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: Bharat Kumar

Roll Number: B20090

Assignment-4

Course- IC272

Mobile No: 6299862503

Branch: CSE

1   a.

	Prediction Outcome	
	93	25
	19	200

(True Label on vertical axis)

Figure 1 KNN Confusion Matrix for K = 1

	Prediction Outcome	
	92	26
	9	210

(True Label on vertical axis)

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	86.94
3	89.61
4	89.32

Inferences:

1. The highest classification accuracy is obtained with K =.3
2. Infer whether increasing the value of K increases the prediction accuracy to some value after that it decreases.
3. Since we select class based on the distances that we get from k datapoints. So, if k is more then it will give more accurate result.
4. No. of diagonal elements increases with increase in accuracy.
5. No. of off-diagonal elements decreases with increase in accuracy.
6. Off-diagonal elements decreases because they are wrongly classified datapoints.

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	96.14
3	97.03
5	97.33

Inferences:

1. Data normalization increases classification accuracy.
2. The reason for increase in classification accuracy after data normalization is that now the distance calculated is more realistic.
3. The highest classification accuracy is obtained with K =.5.
4. Increasing the value of K increases the prediction accuracy.
5. Since we select class based on the distances that we get from k datapoints. So, if k is more then it will give more accurate result.
6. No. of diagonal elements increases with increase in accuracy.
7. No. of off-diagonal elements decreases with increase in accuracy.
8. Off-diagonal elements decreases because they are wrongly classified datapoints.

3

	Prediction Outcome	
True Label	101	7
	12	216

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is   94.35 %.

	0	1
X_Maximum	0.164714	0.434183
Y_Maximum	0.147408	0.124446
Pixels_Areas	0.047598	0.003809
X_Perimeter	0.033848	0.004804
Y_Perimeter	0.011358	0.002320
Sum_of_Luminosity	0.069740	0.005289
Minimum_of_Luminosity	0.272463	0.483694
Maximum_of_Luminosity	0.457679	0.431407
Length_of_Conveyer	0.129623	0.402173
Steel_Plate_Thickness	0.000951	0.232438
Edges_Index	0.127057	0.390740
Empty_Index	0.476330	0.443525
Square_Index	0.589849	0.506224
Outside_X_Index	0.121995	0.020990
Edges_X_Index	0.559508	0.620130
Edges_Y_Index	0.500517	0.829175
Outside_Global_Index	0.268551	0.611000
LogOfAreas	0.675562	0.402095
Log_X_Index	0.629985	0.329262
Log_Y_Index	0.428535	0.305139
Orientation_Index	0.334356	0.566372
Luminosity_Index	0.559709	0.551181
SigmoidOfAreas	0.904506	0.463138

