7055675	5.13441572	-1.0397644	7.17455393	3.7878935	0.73654309	-10.228312	-4.3945841	-9.485397	-4.3589683	-5.6947445	-7.3639611
-2832.3161	-34372354	2315.24829	185.314774	313.843078	147379.793	-263.23941	-252.573	1507.22102	4839.48461	-1.6835653	0.49900403	-1.0898059	-0.1236302	-2.5489209	2.17793581	5.52186974	2.46303721	-1.6609719	4.56295689	7.27255755	-2.0202143	1.79776855
3.38962885	36534.7689	-37.452553	-0.3720684	-1.3455593	-3554.681	1.25855413	0.64891137	-0.812817	-1.6835653	0.09152501	-0.0006192	0.00718962	0.0029095	0.00573767	-0.005633	-0.0180243	-0.0068285	0.0062196	-0.013269	-0.0253706	0.00602743	-0.0007343
-2.4629513	-16500.843	31.8358468	3.60420554	2.599497	3415.54063	0.7646066	-0.0336688	-4.7055675	0.49900403	-0.0006192	0.01925909	-0.0041018	0.00101333	-0.0127655	-0.010491	-0.007875	0.02150552	0.01803361	0.0189089	-0.0027524	0.0026876	0.0213124
11.6303512	-26651.402	-107.65827	-7.9980781	-6.4127763	-11365.915	0.29934789	-0.6267745	5.13441572	-1.0898059	0.00718962	-0.0041018	0.07921745	-0.0035408	0.02160555	0.01592075	-0.0116421	-0.0504217	-0.0218775	-0.031296	-0.0166797	-0.0011201	-0.0267006
1.20702483	18243.7641	69.8587161	4.80708968	1.40340759	7414.94048	-0.1577692	0.15812134	-1.0397644	-0.1236302	0.00029095	0.00101333	-0.0035408	0.00307401	0.00195839	-0.0059908	-0.005786	0.01366705	0.01298821	0.00149777	-0.0094552	-0.0002132	0.00493196
8.39175576	54437.8942	-87.930551	-4.1746181	-8.1855518	-8940.0169	0.23707928	0.83431312	7.17455393	-2.5489209	0.00573767	-0.0127655	0.02160555	0.00195839	0.06479208	-0.0136109	-0.0660066	-0.0602551	0.01281467	-0.0828514	-0.1014352	0.00366659	-0.0422166
-4.1327484	-29077.053	-125.61698	-10.055025	-2.7111152	-13523.878	-1.2046988	-1.4206789	3.7878935	2.17793581	-0.005633	-0.010491	0.01592075	-0.0059908	-0.0136109	0.04840095	0.06550003	-0.0269553	-0.0573723	0.02195446	0.08604196	-0.0092573	-0.0153896
-10.083992	-74054.472	30.5783246	-3.3170581	6.35527856	2549.55838	-2.8329098	-2.3621483	0.73654309	5.52186974	-0.0180243	-0.007875	-0.0116421	-0.005786	-0.0660066	0.06550003	0.22663226	0.04231248	-0.0750401	0.11048117	0.2306543	-0.0182531	0.02057885
-15.464289	74366.2716	692.874031	37.998743	28.17901	71815.0233	-4.855393	-0.8789942	-10.228312	2.46303721	-0.0068285	0.02150552	-0.0504217	0.01366705	-0.0602551	-0.0269553	0.04231248	0.26084036	0.1143878	0.16374976	0.06231963	-0.0187757	0.14027205
1.16597031	89905.8487	377.582817	24.9050097	10.6683422	39675.6919	-1.1122046	1.21832317	-4.3945841	-1.6609719	0.0062196	0.01803361	-0.0218775	0.01298821	0.01281467	-0.0573723	-0.0750401	0.1143878	0.11620742	0.01332104	-0.103683	0.00073869	0.06081141
-18.575736	-28476.239	342.989992	16.1432399	19.7474841	35077.3895	-3.1843305	-2.1113846	-9.485397	4.56295689	-0.013269	0.0189089	-0.031296	0.00149777	-0.0828514	0.02195446	0.11048117	0.16374976	0.01332104	0.16536017	0.16291606	-0.0175153	0.09645976
-23.356581	-116644.65	17.3937558	-5.7225648	9.9134259	888.039931	-2.8037773	-3.4066518	-4.3589683	7.27255755	-0.0253706	-0.0027524	-0.0166797	-0.0094552	-0.1014352	0.08604196	0.2306543	0.06231963	-0.103683	0.16291606	0.30203792	-0.0213964	0.03861205
-10.16544	-13913.783	-31.139554	-1.0182258	-1.4954339	-2320.9565	3.94402157	2.9141063	-5.6947445	-2.0202143	0.00602743	0.0026876	-0.0011201	-0.0002132	0.00366659	-0.0092573	-0.0182531	-0.0187757	0.00073869	-0.0175153	-0.0213964	0.02601085	-0.0085037
-14.897115	-2798.6542	225.094957	15.1152837	12.2662191	23386.7242	-1.9063732	-0.7099344	-7.3639611	1.79776855	-0.0007343	0.0213124	-0.0267006	0.00493196	-0.0422166	-0.0153896	0.02057885	0.14027205	0.06081141	0.09645976	0.03861205	-0.0085037	0.09779023

For better view open 'c1.csv' in zip

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.

IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – IV

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

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	**89.58**
2.	KNN on normalized data	**96.72**
3.	Bayes	**95.83%.**

Inferences:

1. KNN on normalized data has highest accuracy while KNN has lowest accuracy.