



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: Mayank Bansal

Mobile No: +919636993445

Roll Number: B20156

Branch:CSE

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	83.9
3	88.9
5	89.5

#### 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 more no training samples are involved.
4. Diagonal elements increase with rise in accuracy.
5. Diagonal elements correspond to correct predictions. Hence by increase in accuracy diagonal elements also increase.
6. As the classification accuracy increases with the increase in value of K the number of off-diagonal elements decrease.
7. Total no of test elements is same. Diagonal elements correspond to correct predictions. Hence by increase in accuracy non-diagonal elements decrease.

2

	Prediction Outcome	
True Label	108	0
	0	228

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

	Prediction Outcome	
True Label	108	0
	0	228

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

	Prediction Outcome	
True Label	108	0
	0	228

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	100
3	100
5	100

#### Inferences:

1. Data normalization increases accuracy .
2. Accuracy increases because after normalization attributes with bigger range come in same range with that of the other attributes. Euclidean distances can't have different ranges for different attributes.
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 more no training samples are involved.
6. Diagonal elements increase with rise in accuracy.
7. Diagonal elements correspond to correct predictions. Hence by increase in accuracy diagonal elements also increase.
8. As the classification accuracy increases with the increase in value of K the number of off-diagonal elements decrease.
9. Total no of test elements is same. Diagonal elements correspond to correct predictions. Hence by increase in accuracy non-diagonal elements decrease.

3

	Prediction Outcome	
True Label	100	8
	10	218

## IC 272: DATA SCIENCE - III

### LAB ASSIGNMENT – IV

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

**Figure 7 Confusion Matrix obtained from Bayes Classifier**

The classification accuracy obtained from Bayes Classifier is 94.6%.

	A	B
1	Means in Class1	0
2	X_Minimum	718.104
3	X_Maximum	746.584
4	Y_Minimum	1445930.352
5	Y_Maximum	1445963.75
6	Pixels_Areas	583.512
7	X_Perimeter	52.184
8	Y_Perimeter	43.112
9	Sum_of_Luminosity	61552.412
10	Minimum_of_Luminosity	94.804
11	Maximum_of_Luminosity	130.184
12	Length_of_Conveyer	1486.63
13	Steel_Plate_Thickness	100.434
14	Edges_Index	0.3888644
15	Empty_Index	0.4186428
16	Square_Index	0.5103224
17	Outside_X_Index	0.0198538
18	Edges_X_Index	0.6256006
19	Edges_Y_Index	0.837443
20	Outside_Global_Index	0.611
21	LogOfAreas	2.2643112
22	Log_X_Index	1.2140754
23	Log_Y_Index	1.2994936
24	Orientation_Index	0.131946
25	Luminosity_Index	-0.122632
26	SigmoidOfAreas	0.5270244

	A	B
1	Means in Class 0	0
2	X_Minimum	137.0742
3	X_Maximum	286.3322
4	Y_Minimum	1711389
5	Y_Maximum	1711478
6	Pixels_Areas	7268.032
7	X_Perimeter	355.6148
8	Y_Perimeter	207.1555
9	Sum_of_Luminosity	808615.7
10	Minimum_of_Luminosity	53.40283
11	Maximum_of_Luminosity	135.8587
12	Length_of_Conveyer	1382.516
13	Steel_Plate_Thickness	40.24735
14	Edges_Index	0.126447
15	Empty_Index	0.449608
16	Square_Index	0.593253
17	Outside_X_Index	0.108173
18	Edges_X_Index	0.565851
19	Edges_Y_Index	0.524692
20	Outside_Global_Index	0.268551
21	LogOfAreas	3.599567
22	Log_X_Index	2.048011
23	Log_Y_Index	1.825003
24	Orientation_Index	-0.32807
25	Luminosity_Index	-0.10907
26	SigmoidOfAreas	0.91587

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.