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

In Fig. 8 and 9 representing covariance matrices for class 1 and class 0 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	0.088452	0.007506	-7.39E-05	1.46E-05	-6.13E-05	-0.0001	-0.00354	-0.0032	0.018771	-0.00638	0.001994	-0.00153	0.006866	0.000808	0.004985	-0.00254	-0.0059	-0.00185	0.000246	-0.00255	-0.0069	-0.00374	-0.0099
2	0.007506	0.024659	0.000287	0.000241	4.43E-05	0.000395	-0.00158	0.00024	-0.0003	-0.01143	0.003175	-0.00151	-0.00232	0.001804	0.004777	-0.00264	-0.0064	0.001317	0.002804	-0.00058	-0.00509	-0.00076	-0.00027
3	-7.39E-05	0.000287	0.00022	0.000127	4.89E-05	0.000301	-0.00051	8.38E-05	-0.0005	5.83E-05	-0.00025	0.000221	-0.00071	0.000523	-0.00058	-0.00086	0.0002	0.00093	0.000892	0.000528	5.75E-05	-0.00013	0.001674
4	1.46E-05	0.000241	0.000127	9.94E-05	3.04E-05	0.000175	-0.00026	9.04E-05	-0.00053	6.82E-05	-3.58E-05	0.000366	-0.00077	0.000526	-0.00041	-0.00101	-0.00032	0.000745	0.000086	0.000363	-0.00028	-6.13E-05	0.001642
5	-6.13E-05	4.43E-05	4.89E-05	3.04E-05	1.52E-05	6.67E-05	-0.00015	-5.93E-06	-0.00018	6.65E-05	-7.45E-05	0.000152	-0.00036	8.84E-05	-0.00046	-0.00016	0.00035	0.000318	0.000212	0.000363	0.000275	-5.18E-05	0.000767
6	-0.0001	0.000395	0.000301	0.000175	6.67E-05	0.000414	-0.00064	0.000159	-0.00074	4.89E-05	-0.00031	0.000312	-0.00099	0.000732	-0.00078	-0.00123	0.00022	0.001269	0.001234	0.000711	3.86E-05	-0.00013	0.00229
7	-0.00354	-0.00158	-0.00051	-0.00026	-0.00015	-0.00064	0.020176	0.008468	-0.0149	-0.00517	0.006452	0.004133	0.00154	-0.00092	0.001227	-0.00646	-0.01445	-0.00507	-0.00205	-0.00381	-0.00721	0.012657	-0.01104
8	-0.0032	0.00024	8.38E-05	9.04E-05	-5.93E-06	0.000159	0.008468	0.009735	-0.00659	-0.0045	0.003019	-0.00017	-0.00293	0.000837	0.003919	-0.00691	-0.01094	-0.00083	0.002034	-0.0023	-0.00795	0.008486	-0.00373
9	0.018771	-0.0003	-0.0005	-0.00053	-0.00018	-0.00074	-0.0149	-0.00659	0.164573	0.015175	-0.00214	-0.01305	0.013553	-0.00311	0.019056	0.01042	0.001928	-0.00548	-0.00415	-0.00583	-0.00576	-0.00938	-0.02188
10	-0.00638	-0.01143	5.83E-05	6.82E-05	6.65E-05	4.89E-05	-0.00517	-0.0045	0.015175	0.07159	-0.00651	0.002033	-0.00423	-0.00054	-0.00995	0.008803	0.021238	0.00194	-0.0023	0.004121	0.014108	-0.00489	0.007848
11	0.001994	0.003175	-0.00025	-3.58E-05	-7.45E-05	-0.00031	0.006452	0.003019	-0.00214	-0.00651	0.09241	-0.00066	0.007285	0.000334	0.00585	-0.00595	-0.01811	-0.00141	0.002254	-0.00313	-0.01286	0.00381	-0.00084
12	-0.00153	-0.00151	0.000221	0.000366	0.000152	0.000312	0.004133	-0.00017	-0.01305	0.002033	-0.00066	0.021616	-0.00438	0.001228	-0.01372	-0.01168	-0.00834	0.004666	0.00689	0.004704	-0.00147	0.001791	0.025629
13	0.006866	-0.00232	-0.00071	-0.00077	-0.00036	-0.00099	0.00154	-0.00293	0.013553	-0.00423	0.007285	-0.00438	0.080549	-0.00408	0.022105	0.016871	-0.01174	-0.01041	-0.00796	-0.00741	-0.00848	-0.00071	-0.03056
14	0.000808	0.001804	0.000523	0.000526	8.84E-05	0.000732	-0.00092	0.000837	-0.00311	-0.00054	0.000334	0.001228	-0.00408	0.004021	0.002272	-0.0072	-0.00662	0.003201	0.005356	0.000402	-0.00545	-0.00015	0.006402
15	0.004985	0.004777	-0.00058	-0.00041	-0.00046	-0.00078	0.001227	0.003919	0.019056	-0.00995	0.00585	-0.01372	0.022105	0.002272	0.066699	-0.01451	-0.06697	-0.01252	0.004689	-0.01974	-0.05191	0.00234	-0.04862
16	-0.00254	-0.00264	-0.00086	-0.00101	-0.00016	-0.00123	-0.00646	-0.00691	0.01042	0.008803	-0.00595	-0.01168	0.016871	-0.0072	-0.01451	0.05345	0.068831	-0.0058	-0.02174	0.005417	0.045604	-0.00612	-0.01836
17	-0.0059	-0.0064	0.0002	-0.00032	0.00035	0.0022	-0.01445	-0.01094	0.001928	0.021238	-0.01811	-0.00834	-0.01174	-0.00662	-0.06697	0.068831	0.226632	0.008666	-0.02706	0.025942	0.116333	-0.01148	0.023359
18	-0.00185	0.001317	0.00093	0.000745	0.000318	0.001269	-0.00507	-0.00083	-0.00548	0.00194	-0.00141	0.004666	-0.01041	0.003201	-0.01252	-0.0058	0.008666	0.010941	0.008448	0.007875	0.006437	-0.00242	0.032609
19	0.000246	0.002804	0.000892	0.00086	0.000212	0.001234	-0.00205	0.002034	-0.00415	-0.0023	0.002254	0.00689	-0.00796	0.005356	0.004689	-0.02174	-0.02706	0.008448	0.015111	0.001128	-0.01886	0.000168	0.024891
20	-0.00255	-0.00058	0.000528	0.000363	0.000275	0.000711	-0.00381	-0.0023	-0.00583	0.004121	-0.00313	0.004704	-0.00741	0.000402	-0.01974	0.005417	0.025942	0.007875	0.001128	0.009118	0.019294	-0.00259	0.025709
21	-0.0069	-0.00509	5.75E-05	-0.00028	0.000275	3.86E-05	-0.00721	-0.00795	-0.00576	0.014108	-0.01286	-0.00147	-0.00848	-0.00545	-0.05191	0.045604	0.116333	0.006437	-0.01886	0.019294	0.076833	-0.00679	0.022105
22	-0.00374	-0.00076	-0.00013	-6.13E-05	-5.18E-05	-0.00013	0.012657	0.008486	-0.00938	-0.00489	0.00381	0.001791	-0.00071	-0.00015	0.00234	-0.00612	-0.01148	-0.00242	0.000168	-0.00259	-0.00679	0.010291	-0.00607
23	-0.0099	-0.00027	0.001674	0.001642	0.000767	0.00229	-0.01104	-0.00373	-0.02188	0.007848	-0.00084	0.025629	-0.03056	0.006402	-0.04862	-0.01836	0.023359	0.032609	0.024891	0.025709	0.022105	-0.00607	0.125992

