



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: **Priyanka Kumari**

Mobile No: **8328354314**

Roll Number: **B20307**

Branch: **Electrical Engineering**

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	86.9
3	89.6
5	89.3

Inferences:

1. The highest classification accuracy is obtained with K = 3
2. The k value initially increases from k=1 and k=3 and then decreases for k=5.
3. On increasing the value of k the data gets more accurate on increasing the larger values of k it
4. As the classification accuracy increases with the increase in value of K the number of diagonal elements increase.
5. In general diagonal elements tells how the data is correctly assigned to that particular class such that the accuracy increases it means number of diagonal elements increases.
6. As the classification accuracy increases with the increase in value of k the number of off-diagonal elements decreases as the off-diagonal elements indicate the elements which are not correctly predicted.
7. The off-diagonal elements indicate which are not correctly predicted so accuracy increases the off-diagonal elements decreases.

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	113	5
	4	215

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

	Prediction Outcome	
True Label	109	9
	4	215

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	96.1
3	97.3
5	96.1

Inferences:

1. The data normalization increases classification accuracy.
2. Data normalization centers the data and removes the noise, so the accuracy increases. But it is not true for every case. Sometimes , it removes the important features from the data and results in accuracy going down as can we see for K = 5.
3. The highest classification accuracy is obtained with K = 3
4. We can see an increase in the prediction accuracy with increasing value of K, with some exceptions like K=5
5. Increasing the value of K increases the prediction accuracy because the number of nearest neighbors, we can get the more number of classification as we increase the value of K so in Result its improving the accuracy.
6. With the increase in value of K the number of diagonal elements also increase.
7. As the accuracy is increasing with the increase in K, the True positive and True Negative are also increasing, which are the diagonal elements.
8. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements are decrease.
9. As the accuracy increase the False positive and False negative decreases, which are the off-diagonal elements.

3

	Prediction Outcome	
True Label	109	9
	9	210

Figure 7 Confusion Matrix obtained from Bayes Classifier

IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – IV

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

The classification accuracy obtained from Bayes Classifier is 94.7 %.

Table 3 Mean for class 0 and class 1

S. No.	Attribute Name	Mean	
		Class 0	Class 1
1.	X_Minimum	-	
2.	X_Maximum	273.418	723.656
3.	Y_Minimum	-	
4.	Y_Maximum	1583169.659	1431588.690
5.	Pixels_Areas	273.183	585.967
6.	X_Perimeter	843350.275	54.491
7.	Y_Perimeter	53.326	45.658
8.	Sum_of_Luminosity	135.762	62191.126
9.	Minimum_of_Luminosity	1382.762	96.236
10.	Maximum_of_Luminosity	40.073	130.452
11.	Length_of_Conveyer	0.123	1480.018
12.	TypeOfSteel_A300	-	-
13.	TypeOfSteel_A400	-	-
14.	Steel_Plate_Thickness	40.073	104.214
15.	Edges_Index	0.123	0.385
16.	Empty_Index	0.459	0.427
17.	Square_Index	0.592	0.513
18.	Outside_X_Index	0.108	0.020
19.	Edges_X_Index	0.550	0.608
20.	Edges_Y_Index	0.523	0.831
21.	Outside_Global_Index	0.288	0.608
22.	LogOfAreas	3.623	2.287
23.	Log_X_Index	2.057	1.227
24.	Log_Y_Index	1.848	1.318
25.	Orientation_Index	-0.314	0.136
26.	Luminosity_Index	-0.115	-0.116
27.	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.



