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: **Mobile No:**

Roll Number: **Branch:**

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	**0.839**
3	**0.890**
5	**0.896**

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 because we predict from larger amount of conclusions.
4. As the classification accuracy increases with the increase in value of K the number of diagonal elements increase .
5. Reason for increase in diagonal elements is with the increase of accuracy more no.s of positive and negative values will be predicted correctly.
6. As the classification accuracy increases with the increase in value of K , the number of off-diagonal elements decrease .
7. This is because with increase in accuracy there will be less false prediction.

2 a.

	Prediction Outcome	
True Label	104	4
	9	219

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

	Prediction Outcome	
True Label	105	3
	7	221

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

	Prediction Outcome	
True Label	104	4
	7	221

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	**0.961**
3	**0.970**
5	**0.967**

Inferences:

1. Data normalization increases classification accuracy.
2. Because in knn classification we measure Euclidian distance which get more accurate with normalization
3. The highest classification accuracy is obtained with K =2.
4. Accuracy initially increases with K but then decreases
5. increasing the value of K increases/decreases the prediction accuracy.
6. As the classification accuracy increases/decreases with the increase in value of K the number of diagonal elements increase.
7. With more accuracy more values will be predicted correctly and hence add up to diagonal elements

8. State the reason for increase/decrease in off-diagonal elements.
   Note: Dummy values have been filled in the confusion matrix. Replace it with values obtained by you.

3

		Prediction Outcome
True Label	74	15
	2	133

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 93.992%.

Table 3 Mean for class 0 and class 1

S. No.	Attribute Name	Mean	
		Class 0	Class 1
1.	X_Minimum	285.061	731.24
2.	X_Maximum		
3.	Y_Minimum		
4.	Y_Maximum	1649336.19	1511122.61
5.	Pixels_Areas	7622.655	567.684
6.	X_Perimeter	384.657	52.212
7.	Y_Perimeter	255.35	44.927
8.	Sum_of_Luminosity	835433.60	59982.029
9.	Minimum_of_Luminosity	55.061	95.166
10.	Maximum_of_Luminosity	137.11	130.192
11.	Length_of_Conveyer	1384.028	1480.023
12.	TypeOfSteel_A300		
13.	TypeOfSteel_A400		
14.	Steel_Plate_Thickness	40.179	101.229
15.	Edges_Index	0.13	0.383
16.	Empty_Index	0.457	0.422
17.	Square_Index	0.603	0.513
18.	Outside_X_Index	0.107	0.019
19.	Edges_X_Index	0.552	0.614
20.	Edges_Y_Index	0.527	0.842
21.	Outside_Global_Index	0.307	0.622
22.	LogOfAreas	3.599	2.287
23.	Log_X_Index	2.041	1.217
24.	Log_Y_Index	1.83	1.323
25.	Orientation_Index	-0.3	0.15
26.	Luminosity_Index	-0.102	-0.12
27.	SigmoidOfAreas	0.916	0.541

