



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: Gayatri Shridhar Kapse

Mobile No: 8623916561

Roll Number: B20199

Branch:EE

---

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	86.9436
3	89.6142
5	89.3175

#### Inferences:

1. The highest classification accuracy is obtained with K = 3
2. Increasing the value of K at first it increases and then decreases again.
3. By increasing the value of K we are actually considering the Euclidian distances of K points from the test point that's why we get high accuracy on increasing the K value.
4. As the classification accuracy increases with the increase in value of K the number of diagonal elements increase.
5. The reason for increase in the number of diagonal element is because as the accuracy increase more number of predicted values are equal to the actual value or we can say our prediction is right.

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	96.1424
3	97.3293
5	96.1424

**Inferences:**

1. Normalization increases classification accuracy.
2. There is an increase in classification accuracy after normalization because as we are scaling the values of the attributes so that their will not be any data point that will be dominating the output.
3. The highest classification accuracy is obtained with K =3.
4. The classification accuracy first increases and the decreases with the increase in value of K.
5. By increasing the value of K we are actually considering the Euclidian distances of K points from the test point that's why we get high accuracy on increasing the K value thereby calculating the eulcidian distance between the test data point and K number of train data points.

3

	Prediction Outcome	
True Label	105	13
	6	213

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 94.4%.

## IC 272: DATA SCIENCE - III

### LAB ASSIGNMENT – IV

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

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	7779.663	585.967
6.	X_Perimeter	393.835	54.491
7.	Y_Perimeter	273.183	45.658
8.	Sum_of_Luminosity	843350.275	62191.126
9.	Minimum_of_Luminosity	53.326	96.236
10.	Maximum_of_Luminosity	135.762	130.452
11.	Length_of_Conveyer	1382.762	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

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	-6.925295	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	-569890.139	-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_Perime	-15750.50812	83317353.38	6692648.9	442770.572	706256.5009	557116030.4	-7764.04533	769.5856092	771.6039916	31.92388	-24.05284	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_Perime	-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_L	-32609924.84	48997689854	9008476632	557116030.4	807551258.1	8.15946E+11	-16498427.92	77671.2936	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	-569890.139	-154934.0074	-7764.04533	-6894.471693	-16498427.92	1458.213181	439.2359944	-153.833859	-1.972501	3.9231511	-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	-1.52943	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	4.696192938	-59251.27802	200.1953309	13.92388494	10.20712131	49759.90627	-1.97250539	-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_In	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_In	-6.92529523	21948.20811	585.2306148	38.1611117	44.18237773	58474.64332	-1.750044594	-0.22158645	0.806365021	0.0147	-0.009277	0.015302	0.00472	0.004944	-0.017655	-0.011399	0.003021	0.051673	0.030409	0.036164	0.005163	-0.00347	0.015267
14	Square_In	4.696192938	-59251.27802	200.1953309	13.92388494	10.20712131	49759.90627	-1.97250539	-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
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	-1121.192853	-67.82368627	-65.41729157	-123180.7703	3.738841272	-0.14245473	-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	-354.5731706	-13.28402598	13.41058397	-50984.93259	4.623318004	1.57515014	-0.53420574	-0.01538	0.022417	-0.015599	-0.000653	-0.007306	0.022848	0.030681	0.024941	-0.099278	-0.062596	-0.044852	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	183.05731	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_In	-47.78156152	111447.6991	1427.025894	68.411727	44.05483883	157340.8395	-12.8606842	-4.35840986	-0.00187236	0.041366	-0.053519	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_In	-31.14733373	73014.3565	2840.741336	169.1285737	207.7916999	278177.3419	-10.74722963	-1.52924345	2.644925298	0.019269	-0.037595	0.036164	0.023186	0.01388	-0.073084	-0.044552	0.019258	0.246904	0.124113	0.156846	0.029178	-0.02546	0.064575
22	Orientatio	27.67876208	-82046.97983	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	0.030317	-0.067013	-0.044084	-0.025463	0.030895	0.027438	-0.02644
24	SigmoidOf	-30.09314133	73811.60519	575.0403684	28.52110558	19.50568072	62063.2628	-6.55740735	-2.73710077	0.210989671	0.005483	-0.02755	0.015267	-0.009701	0.007482	-0.02567	-0.031098	0.023251	0.135218	0.081643	0.064575	-0.02766	-0.02644	0.049322