IC 272: DATA SCIENCE - III LAB ASSIGNMENT – IV

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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
2	X_Maximu	46733.77351	-60848696.53	-320672.3293	-15750.50812	-12943.76426	-32609924.84	3686.072923	2040.904937	1237.643908	16.734	25.36021	-6.929295	4.696193	-1.515867	16.65354	22.50463	30.83904	-76.31962	-47.78156	-31.14733	27.67876	18.08286	-30.0931
3	Y_Maximu	-60848696.53	1.82181E+12	1027980976	83317353.38	160209448.9	48997689854	-5669890.139	-6007837.24	-7505510.38	-114611.2	-47711.37	21948.27	-59251.28	4294.736	-19165.63	-35306.43	-86404.07	168069.8	111447.7	73014.36	-82046.9	-50711.2	73811.61
4	Pixels_Are	-320672.3293	1027980976	104771842.6	6692648.9	10371695.26	9008476632	-154934.0074	6294.463585	10070.20623	547.0101	-492.1134	585.2306	200.1953	223.0561	-1121.193	-354.5732	556.0752	3456.879	1427.026	2840.741	980.3329	-300.211	575.0404
5	X_Perimet	-15750.50812	83317353.38	6692648.9	442770.572	706256.5009	557116030.4	-7764.04533	769.5856092	771.6039916	31.92388	-24.09284	38.16111	10.59581	10.99425	-67.82369	-13.28403	45.34169	183.0575	68.41173	169.1286	72.43566	-15.7026	28.52111
6	Y_Perimet	-12943.76426	160209448.9	10371695.26	706256.5009	1206390.51	807551258.1	-6894.471693	1492.073179	-1364.1952	10.20712	-17.5711	44.18238	-16.55017	6.495981	-65.41729	13.41058	63.25045	176.6405	44.05484	207.7917	105.1195	-21.062	19.50566
7	Sum_of_U	-32609924.84	48997689854	9008476632	557116030.4	807551258.1	8.19346E+11	-16498427.92	777671.2936	2214134.327	49759.91	-53267.33	58474.64	44601.85	25470.52	-123180.8	-50984.93	60033.13	361544.8	157340.8	278177.3	96509.49	-22290.5	62063.26
8	Minimum	3686.072923	-5669890.139	-154934.0074	-7764.04533	-6894.471693	-16498427.92	1458.213181	439.2359944	-153.833859	-1.972501	3.931511	-1.750045	1.077743	-1.455289	3.738841	4.623318	4.758855	-22.18673	-12.86067	-10.74723	3.816648	4.448267	-6.55741
9	Maximum	2040.904937	-6007837.239	6294.463585	769.5856092	1492.073179	777671.2936	439.2359944	333.3806022	2.285014006	-0.791317	1.768683	-0.221586	2.057703	-0.352958	-0.142446	1.57515	4.206583	-5.859388	-4.35841	-1.529243	4.136383	2.716174	-2.7371
10	Length_of	1237.643908	-7505510.376	10070.20623	771.6039916	-1364.195203	2214134.327	-153.8338585	2.285014006	2521.557073	-1.820728	1.321957	0.806365	3.925976	-0.192474	-2.696655	-0.534206	4.535627	2.03005	-0.001872	2.644925	4.369843	-0.4847	0.21099
11	Steel_Plat	16.73400129	-114611.1882	547.0100733	31.92388494	10.20712131	49759.90627	-1.972500539	-0.79131653	-1.82072829	0.729907	-0.008741	0.0147	-0.015494	0.019054	0.003184	-0.01538	-0.021143	0.041098	0.041366	0.019269	-0.02246	-0.0077	0.005483
12	Edges_Ind	25.36020808	-47711.36683	-492.1134083	-24.09283898	-17.57109505	-53267.33043	3.931510572	1.768682791	1.321956688	-0.008741	0.029323	-0.009277	0.007154	-0.006048	0.014692	0.022417	0.026357	-0.084016	-0.053519	-0.037595	0.024297	0.015975	-0.02755