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

Covariance matrix for class 0

	X_Minimum	X_Maximum	Y_Minimum	Y_Maximum	Pixels_Areas	X_Perimeter	Y_Perimeter	Sum_of_Luminosity	Minimum_of_L	Maximum_of_L	Length_of_Conveyer	TypeOfSteel_A	TypeOfSteel_A	Steel_Plate_Thickness	Edges_Index	Empty_Index	Square_Index	Outside_X_Index	Edges_X_Index	Edges_Y_Index	Outside_Global_Index	LogOfAreas	Log_X_Index	Log_Y_Index	Orientation_Index	Luminosity_Index	SigmoidOfAreas
X_Minimum	70682	58002	-9E+07	-9E+07	-708835	-34840	-22963	-8E+07	6515.1	3109.9	2381.1	3.7882	-3.7882	141.15	39.125	-18.026	12.745	-9.2509	26.024	39.056	44.132	-132.52	-86.775	-59.002	42.468	27.761	-45.436
X_Maximum	58002	55930	-8E+07	-8E+07	-365274	-18028	-13109	-4E+07	4210.8	2536.1	2209.9	3.4691	-3.4691	148.27	29.804	-9.7743	6.125	-1.5999	22.899	27.729	29.934	-87.664	-54.246	-37.19	28.16	21.631	-33.868
Y_Minimum	-9E+07	-8E+07	2E+12	2E+12	4E+08	4E+07	1E+08	-4E+09	-7E+06	-8E+06	-9E+06	-2493.8	2493.8	-224331	-60533	18850	-79612	4085	-12532	-44847	-116709	200578	139375	70080	-121797	-64931	91950
Y_Maximum	-9E+07	-8E+07	2E+12	2E+12	4E+08	4E+07	1E+08	-3E+09	-7E+06	-8E+06	-9E+06	-2494	2494	-224343	-60542	18877	-79632	4087.4	-12568	-44832	-116669	200681	139396	70211	-121730	-64946	91961
Pixels_Areas	-708835	-365274	4E+08	4E+08	8E+07	5E+06	8E+06	7E+09	-161145	-3709.5	12622	-17.866	17.866	-178.25	-538.34	537.55	257.97	245.74	-1070.1	-490.16	401.54	3442.1	1557.5	2565.5	725.45	-318	635.42
X_Perimeter	-34840	-18028	4E+07	4E+07	5E+06	333154	507708	4E+08	-7868.7	211.1	1040.4	-0.9017	0.9017	-4.2205	-26.265	34.427	15.574	11.993	-63.899	-21.828	36.827	177	74.346	147.77	55.789	-15.758	31.339
Y_Perimeter	-22963	-13109	1E+08	1E+08	8E+06	507708	851311	6E+08	-6297.6	870.03	-614.44	-0.565	0.565	-11.986	-17.818	35.659	-6.2354	7.1171	-55.932	1.8796	48.484	152.89	46.477	163.31	77.489	-17.474	20.489
Sum_of_Luminosity	-8E+07	-4E+07	-4E+09	-3E+09	7E+09	4E+08	6E+08	7E+11	-2E+07	-179222	2E+06	-2015.7	2015.7	-32870	-58811	56062	47317	28177	-120352	-63220	47455	369489	172825	262100	75932	-26436	69307
Minimum_of_Luminosity	6515.1	4210.8	-7E+06	-7E+06	-161145	-7868.7	-6297.6	-2E+07	1567.6	491.28	-135.35	-0.0027	0.0027	-2.2418	4.776	-2.1281	1.3296	-1.6575	4.2475	5.2214	4.6234	-24.323	-14.189	-11.818	3.9902	4.9996	-7.354
Maximum_of_Luminosity	3109.9	2536.1	-8E+06	-8E+06	-3709.5	211.1	870.03	-179222	491.28	388.42	-6.2211	-0.1131	0.1131	-5.4813	2.3782	-0.4632	2.1855	-0.421	0.254	1.9959	4.238	-7.0727	-5.1211	-2.1444	4.34	3.136	-3.1645
Length_of_Conveyer	2381.1	2209.9	-9E+06	-9E+06	12622	1040.4	-614.44	2E+06	-135.35	-6.2211	2531.7	0.7769	-0.7769	29.072	1.1807	0.5429	4.0316	-0.2604	-2.7068	-0.0898	5.4426	1.1839	-0.7654	2.4166	5.0988	-0.4979	0.1123
TypeOfSteel_A300	3.7882	3.4691	-2493.8	-2494	-17.866	-0.9017	-0.565	-2015.7	-0.0027	-0.1131	0.7769	0.0026	-0.0026	0.1021	-0.0002	-0.0007	0.0003	-0.0002	0.0005	0.0011	0.0018	-0.002	-0.0017	-0.0008	0.0015	-0.0008	0.0001
TypeOfSteel_A400	-3.7882	-3.4691	2493.8	2494	17.866	0.9017	0.565	2015.7	0.0027	0.1131	-0.7769	-0.0026	0.0026	-0.1021	0.0002	0.0007	-0.0003	0.0002	-0.0005	-0.0011	-0.0018	0.002	0.0017	0.0008	-0.0015	0.0008	-0.0001