class\_0

		X_Maximum	Y_Maximum	Pixels_Area	X_Perimeter	Y_Perimeter	Sum_of_Lumin	Minimum_of_Maximum	Length_of_Steel_Plate	Edges_In	Empty_In	Square_In	Outside_X	Edges_X	Edges_Y	Outside_C	LogOfArea	Log_X_In	Log_Y_In	Orientatio	Luminosit	SigmoidOf	Area	
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	1.12E+08	322720784	20351188	4969651.77	32954294851	3631824.7	43295.9	3999505.64	36154262.6	2356.3	19251	-38009.7	13457.3	64532.97	-22198.8	-74705.2	15298.09	64300.3	63426.8	-119870	-14717.9	-37674.9
4	Pixels_Are	-22254.6237	3.23E+08	4714217.3	178492.15	189451.109	488874179.5	-15631.976	-300.304	-2384.665	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_Perime	1101.07865	20351188	178492.15	9807.2032	5546.89855	18662200.1	1070.11602	301.4967	-1446.8768	282.1131174	1.33167	4.155596	-7.3181	3.971901	-4.84985	-9.17608	-2.1516	36.6199	25.5571	16.8636	-3.75763	-1.1861	15.5083
6	Y_Perime	-1973.56461	4659662	129451.11	5546.8985	5000.64669	13453352.78	-557.42319	-79.1464	-1139.3109	438.5595959	-2.24421	2.951694	6.49605	8.61151	-2.36737	7.109846	79.07755	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	5652.58	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	97.146408	84723.02772	348.044835	406.4608	-381.09925	-205.3942	0.429118	-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_Plat	-1932.61914	36154263	4262.208	282.11312	438.559596	343512.3962	-204.83602	-205.394	1243.44306	5645.306414	-1.3306	0.899194	1.13384	-0.16545	-3.44259	2.058128	6.623469	3.626633	-1.37643	5.40272	7.84601	-1.6621	2.39033
12	Edges_In	8.91391616	23556.3	47.645532	-1.3316684	-2.24420986	4688.897042	1.06636816	0.429118	-0.0904651	0.33006088	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_In	-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	10.8926677	-38009.67	90.633582	-7.3180958	-6.49604556	-9652.57328	0.77518161	-0.26703	2.46817139	1.13384014	0.010929	-0.00202	0.082373	-0.00291	0.019744	0.014881	-0.01558	-0.05315	-0.02053	-0.03335	-0.02057	0.00137	-0.02827
15	Outside_X	1.50432796	13457.3	52.908645	3.9719011	1.20446853	557.969289	1.0514255	0.04392	-0.697556	0.16455127	6.45E-05	0.001242	-0.00291	0.002467	-0.001752	-0.00529	-0.0052	0.011616	0.0115	0.00132	-0.00839	-0.00022	0.00464
16	Edges_X	6.69478626	64532.97	101.64279	4.8498518	1.61544592	10453.58498	0.42720993	0.877571	6.59105218	3.44259172	0.008301	-0.01249	0.019744	0.01752	0.065074	-0.01386	-0.06755	-0.06618	0.01098	-0.08629	-0.10253	0.00434	-0.04488
17	Edges_Y	-5.01836115	-27198.8	-96.056608	9.1760789	2.36736536	-1071.85642	-0.8326461	-1.08968	-1.9724953	2.058128083	-0.00333	-0.0101	0.014881	-0.00529	0.01386	0.049202	0.064327	-0.02518	-0.05805	0.02378	0.08641	-0.00723	-0.01687
18	Outside_C	-16.5642543	74705.16	55.178729	-2.1561079	1.0984561	5462.295447	-2.2343347	-2.01841	-1.3177411	6.623468512	0.01658	-0.00752	0.01558	-0.0052	-0.06755	0.064327	0.227474	0.047656	-0.07282	0.11336	0.22798	-0.01479	-0.02182
19	LogOfArea	13.78136036	15298.09	65305.314	36.618999	30.0275535	67782.6567	-150427	-9532299	3.62663276	0.01211	0.026336	-0.05315	0.011616	-0.06618	-0.02518	0.047656	0.270784	0.11641	0.17702	0.0729	0.01936	0.14744	
20	log_X_In	5.30599098	64300.31	330.77912	23.55709	10.6809231	34740.28603	1.2992892	6.878254	1.4397173	1.3764265	0.004646	0.016636	-0.02053	0.011505	0.010977	-0.05805	-0.07282	0.116409	0.11864	0.0736	-0.10068	-0.0004	0.06466
21	log_Y_In	-21.2042167	63426.82	65.511461	16.863627	12.046531	36734.7789	-2.865781	-2.16518	-1.0562781	4.547175518	-0.01652	0.00167	-0.0335	-0.00137	-0.08629	0.023781	0.113361	0.177016	0.01376	0.17378	0.16863	-0.01723	0.1025
22	Orientatio	-25.8956458	119869.7	65.419425	3.757626	1.0454571	6356.11894	-2.502995	-2.8738	-7.4308047	7.84601341	-0.02434	-0.0145	0.02057	-0.00839	-0.10253	0.086409	0.229824	0.072903	0.10068	0.16865	-0.30151	-0.01872	-0.0412
23	Luminosit	-8.45195127	14717.93	32.383836	-1.186086	-1.5563557	-2282.381369	3.68376204	2.786478	4.5467858	-1.66209834	0.004642	-0.00021	0.001372	-0.00022	0.004338	-0.00673	-0.01479	-0.01936	-0.0004	-0.01723	-0.01872	0.02452	-0.00898
24	SigmoidOf	-14.2210892	-37674.92	218.94798	15.508343	13.0139537	2884.864732	-1.9835536	-0.96	5.967605	2.3903055	-0.00405	0.00283	-0.02827	0.004643	-0.04488	-0.01672	0.021824	0.147443	0.06466	0.1025	0.0412	-0.00898	0.10227

4

Table 4 Comparison between classifiers based upon classification accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN	89 . 6
2.	KNN on normalized data	97 . 3
3.	Bayes	94 . 4

**Inferences:**

1. KNN classifier with normalized data has highest accuracy and KNN classifier has the lowest accuracy.
2.  $KNN < Bayes < KNN \text{ on normalized data}$ .
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.