13	Empty_Inc	-6.929295253	21948.26811	585.2306148	38.161117	44.18237773	58474.64332	-1.750044594	-0.22158645	0.806365021	0.0147	-0.009277	0.015302	0.004472	0.004944	-0.017655	-0.011599	0.003021	0.051673	0.030409	0.036164	0.005163	-0.00347	0.015267
14	Square_In	4.696192938	-59251.27802	200.1953395	10.59581308	-16.55017007	44601.84544	1.077474384	2.057702924	3.925975718	-0.015494	0.007154	0.00472	0.064486	-0.004106	-0.036326	-0.000653	0.070297	0.001334	-0.019666	0.021386	0.068654	0.016339	-0.0097
15	Outside_X	-1.51586711	4294.736092	223.0561229	10.9942463	6.495980628	25470.51969	-1.455288857	-0.35295772	-0.19247353	0.019054	-0.006048	0.004944	-0.004106	0.004743	-0.002219	-0.007306	-0.009753	0.029154	0.020886	0.01388	-0.00952	-0.00376	0.007482
16	Edges_X_I	16.65353756	-19165.62798	-1121.192855	-67.8236827	-65.41729197	-123180.7703	3.738841272	-0.14244573	-2.69665455	0.003184	0.014692	-0.017655	-0.036326	-0.002219	0.056908	0.022848	-0.038558	-0.098413	-0.039256	-0.073084	-0.04451	0.002776	-0.02567
17	Edges_Y_I	22.50462672	-35306.42553	-354.5731798	-13.28402598	13.41058397	-50984.93259	4.623318004	1.57515014	-0.53420574	-0.01538	0.022417	-0.011599	-0.000653	-0.007306	0.022848	0.030681	0.024941	-0.099278	-0.062596	-0.044652	0.023024	0.014378	-0.0311
18	Outside_C	30.83904331	-86404.06896	556.0751993	45.34168956	63.25045114	60033.13395	4.758854503	4.206582633	4.535626751	-0.021143	0.026357	0.003021	0.070297	-0.009753	-0.038558	0.024941	0.020859	-0.057832	-0.072752	0.019258	0.138071	0.033017	-0.03252
19	LogOfArea	-76.31961832	168069.821	3456.878751	183.05747	176.6404729	361544.7547	-22.18673088	-5.85938831	2.030050298	0.041098	-0.084016	0.051673	0.001334	0.029154	-0.098413	-0.099278	-0.057832	0.471457	0.266901	0.246904	-0.04394	-0.06701	0.135218
20	Log_X_Ind	-47.78156352	111447.6991	1247.025894	68.411727	44.05483883	157340.8395	-12.8606842	-4.35840986	-0.00187236	0.041366	-0.053159	0.030409	-0.019666	0.020886	-0.039256	-0.062596	-0.072752	0.266901	0.167866	0.124113	-0.06631	-0.04408	0.081643
21	Log_Y_Ind	-31.14733373	73014.3565	2840.741336	169.1285737	207.7916999	278177.3419	-10.74722663	-1.52924345	2.644925298	0.019269	-0.037595	0.036164	0.023186	0.01388	-0.073084	-0.044652	0.019258	0.246904	0.124113	0.156846	0.029178	-0.02546	0.064575
22	Orientatio	27.67876208	-82046.87983	980.3328791	72.43565577	105.1195171	96509.49238	3.816647634	4.136382598	4.369842525	-0.022463	0.024297	0.005163	0.068654	-0.009525	-0.044513	0.023024	0.138071	-0.043944	-0.066308	0.029178	0.133168	0.030895	-0.02766
23	Luminosit	18.08285838	-50711.211	-300.2109746	-15.70264171	-21.06203908	-22290.54261	4.4482667	2.716173704	-0.4847024	-0.007703	0.015975	-0.003468	0.016339	-0.003762	0.002776	0.014378	0.033017	-0.067013	-0.044084	-0.025463	0.030895	0.027438	-0.02644
24	SigmoidOf	-30.09314133	73811.60519	575.0403684	28.52110558	19.50566072	62063.2628	-6.55740735	-2.73710077	0.210989671	0.005483	-0.02755	0.015267	-0.009701	0.007482	-0.02567	-0.031098	-0.032521	0.135218	0.081643	0.064575	-0.02766	-0.02644	0.049322