Steel_Plate_Thickness	141.15	148.27	-224331	-224343	-178.25	-4.2205	-11.986	-32870	-2.2418	-5.4813	29.072	0.1021	-0.1021	4.8397	-0.0172	-0.0146	-0.0019	0.0052	0.0138	0.0301	0.0475	-0.0369	-0.0302	-0.011	0.041	-0.0409	0.0119
Edges_Index	39.125	29.804	-60533	-60542	-538.34	-26.265	-17.818	-58811	4.776	2.3782	1.1807	-0.0002	0.0002	-0.0172	0.0346	-0.0115	0.0082	-0.0068	0.0185	0.0267	0.0271	-0.096	-0.0611	-0.0436	0.0263	0.021	-0.0325
Empty_Index	-18.026	-9.7743	18850	18877	537.55	34.427	35.659	56062	-2.1281	-0.4632	0.5429	-0.0007	0.0007	-0.0146	-0.0115	0.0168	0.0024	0.0059	-0.0181	-0.0144	0.0011	0.055	0.0344	0.0367	0.0014	-0.0054	0.0173
Square_Index	12.745	6.125	-79612	-79632	257.97	15.574	-6.2354	47317	1.3296	2.1855	4.0316	0.0003	-0.0003	-0.0019	0.0082	0.0024	0.0656	-0.0051	-0.036	0.001	0.0736	-0.0049	-0.025	0.0204	0.0712	0.0168	-0.0126
Outside_X_Index	-9.2509	-1.5999	4085	4087.4	245.74	11.993	7.1171	28177	-1.6575	-0.421	-0.2604	-0.0002	0.0002	0.0052	-0.0068	0.0059	-0.0051	0.0055	-0.002	-0.0082	-0.0106	0.0323	0.0235	0.0155	-0.0107	-0.0044	0.0084
Edges_X_Index	26.024	22.899	-12532	-12568	-1070.1	-63.899	-55.932	-120352	4.2475	0.254	-2.7068	0.0005	-0.0005	0.0138	0.0185	-0.0181	-0.036	-0.002	0.0609	0.0266	-0.04	-0.105	-0.0428	-0.0754	-0.0438	0.0059	-0.029
Edges_Y_Index	39.056	27.729	-44847	-44832	-490.16	-21.828	1.8796	-63220	5.2214	1.9959	-0.0898	0.0011	-0.0011	0.0301	0.0267	-0.0144	0.001	-0.0082	0.0266	0.0339	0.0244	-0.1091	-0.0686	-0.051	0.0234	0.018	-0.0347
Outside_Global_Index	44.132	29.934	-116709	-116669	401.54	36.827	48.484	47455	4.6234	4.238	5.4426	0.0018	-0.0018	0.0475	0.0271	0.0011	0.0736	-0.0106	-0.04	0.0244	0.2107	-0.06	-0.0753	0.0175	0.141	0.0327	-0.0305
LogOfAreas	-132.52	-87.664	200578	200681	3442.1	177	152.89	369489	-24.323	-7.0727	1.1839	-0.002	0.002	-0.0369	-0.096	0.055	-0.0049	0.0323	-0.105	-0.1091	-0.06	0.4962	0.2842	0.2561	-0.0516	-0.0763	0.1454
Log_X_Index	-86.775	-54.246	139375	139396	1557.5	74.346	46.477	172825	-14.189	-5.1211	-0.7654	-0.0017	0.0017	-0.0302	-0.0611	0.0344	-0.025	0.0235	-0.0428	-0.0686	-0.0753	0.2842	0.1804	0.1325	-0.0715	-0.0502	0.0879
Log_Y_Index	-59.002	-37.19	70080	70211	2565.5	147.77	163.31	262100	-11.818	-2.1444	2.4166	-0.0008	0.0008	-0.011	-0.0436	0.0367	0.0204	0.0155	-0.0754	-0.051	0.0175	0.2561	0.1325	0.1566	0.0224	-0.0297	0.0699
Orientation_Index	42.468	28.16	-121797	-121730	725.45	55.789	77.489	75932	3.9902	4.34	5.0988	0.0015	-0.0015	0.041	0.0263	0.0014	0.0712	-0.0107	-0.0438	0.0234	0.141	-0.0516	-0.0715	0.0224	0.1335	0.0325	-0.0286
Luminosity_Index	27.761	21.631	-64931	-64946	-318	-15.758	-17.474	-26436	4.9996	3.136	-0.4979	-0.0008	0.0008	-0.0409	0.021	-0.0054	0.0168	-0.0044	0.0059	0.018	0.0327	-0.0763	-0.0502	-0.0297	0.0325	0.0297	-0.03
SigmoidOfAreas	-45.436	-33.868	91950	91961	635.42	31.339	20.489	69307	-7.354	-3.1645	0.1123	0.0001	-0.0001	0.0119	-0.0325	0.0173	-0.0126	0.0084	-0.029	-0.0347	-0.0305	0.1454	0.0879	0.0699	-0.0286	-0.03	0.0543