## IC 272: DATA SCIENCE - III LAB ASSIGNMENT – IV

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

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0	73410.76	60549.17	-90087769	-90093815	-670150	-31930.69	-17452.27	-74756359	6342.261	2694.95	2828.869	195.088	35.08571	-17.97702	13.56639	-9.395472	24.49329	38.95188	46.25837	-131.5098	-87.22761	-57.24686	44.99933	24.74246	-45.08064
1	60549.17	57593.63	-86697461	-86700603	-349303.7	-15539.49	-8064.197	-38068098	4246.334	2211.845	2606.644	204.7402	26.17266	-9.754506	7.642827	-2.229997	20.45434	28.00765	34.63566	-87.72956	-55.97149	-35.52575	32.66325	19.371	-33.45988
2	-90087769	-86697461	2.63E+12	2.63E+12	-7.55E+08	-38599976	-29528899	-9.8E+10	-4663742	-8054009	-10798321	-325682.7	-55553.89	14526.26	-93640.68	3189.947	6719.661	-38617.3	-133546.7	183134	137789.1	46344.19	-141244.4	-57050.28	95433.37
3	-90093815	-86700603	2.63E+12	2.63E+12	-7.54E+08	-38584596	-29519236	-9.8E+10	-4665084	-8053965	-10797723	-325686.4	-55558.11	14531.11	-93632.71	3191.985	6707.551	-38623.71	-133538.2	183163.6	137803.1	46364.04	-141235.6	-57051.38	95439.94
4	-670150	-349303.7	-7.55E+08	-7.54E+08	28362934	1395371	857469.9	3.37E+09	-130039.3	-4383.886	30347.18	-158.4831	-476.9369	368.7523	529.9781	228.2041	-931.4997	-654.2399	290.1563	2816.525	1451.628	1686.892	371.9959	-158.5223	605.0511
5	-31930.69	-15539.49	-38599976	-38584596	1395371	74685.83	45819.84	1.67E+08	-6114.624	45.13686	2140.324	1.372203	-22.569	22.28837	32.94668	11.61067	-52.15125	-33.65195	22.92827	135.711	69.40658	86.51528	26.97984	-5.82787	28.83417
6	-17452.27	-8064.197	-29528899	-29519236	857469.9	45819.84	28599.26	1.03E+08	-3579.286	186.0079	1535.583	-4.613062	-12.42497	13.37403	22.38895	6.618315	-32.587	-19.54701	19.01129	79.72314	39.188	52.72424	20.9153	-2.337009	16.39452
7	-74756359	-38068098	-9.8E+10	-9.8E+10	3.37E+09	1.67E+08	1.03E+08	4.03E+11	-14678207	10270.49	3727268	-38801.98	-53411.29	43540.79	69465.53	26038.01	-112302.2	-74739.59	44593.87	321540.3	162501.5	197432.1	54471.51	-14263.37	67039.14
8	6342.261	4246.334	-4663742	-4665084	-130039.3	-6114.624	-3579.286	-14678207	1435.624	454.1635	-143.8008	-2.688645	4.151371	-2.060208	1.110997	-1.50741	4.217815	4.825914	3.304557	-23.06008	-13.28703	-11.31091	2.997309	4.691634	-7.150363
9	2694.95	2211.845	-8054009	-8053965	-4383.886	45.13686	186.0079	10270.49	454.1635	359.4764	-7.735333	-7.269879	1.958658	-0.349753	2.293221	-0.056177	-0.052436	1.563509	3.839511	-6.09022	-4.447025	-1.785354	3.952615	2.95132	-2.910464
10	2828.869	2606.644	-10798321	-10797723	30347.18	2140.324	1535.583	3727268	-143.8008	-7.735333	2489.102	40.58116	1.088053	0.403797	3.902723	-0.291321	-2.618431	0.068471	4.977984	1.110105	-0.94312	2.477846	5.153582	-0.476642	0.079518
11	195.088	204.7402	-325682.7	-325686.4	-158.4831	1.372203	-4.613062	-38801.98	-2.688645	-7.269879	40.58116	6.67619	-0.022884	-0.018332	-0.000332	0.007042	0.015516	0.04225	0.075182	-0.05118	-0.043492	-0.01177	0.063571	-0.054797	0.016415