Figure 8: Covariance matrix for class 0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
1		X_Maximu	Y_Maximu	Pixels_Area	X_Perimet	Y_Perimet	Sum_of_Lumin	Minimum_of	Maximum	Length_of_C	Steel_Plate	TE	Edges_Ind	Empty_Inc	Square_In	Outside_X	Edges_X_I	Edges_Y_I	Outside_C	LogOfArea	Log_X_Inc	Log_Y_Inc	Orientatio	Luminosit	SigmoidOf
2	X_Maximu	256526.309	1.12E+08	-22254.624	1101.0786	-1973.56461	-2334975.575	-1224.8085	-744.043	13220.0789	-1932.61914	8.913916	-3.8064	10.89266	1.504328	6.694786	-5.01836	-16.5642	-13.7813	5.30599	-21.2042	-25.8957	-8.45195	-14.2211	
3	Y_Maximu	111783525	3.12E+12	322720784	20351188	4659661.77	32954294851	-3631824.7	-43295.9	3999505.64	-36154262.6	23556.3	-19251	-38009.7	13457.3	64532.97	-22198.8	-74705.2	15298.09	64300.3	-63426.8	-119870	-14717.9	-37674.9	
4	Pixels_Area	-22254.6237	3.23E+08	4714217.3	178492.15	129451.109	488874179.5	-15631.976	-300.304	-23834.665	4262.207954	-47.6455	35.6195	-90.6336	52.90864	-101.643	-96.0566	55.17783	653.0513	330.779	355.115	65.4194	-32.3838	218.948	
5	X_Perimet	1101.07865	20351188	178492.15	9807.2032	5546.89855	18662200.1	-570.11602	30.14967	-1446.8768	282.1131174	-1.33167	4.155596	-7.3181	3.971901	-4.84985	-9.17608	-2.1516	36.6199	23.5571	16.8636	-3.75763	-1.11861	15.5083	
6	Y_Perimet	-1973.56461	4659661.77	129451.11	5546.8985	5000.64669	13453352.78	-557.42319	-79.1464	-1139.3109	438.5595695	-2.24421	2.951694	-6.49605	1.204469	-8.61151	-2.36737	7.109846	29.02755	10.6809	21.0247	11.0455	-1.55636	13.014	
7	Sum_of_L	-2334975.57	3.3E+10	488874179	18662200	13453352.8	50945346301	-1463160.7	84723.03	-2735155.1	343512.3962	-4688.9	3985.075	-9652.58	5577.969	-10534.6	-10271.9	5462.295	67782.66	34740.3	36734.8	6364.12	-2282.38	22864.8	
8	Minimum	-1224.80855	-3631825	-15631.976	-570.11602	-557.423186	-1463160.736	733.908876	348.0448	-993.31126	-204.836019	1.066368	0.591072	0.775182	-0.15145	0.427209	-0.83326	-2.22434	-5.04259	-1.29929	-3.28658	-2.50299	3.68376	-1.98355	
9	Maximum	-744.043156	-43295.9	-300.30378	30.149672	-79.146408	84723.02772	348.044835	406.4608	-381.09265	-205.3942	4.291918	-0.02454	-0.26703	0.04392	0.877571	-1.08968	-2.01841	-1.50427	0.67825	-2.16518	-2.8738	2.78648	-0.96	
10	Length_of	13220.0789	3999506	-23834.665	-1446.8768	-1139.31087	-2735155.116	-993.31126	-381.093	23100.7694	1243.443056	-0.09047	-5.15952	2.468171	-0.69776	6.591052	1.97125	-3.13774	-7.95323	-1.43972	-10.5673	-7.4308	-4.54679	-5.96676	
11	Steel_Plate	-1932.61914	-36154263	4262.208	282.11312	438.559569	343512.3962	-204.83602	-205.394	1243.44306	5645.306414	-1.3306	0.699194	-1.13384	-0.16545	-3.44259	2.058128	6.23469	3.626633	-1.37643	5.40272	7.84601	-1.6621	2.39033	
12	Edges_Ind	8.91391616	23556.3	-47.645532	-1.3316684	-2.24420986	-4688.897042	1.06636816	0.429118	-0.0904651	-1.33060088	0.08965	-0.00063	0.010929	6.45E-05	0.008301	-0.00333	-0.01658	-0.01211	0.00465	-0.01652	-0.02434	0.00464	-0.00405	
13	Empty_Inc	-3.80639693	-19251	35.619499	4.1555961	2.95169386	3985.075354	0.59107157	-0.02454	-5.1595223	0.699193586	-0.00063	0.020283	-0.00202	0.001242	-0.01249	-0.01101	-0.00752	0.026336	0.02169	0.02161	-0.00415	0.0021	0.02383	
14	Square_In	10.8926577	-38009.67	-90.633582	-7.3180958	6.49605456	-9652.577328	0.77518161	-0.26703	2.46817139	-1.13384041	0.010929	-0.00202	0.008273	-0.00291	0.019744	0.14881	-0.01558	-0.05315	-0.02053	-0.03335	-0.02057	0.00137	-0.02827	