Covariance matrix for class 1

	X_Minimum	X_Maximum	Y_Minimum	Y_Maximum	Pixels_Areas	X_Perimeter	Y_Perimeter	Sum_of_Luminosity	Minimum_of_Lumino	Maximum_of_Lumi	Length_of_Conveyer	TypeOfSteel_	TypeOfSteel_A400	Steel_Plate_Thickness	Edges_Index	Empty_Index	Square_Index	Outside_X_Ind	Edges_X_Ind	Edges_Y_Ind	Outside_Global	LogOfAreas	Log_X_Index	Log_Y_Index	Orientation_Index	Luminosity_In	SigmoidOfAre
X_Minimum	259850.0075	257093.9577	116322877.6	116319422.6	-108178.7691	-5394.43771	-4534.67636	-11143359.7	-951.823278	-1271.6454	11924.55184	29.6757	-29.675743	-2632.051308	7.80472	-4.6377	16.2541	-2.20775	3.98184	3.03354	-0.7494	-38.747	-17.697	-24.54	-7.9225987	-9.22488	-22.965
X_Maximum	257093.9577	258659.883	134056676.1	134053472.3	-37918.81391	-390.160962	-3021.65098	-3681264.498	-1138.2766	-1083.6103	11200.64775	32.6147	-32.61471	-2812.580666	8.0874	-3.2329	12.3406	0.937972	6.35496	-4.0957	-8.1544	-23.165	-2.3286	-22.975	-19.57775	-9.43195	-16.813
Y_Minimum	116322877.6	134056676.1	3.45005E+12	3.45005E+12	283390856.3	17647756.35	1458293.234	28989356104	-3455420.5	-360599.27	-4976242.019	156365	-156364.68	-36475421.25	37415	-19484	438.149	13024.92	65784.8	-27259	-80188	18896.9	67484.5	-64026	-129745.76	-16609.7	-23240
Y_Maximum	116319422.6	134053472.3	3.45005E+12	3.45004E+12	283450218.6	17650376.1	1461527.299	28995467841	-3455883.07	-360704.42	-4976981.311	156361	-156360.86	-36474927.9	37413.7	-19482	432.54	13025.11	65776.7	-27258	-80180	18916.7	67488.3	-64008	-129732.13	-16611.3	-23230
Pixels_Areas	-108178.7691	-37918.81391	283390856.3	283450218.6	3861950.94	156138.547	109980.3937	4012657718	-14200.6572	1670.51889	-24199.88807	1.35197	-1.3519696	4833.067072	-33.468	33.1659	-98.887	51.49464	-91.541	-92.01	33.4387	609.583	313.572	325.813	43.7033491	-30.4821	210.537
X_Perimeter	-5394.437705	-390.1609618	17647756.35	17650376.1	156138.547	8760.084488	4986.715559	16371117.53	-530.098482	155.680034	-1588.897656	-0.3978	0.39776214	291.9555338	-0.6003	3.70804	-7.5039	3.645861	-4.7654	-8.4057	-2.7025	34.5524	21.8805	16.0219	-3.7380221	-1.04302	14.6566
Y_Perimeter	-4534.676359	-3021.650976	1458293.234	1461527.299	109980.3937	4986.715559	4648.416976	11401566.26	-566.294623	-48.405513	-1080.508614	-3.8074	3.80739944	455.9438211	-1.532	2.80255	-6.9535	1.103055	-8.6399	-1.8636	6.6134	27.9726	9.79272	20.6883	11.0490339	-1.63647	12.6968
Sum_of_Luminosity	-11143359.7	-3681264.498	28989356104	28995467841	4012657718	16371117.53	11401566.26	41955304743	-1307025.16	320682.091	-2739601.778	-883.04	883.041173	410738.0965	-2999	3625.93	-10439	5468.395	-9306	-9947.7	2853.31	63118.1	33003.2	33381.4	3664.59415	-1983.36	21875.2
Minimum_of_Luminosity	-951.8232783	-1138.276598	-3455420.504	-3455883.071	-14200.65718	-530.098482	-566.294623	-1307025.159	783.1401496	372.07941	-961.1125523	-3.5608	3.560808	-214.8208485	1.26054	0.70863	0.62012	-0.1286	0.49108	-1.0029	-2.5134	-5.1218	-1.1253	-3.4568	-2.9065472	3.966339	-1.9926
Maximum_of_Luminosity	-1271.645381	-1083.610253	-360599.2656	-360704.4187	1670.518887	155.6800339	-48.4055126	320682.0907	372.0794096	453.343985	-429.1640567	-1.1333	1.13326632	-207.067559	0.70607	0.06723	-0.3989	0.132993	0.86146	-1.2098	-2.2414	-0.9315	1.14022	-1.977	-3.2126129	2.972325	-0.6765