12	35.08571	26.17266	-55553.89	-55558.11	-476.9369	-22.569	-12.42497	-53411.29	4.151371	1.958658	1.088053	-0.022884	0.031376	-0.010696	0.008443	-0.006652	0.016943	0.024762	0.025106	-0.089475	-0.057231	-0.040142	0.02475	0.017144	-0.030313
13	-17.97702	-9.754506	14526.26	14531.11	368.7523	22.28837	13.37403	43540.79	-2.060208	-0.349753	0.403797	-0.018332	-0.010696	0.015879	0.003162	0.005884	-0.01716	-0.014905	-0.001551	0.055166	0.035188	0.034454	-0.000615	-0.004472	0.016978
14	13.56639	7.642827	-93640.68	-93632.71	529.9781	32.94668	22.38895	69465.53	1.110997	2.293221	3.902723	-0.000332	0.008443	0.003162	0.064938	-0.00461	-0.036792	0.001585	0.070142	-0.002032	-0.024245	0.024275	0.072524	0.016203	-0.013461
15	-9.395472	-2.229997	3189.947	3191.985	228.2041	11.61067	6.618315	26038.01	-1.50741	-0.356177	-0.291321	0.007042	-0.00652	0.005884	-0.00461	0.005192	-0.002687	-0.007895	-0.008767	0.031563	0.022657	0.015489	-0.009319	-0.003908	0.008422
16	24.49329	20.45434	6719.661	6707.551	-931.4997	-52.15125	-32.587	-112302.2	4.217815	-0.052436	-2.618431	0.015516	0.016943	-0.01716	-0.036792	-0.002687	0.057628	0.026556	-0.035449	-0.103878	-0.043683	-0.072027	-0.040273	0.003847	-0.026872
17	38.95188	28.00765	-38617.3	-38623.71	-654.2399	-33.65195	-19.54701	-74739.59	4.825914	1.563509	0.068471	0.04225	0.024762	-0.014905	0.001585	-0.007895	0.026556	0.032364	0.021446	-0.108107	-0.066749	-0.052792	0.020181	0.015404	-0.033496
18	46.25837	34.63566	-133546.7	-133538.2	290.1563	22.92827	19.01129	44593.87	3.304557	3.839511	4.977984	0.075182	0.025106	-0.001551	0.070142	-0.008767	-0.035449	0.021446	0.193582	-0.048175	-0.065507	0.016604	0.127894	0.028636	-0.029729
19	-131.5098	-87.72956	183134	183163.6	2816.525	135.711	79.72314	321540.3	-23.06008	-6.09022	1.110105	-0.05118	-0.089475	0.055166	-0.002032	0.031563	-0.103878	-0.108107	-0.048175	0.497087	0.28442	0.253712	-0.045113	-0.066846	0.147085
20	-87.22761	-55.97149	137789.1	137803.1	1451.628	69.40658	39.188	162501.5	-13.28703	-4.447025	-0.94312	-0.043492	-0.057231	0.035188	-0.024245	0.022657	-0.043683	-0.066749	-0.065507	0.28442	0.178677	0.134332	-0.064281	-0.044567	0.088635
21	-57.24686	-35.52575	46344.19	46364.04	1686.892	86.51528	52.72424	197432.1	-11.31091	-1.785354	2.477846	-0.01177	-0.040142	0.034454	0.024275	0.015489	-0.072027	-0.052792	0.016604	0.253712	0.134332	0.146629	0.018411	-0.024791	0.070343
22	44.99933	32.66325	-141244.4	-141235.6	371.9959	26.97984	20.9153	54471.51	2.997309	3.952615	5.153582	0.063571	0.02475	-0.000615	0.072524	-0.009319	-0.040273	0.020181	0.127894	-0.045113	-0.064281	0.018411	0.122956	0.029404	-0.02825
23	24.74246	19.371	-57050.28	-57051.38	-158.5223	-5.82787	-2.337009	-14263.37	4.691634	2.95132	-0.476642	-0.054797	0.017144	-0.004472	0.016203	-0.003908	0.003847	0.015404	0.028636	-0.066846	-0.044567	-0.024791	0.029404	0.025836	-0.027681
24	-45.08064	-33.45988	95433.37	95439.94	605.0511	28.83417	16.39452	67039.14	-7.150363	-2.910464	0.079518	0.016415	-0.030313	0.016978	-0.013461	0.008422	-0.026872	-0.033496	-0.029729	0.147085	0.088635	0.070343	-0.02825	-0.027681	0.053956