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0.01974	-0.00439	-0.00134	-0.00087	-0.00026	-0.00192	0.01268	0.006	0.004	0.00046	0.0154	-0.00605	0.00451	-0.00149	0.01215	0.01723	0.02028	-0.01052	-0.01182	-0.00488	0.00965	0.00713	-0.02224
2	-0.00439	0.0197	-0.00043	-0.00032	-0.00014	-0.00073	-0.00206	-0.00322	-0.00244	-0.00011	-0.00483	0.00133	-0.00817	0.00032	0.00059	-0.00351	-0.01155	0.00324	0.0043	0.00094	-0.00616	-0.0031	0.00937
3	-0.00134	-0.00043	0.00122	0.00087	0.00031	0.00191	-0.00435	-0.00013	0.0052	-3.99E-06	-0.00314	0.00256	0.0035	0.00171	-0.00619	-0.0045	0.0019	0.00378	0.00343	0.00259	0.00123	-0.00065	0.0045
4	-0.00087	-0.00032	0.00087	0.00068	0.00024	0.00138	-0.00299	2.00E-05	0.00054	5.05E-07	-0.00217	0.00226	0.00318	0.00127	-0.00506	-0.00339	0.00219	0.00266	0.0024	0.00194	0.0013	-0.00035	0.00313
5	-0.00026	-0.00014	0.00031	0.00024	8.68E-05	0.00049	-0.00101	4.74E-05	0.00022	-9.77E-06	-0.00069	0.00078	0.00124	0.00042	-0.00182	-0.00113	0.00105	0.0009	0.00078	0.00068	0.00058	-8.10E-05	0.00103
6	-0.00192	-0.00073	0.00191	0.00138	0.00049	0.003	-0.00646	4.10E-06	0.00084	-1.29E-05	-0.00463	0.00398	0.00604	0.00257	-0.00983	-0.00678	0.00385	0.00568	0.00506	0.004	0.00237	-0.00077	0.00656
7	0.01268	-0.00206	-0.00435	-0.00299	-0.00101	-0.00646	0.03737	0.01073	-0.00192	-5.28E-05	0.02128	-0.01114	0.00572	-0.0088	0.02183	0.02587	0.01686	-0.0241	-0.02445	-0.01355	0.00771	0.01506	-0.04141
8	0.006	-0.00322	-0.00013	2.00E-05	4.74E-05	4.10E-06	0.01073	0.0077	-9.37E-05	-0.00013	0.00911	-0.00172	0.01071	-0.00189	-0.00025	0.00761	0.01778	-0.00577	-0.00742	-0.00194	0.00923	0.00859	-0.01529
9	0.004	-0.00244	0.00052	0.00054	0.00022	0.00084	-0.00192	-9.37E-05	0.01706	0.00041	0.00286	0.00112	0.0103	-0.00087	-0.00695	0.00019	0.01303	0.0006	-0.00089	0.00152	0.0068	-0.00078	0.00024
10	0.00046	-0.00011	-3.99E-06	5.05E-07	-9.77E-07	-1.29E-05	-5.28E-05	-0.00013	0.00041	9.88E-05	-8.84E-05	-7.47E-05	-1.29E-06	3.10E-05	6.05E-05	0.00017	0.00029	-4.03E-05	-6.03E-05	-1.06E-05	0.00012	-0.00013	7.17E-05
11	0.0154	-0.00483	-0.00314	-0.00217	-0.00069	-0.00463	0.02128	0.00911	0.00286	-8.84E-05	0.03168	-0.01139	0.00855	-0.00749	0.01727	0.02615	0.02523	-0.01841	-0.02074	-0.00947	0.01254	0.01084	-0.03457
12	-0.00605	0.00133	0.00256	0.00226	0.00078	0.00398	-0.01114	-0.00172	0.00112	-7.47E-05	-0.01139	0.01782	0.00338	0.00713	-0.01845	-0.01659	-0.00164	0.01197	0.01344	0.00857	-0.00033	-0.00298	0.02042
13	0.00451	-0.00817	0.0035	0.00318	0.00124	0.00604	0.00572	0.01071	0.0103	-1.29E-06	0.00855	0.00338	0.06603	-0.00532	-0.03764	0.00168	0.07073	-0.00042	-0.00882	0.00575	0.03688	0.01028	-0.01541
14	-0.00149	0.00032	0.00171	0.00127	0.00042	0.00257	-0.0088	-0.00189	-0.00087	3.10E-05	-0.00749	0.00713	-0.00532	0.00679	-0.00312	-0.00949	-0.01003	0.00739	0.00934	0.00416	-0.00538	-0.00281	0.01093
15	0.01215	0.00059	-0.00619	-0.00506	-0.00182	-0.00983	0.02183	-0.00025	-0.00695	6.05E-05	0.01727	-0.01845	-0.03764	-0.00312	0.05932	0.02831	-0.03597	-0.02159	-0.01598	-0.01716	-0.02061	0.00246	-0.03095