15	Outside_X	1.50432796	13457.3	52.908645	36719011	1.20446853	5577.96928	-0.1514255	0.43932	-0.6977556	-0.16545127	6.45E-05	0.001242	-0.00291	0.002467	-0.001752	-0.00529	-0.00052	0.011616	0.0115	0.00132	-0.00839	-0.00022	0.00646	
16	Edges_X_I	6.69478626	63452.97	-101.64279	-4.8498518	-8.61151492	-10534.58498	1.067240931	0.877571	6.59105182	-3.42459272	0.008301	-0.01249	0.019744	0.001752	0.065074	-0.01386	-0.06755	-0.06618	0.01098	-0.08629	-0.10253	0.00434	-0.04488	
17	Edges_Y_I	-5.01836115	-22198.76	-96.056608	-9.1760789	-2.36736536	-10271.86462	-0.8332641	-1.08968	1.97124953	2.058128085	-0.00333	-0.01101	0.014881	-0.00529	-0.01386	0.049202	0.064322	-0.02518	-0.05805	0.02378	0.08641	-0.00723	-0.01687	
18	Outside_C	-6.56451453	-74705.16	55.177829	-2.1516019	7.10984561	5462.295442	-2.2243437	-0.01841	-3.1377411	6.623468512	-0.01658	-0.00752	-0.01558	-0.0052	-0.06755	0.064322	0.227474	0.47656	-0.07282	0.11336	0.22928	-0.01479	0.02182	
19	LogOfArea	5.37813062	15298.09	653.0514	36.619899	29.027535	67782.652	-0.5045887	-1.50427	-7.9532299	3.626632766	-0.01211	0.026336	-0.05315	0.011616	-0.06618	-0.02518	0.06756	0.270784	0.11641	0.17702	0.0729	-0.01936	0.14744	
20	Log_X_Inc	5.30599098	64300.31	330.77912	23.55709	16.8609231	34740.28603	-1.2992892	6.78254	-1.4397173	1.37662465	0.004646	0.021686	-0.02053	0.011505	0.010977	-0.05805	-0.07282	0.116409	0.11864	0.01736	-0.10068	-0.0004	0.06466	
21	Log_Y_Inc	-21.2042167	-63426.82	355.11461	16.863627	21.0246531	36734.77789	-3.2865781	-2.16518	-10.567281	5.402715518	-0.01652	0.021607	-0.03335	0.001317	-0.08629	0.023781	0.113361	0.177016	0.01736	0.17785	0.16863	-0.01723	0.1025	
22	Orientatio	-25.8956548	-119869.7	65.419425	-3.757626	11.0454571	6364.118791	-2.502995	-2.8738	-7.4308047	7.846013411	-0.02434	-0.00415	-0.02057	-0.00839	-0.10253	0.086409	0.229284	0.072903	-0.10068	0.16863	0.30151	-0.01872	0.0412	
23	Luminosit	-8.4519527	-14717.93	-32.383836	-1.1186086	-1.55635527	-2282.381369	3.68376204	2.786478	-4.5467858	-1.66209186	0.004642	0.0021	0.001372	-0.00202	0.004337	-0.00723	-0.01479	-0.01936	-0.0004	-0.01723	-0.01872	-0.04252	-0.00898	
24	SigmoidOf	-14.2210892	-37674.92	218.94798	15.508343	1.01339537	22864.84792	-1.9835536	-0.96	-5.9667605	2.39033055	-0.00405	0.02383	-0.02827	0.004643	-0.04488	-0.01687	0.021824	0.147443	0.06466	0.1025	0.0412	-0.00898	0.10227	

Figure 9: Covariance matrix for class 1

Inferences:

1. The accuracy of Bayes Classifier is 94.7. it is lesser than previous classification approaches because it considers normal distribution.
2. Nature of value along the diagonal is high for some and low for others because some particular attributes follow Standard normal distribution, hence the prediction is better for them.
3. Off-diagonal elements have varying values. Two pairs of attributes with maximum cov is (Sum_of_Luminosity, Y_Maximum) and (Pixels_Areas, Sum_of_Luminosity).

4

Table 4 Comparison between classifiers based upon classification accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN	94.7
2.	KNN on normalized data	96.100
3.	Bayes	94.659

Inferences:

1. Highest accuracy is :- KNN on normalized data : 96.100, lowest accuracy:- Bayes: 94.659
2. KNN on normalized data > KNN > Bayes.
3. Usually Bayes classifier has higher accuracy but in this case the KNN classifier used is using Normalised data points that is why its accuracy is high.