Figure 8: Covariance matrix 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	23	24
0	259938	256138	1.2E+08	1.2E+08	-114495	-6294.61	-3823.87	-1.2E+07	-957.687	-1393.79	13373.3	-2666.41	2.94481	-3.85731	16.5186	-2.97471	5.62894	4.18412	-1.98451	-34.3185	-16.8198	-20.57	-10.1206	-9.83842	-21.7521
1	256138	258038	1.5E+08	1.5E+08	-19263.8	261.359	-1901.45	-2032754	-1183.65	-1180.02	12247.4	-2832.32	3.38963	-2.46295	11.6304	1.20702	8.39176	-4.13275	-10.084	-15.4643	1.16597	-18.5757	-23.3566	-10.1654	-14.8971
2	1.2E+08	1.5E+08	3.3E+12	3.3E+12	5.1E+08	2.9E+07	9298938	5.3E+10	-3586050	600188	-1304696	-3.4E+07	36535.8	-16502.5	-26646.6	18243.6	54445.4	-29078	-74062.5	74348.2	89902.4	-28492.3	-116657	-13912.4	-2807.31
3	1.2E+08	1.5E+08	3.3E+12	3.3E+12	5.1E+08	2.9E+07	9302068	5.3E+10	-3586455	600090	-1305441	-3.4E+07	36534.8	-16500.8	-26651.4	18243.8	54437.9	-29077.1	-74054.5	74366.3	89905.8	-28476.2	-116645	-13913.8	-2798.65
4	-114495	-19263.8	5.1E+08	5.1E+08	5121724	201881	135507	5.3E+08	-15218.1	2762.65	-29026.5	2315.25	-37.4526	31.8358	-107.658	69.8587	-87.9306	-125.617	30.5783	692.874	377.583	342.99	17.3938	-31.1396	225.095
5	-6294.61	261.359	2.9E+07	2.9E+07	201881	10847.8	5755.11	2.1E+07	-541.645	203.95	-2125.82	185.315	-0.37207	3.60421	-7.99808	4.80709	-4.17462	-10.055	-3.31706	37.9987	24.905	16.1432	-5.72256	-1.01823	15.1153
6	-3823.87	-1901.45	9298938	9302068	135507	5755.11	5008.47	1.4E+07	-538.583	-23.2411	-1229.69	313.843	-1.34556	2.5995	-6.41278	1.40341	-8.18555	-2.71112	6.35528	28.179	10.6683	19.7475	9.91343	-1.49543	12.2662
7	-1.2E+07	-2032754	5.3E+10	5.3E+10	5.3E+08	2.1E+07	1.4E+07	5.6E+10	-1443015	397727	-3291478	147380	-3554.68	3415.54	-11365.9	7414.94	-8940.02	-13523.9	2549.56	71815	39675.7	35077.4	888.04	-2320.96	23386.7
8	-957.687	-1183.65	-3586050	-3586455	-15218.1	-541.645	-538.583	-1443015	775.076	358.481	-1115.3	-263.239	1.25855	0.76461	0.29935	-0.15777	0.23708	-1.2047	-2.83291	-4.85539	-1.1122	-3.18433	-2.80378	3.94402	-1.90637
9	-1393.79	-1180.02	600188	600090	2762.65	203.95	-23.2411	397727	358.481	454.203	-543.781	-252.573	0.64891	-0.03367	-0.62677	0.15812	0.83431	-1.42068	-2.36215	-0.87899	1.21832	-2.11138	-3.40665	2.91411	-0.70993
10	13373.3	12247.4	-1304696	-1305441	-29026.5	-2125.82	-1229.69	-3291478	-1115.3	-543.781	24015.2	1507.22	-0.81282	-4.70557	5.13442	-1.03976	7.17455	3.78789	0.73654	-10.2283	-4.93458	-9.4854	-4.35897	-5.69474	-7.36396
11	-2666.41	-2832.32	-3.4E+07	-3.4E+07	2315.25	185.315	313.843	147380	-263.239	-252.573	1507.22	4839.48	-1.68357	0.499	-1.08981	-0.12363	-2.54892	2.17794	5.52187	2.46304	4.56296	7.27256	2.02021	1.79777	
12	2.94481	3.38963	36535.8	36534.8	-37.4526	-0.37207	-1.34556	-3554.68	1.25855	0.64891	-0.81282	-1.68357	0.09153	-0.00062	0.00719	0.00029	0.00574	-0.00563	-0.01802	-0.00683	0.00622	-0.01327	-0.02537	0.00603	-0.00073
13	-3.85731	-2.46295	-16502.5	-16500.8	31.8358	3.60421	2.5995	3415.54	0.76461	-0.03367	-4.70557	0.499	-0.00062	0.01926	-0.0041	0.00101	-0.01277	-0.01049	-0.00788	0.02151	0.01803	0.01891	-0.00275	0.00269	0.02131
14	16.5186	11.6304	-26446.6	-26651.4	-107.658	-7.99808	-6.41278	-11365.9	0.29935	-0.62677	5.13442	-1.08981	0.00719	-0.0041	0.07922	-0.00354	0.02161	0.01592	-0.01164	-0.05042	-0.02188	-0.0313	-0.01668	-0.00112	-0.0267