16	0.01723	-0.00351	-0.0045	-0.00339	-0.00113	-0.00678	0.02587	0.00761	0.00019	0.00017	0.02615	-0.01659	0.00168	-0.00949	0.02831	0.03574	0.02254	-0.02327	-0.02529	-0.01303	0.0107	0.01018	-0.03995
17	0.02028	-0.01155	0.0019	0.00219	0.00105	0.00385	0.01686	0.01778	0.01303	0.00029	0.02523	-0.00164	0.07073	-0.01003	-0.03597	0.02254	0.19358	-0.00987	-0.02362	0.0039	0.0645	0.01801	-0.03374
18	-0.01052	0.00324	0.00378	0.00266	0.0009	0.00568	-0.0241	-0.00577	0.0006	-4.03E-05	-0.01841	0.01197	-0.00042	0.00739	-0.02159	-0.02327	-0.00987	0.02085	0.02101	0.0122	-0.00466	-0.00861	0.03419
19	-0.01182	0.0043	0.00343	0.0024	0.00078	0.00506	-0.02445	-0.00742	-0.00089	-6.03E-05	-0.02074	0.01344	-0.00882	0.00934	-0.01598	-0.02529	-0.02362	0.02101	0.02323	0.01137	-0.01169	-0.01011	0.03628
20	-0.00488	0.00094	0.00259	0.00194	0.00068	0.004	-0.01355	-0.00194	0.00152	-1.06E-05	-0.00947	0.00857	0.00575	0.00416	-0.01716	-0.01303	0.0039	0.0122	0.01137	0.00808	0.00218	-0.00366	0.01875
21	0.00965	-0.00616	0.00123	0.0013	0.00058	0.00237	0.00771	0.00923	0.0068	0.00012	0.01254	-0.00033	0.03688	-0.00538	-0.02061	0.0107	0.0645	-0.00466	-0.01169	0.00218	0.03128	0.00933	-0.01617
22	0.00713	-0.0031	-0.00065	-0.00035	-8.10E-05	-0.00077	0.01506	0.00859	-0.00078	-0.00013	0.01084	-0.00298	0.01028	-0.00281	0.00246	0.01018	0.01801	-0.00861	-0.01011	-0.00366	0.00933	0.01022	-0.01976
23	-0.02224	0.00937	0.0045	0.00313	0.00103	0.00656	-0.04141	-0.01529	0.00024	7.17E-05	-0.03457	0.02042	-0.01541	0.01093	-0.03095	-0.03995	-0.03374	0.03419	0.03628	0.01875	-0.01617	-0.01976	0.06952

Inferences:

1. The accuracy of Bayes Classifier is 94.35 % and it is lesser than the previous classifier because it is linear but not the previous one.
2. Diagonal elements are variance of different attributes which signifies the spread of that attribute.
3. Off-diagonal elements are pairwise covariance between different attributes.
4. The off-diagonal elements are symmetric about diagonal.
5. Minimum covariance for class 0 (Y_Maximum, Sum of Luminocsity), Maximum is for (X_Maximum, Y_Maximum).
6. Minimum covariance for class 1 (Steel_plate_thickness, Y_Maximum), Maximum is for (Y_Maximum, Sum of Luminocsity).

Table 4 Comparison between classifiers based upon classification accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN(max)	89.61
2.	KNN on normalized data(max)	97.33
3.	Bayes	94.35

Inferences:

1. The classifier with highest and lowest accuracy are KNN on normalized data with k=5 and KNN on normal data with k=1.
2. Arrange the classifiers in ascending order of classification accuracy. Classifier 1 < Classifier 3 < Classifier 2 .
3. The reason behind Inference 1 and 2 is normalization of data.
4. Even Baye's Classifier on normal data gives very less accuracy.
5. In general Bayes classifier is less accurate compared to KNN because it is a linear classifier.