15	-2.97471	1.20702	18243.6	18243.8	69.8587	4.80709	1.40341	7414.94	-0.15777	0.15812	-1.03976	-0.12363	0.00029	0.00101	-0.00354	0.00307	0.00196	-0.00599	-0.00579	0.01367	0.01299	0.0015	-0.00946	-0.00021	0.00493
16	5.62894	8.39176	54445.4	54437.9	-87.9306	-4.17462	-8.18555	-8940.02	0.23708	0.83431	7.17455	-2.54892	0.00574	-0.01277	0.02161	0.00196	0.06479	-0.01361	-0.06601	-0.06026	0.01281	-0.08285	-0.10144	0.00367	-0.04222
17	4.18412	-4.13275	-29078	-29077.1	-125.617	-10.055	-2.71112	-13523.9	-1.2047	-1.42068	3.78789	2.17794	-0.00563	-0.01049	0.01592	-0.00599	-0.01361	0.0484	0.0655	-0.02696	-0.05737	0.02195	0.08604	-0.00926	-0.01539
18	-1.98451	-10.084	-74062.5	-74054.5	30.5783	-3.31706	6.35528	2549.56	-2.83291	-2.36215	0.73654	5.52187	-0.01802	-0.00788	-0.01164	-0.00579	-0.00601	0.0655	0.22663	0.04231	-0.07504	0.11048	0.03065	-0.01825	0.02058
19	-34.3185	-15.4643	74348.2	74366.3	692.874	37.9987	28.179	71815	-4.85539	-0.87899	-10.2283	2.46304	-0.00683	0.02151	-0.05042	0.01367	-0.06026	-0.02696	0.04231	0.26084	0.11439	0.16375	0.06232	-0.01878	0.14027
20	-16.8198	1.16597	89902.4	89905.8	377.583	24.905	10.6683	39675.7	-1.1122	1.21832	-4.39458	-1.66097	0.00622	0.01803	-0.02188	0.01299	0.01281	-0.05737	-0.07504	0.11439	0.11621	0.10332	-0.10368	0.00074	0.06081
21	-20.57	-18.5757	-28492.3	-28476.2	342.99	16.1432	19.7475	35077.4	-3.18433	-2.11138	-9.4854	4.56296	-0.01327	0.01891	-0.0313	0.0015	-0.08285	0.02195	0.11048	0.16375	0.01332	0.16336	0.16292	-0.01752	0.09646
22	-10.1206	-23.3566	-116657	-116645	17.3938	-5.72256	9.91343	888.04	-2.80378	-3.40665	-4.35897	7.27256	-0.02537	-0.00275	-0.01668	-0.00946	-0.10144	0.08604	0.23065	0.06232	-0.10368	0.16292	0.30204	-0.0214	0.03861
23	-9.83842	-10.1654	-13912.4	-13913.8	-31.1396	-0.101823	-1.49543	-2320.96	3.94402	2.91411	-5.69474	-0.20201	0.00603	0.00269	-0.00112	-0.00021	0.00367	-0.00926	-0.01825	-0.01878	0.00074	-0.01752	-0.0214	0.02601	-0.0085
24	-21.7521	-14.8971	-2807.31	-2798.65	225.095	15.1153	12.2662	23386.7	-1.90637	-0.70993	-7.36396	1.79777	-0.00073	0.02131	-0.0267	0.00493	-0.04222	-0.01539	0.02058	0.14027	0.06081	0.09646	0.03861	-0.0085	0.09773

**Inferences:**

1. Accuracy of Bayes Classifier is 94.6%. It is less because it tries to find a general trend instead of focusing on neighbors as done in KNN.
2. The diagonal elements denote the variance of the attribute or data spread across the median. We can find the dispersion of the attribute by looking at it.
3. Off diagonal elements refer to covariance among the data attributes.

4

**Table 4 Comparison between classifiers based upon classification accuracy**

S. No.	Classifier	Accuracy (in %)
1.	KNN	89.5
2.	KNN on normalized data	100
3.	Bayes	94.6

**Inferences:**

1. KNN Normalized has highest and KNN has lowest accuracy.
2.  $KNN < Bayes < KNN \text{ Normalized}$ .
3. KNN performs better when data is normalized because accuracy increases because after normalization attributes with bigger range come in same range with that of the other attributes. Euclidean distances can't have different ranges for different attributes. The Bayes classifier is less because it tries to find a general trend instead of focusing on neighbors as done in KNN.