Length_of_Conveyer	11924.55184	11200.64775	-4976242.019	-4976981.311	-24199.88807	-1588.89766	-1080.50861	-2739601.778	-961.112552	-429.16406	23315.68721	31.0875	-31.087453	1497.498075	-1.2592	-4.0573	3.0107	-0.75399	6.08813	2.09962	-0.7786	-7.8243	-2.4015	-8.5893	-5.0711032	-4.92738	-5.6304
TypeOfSteel_A300	29.67574293	32.61471008	156364.6793	156360.8628	1.351969557	-0.39776214	-3.80739944	-883.0411729	-3.560808	-1.1332663	31.08745295	0.23538	-0.2353776	0.95174169	0.0097	-0.0103	0.02039	0.001878	0.04098	-0.0045	-0.057	-0.0425	0.00926	-0.06	-0.077711	-0.01879	-0.0368
TypeOfSteel_A400	-29.67574293	-32.61471008	-156364.6793	-156360.8628	-1.351969557	0.397762142	3.807399444	883.0411729	3.560808002	1.13326632	-31.08745295	-0.2354	0.23537759	-0.95174169	-0.0097	0.01035	-0.0204	-0.00188	-0.041	0.00453	0.05699	0.04249	-0.0093	0.05996	0.07771097	0.018791	0.03683
Steel_Plate_Thickness	-2632.051308	-2812.580666	-36475421.25	-36474927.9	4833.067072	291.9555338	455.9438211	410738.0965	-214.820849	-207.06756	1497.498075	0.95174	-0.9517417	5125.563536	-1.5218	0.73824	-1.1786	-0.1359	-2.8403	1.98628	5.73755	3.13783	-1.324	4.97351	7.374493	-1.74604	1.86507
Edges_Index	7.8947208852	8.087398603	37414.99906	37413.68811	-33.46750604	-0.60031522	-1.53201386	-2998.960043	1.260536306	0.70607071	-1.259216121	0.0097	-0.0097002	-1.521782321	0.09148	-0.001	0.00872	0.000203	0.00459	-0.0035	-0.0161	-0.0091	0.0035	-0.0128	-0.0215925	0.006381	-0.0017
Empty_Index	-4.637730446	-3.23286886	-19483.89879	-19481.98446	33.16592548	3.708042027	2.802552218	3625.932952	0.708628129	0.0672281	-4.057277747	-0.0103	0.01034663	0.738238412	-0.001	0.01958	-0.0032	0.001026	-0.012	-0.01	-0.0083	0.02358	0.01979	0.02008	-0.0041598	0.002612	0.02223
Square_Index	16.25411704	12.3405555	438.1494723	432.5402423	-98.88709824	-7.5038786	-6.95354875	-10438.79882	0.620123515	-0.3989065	3.010699777	0.02039	-0.0203865	-1.178639973	0.00872	-0.0032	0.08082	-0.0028	0.02338	0.01384	-0.0156	-0.0528	-0.0184	-0.0363	-0.0241823	0.000779	-0.0284
Outside_X_Index	-2.207752115	0.937972439	13024.92363	13025.11041	51.49464345	3.645861078	1.103055338	5468.394678	-0.12859704	0.13299303	-0.753987351	0.00188	-0.0018783	-0.13589762	0.0002	0.00103	-0.0028	0.002298	0.00165	-0.0051	-0.0052	0.01123	0.01098	0.00123	-0.0082235	-0.00014	0.00439
Edges_X_Index	3.981843891	6.354862501	65784.75511	65776.6648	-91.54081341	-4.76539773	-8.63994904	-9306.013351	0.491079514	0.86145699	6.088133375	0.04098	-0.0409833	-2.840302618	0.00459	-0.012	0.02338	0.001651	0.06497	-0.0132	-0.0643	-0.0654	0.01067	-0.0849	-0.0995926	0.004356	-0.0445
Edges_Y_Index	3.033535936	-4.09565128	-27259.06639	-27257.83588	-92.01048076	-8.40572548	-1.86356806	-9947.677177	-1.00287196	-1.2097725	2.099622592	-0.0045	0.00452688	1.986277282	-0.0035	-0.01	0.01384	-0.00507	-0.0132	0.04651	0.06329	-0.0232	-0.0547	0.02315	0.08367077	-0.00762	-0.0157
Outside_Global_Index	-0.749381207	-8.154434716	-80188.26901	-80179.9869	33.43873475	-2.70250503	6.613401643	2853.310794	-2.51335554	-2.2414295	-0.778633591	-0.057	0.05698849	5.737549503	-0.0161	-0.0083	-0.0156	-0.00523	-0.0643	0.06329	0.22489	0.04151	-0.0743	0.10804	0.22563789	-0.01612	0.01807
LogOfAreas	-38.7466573	-23.16531419	18896.91936	18916.66407	609.5825765	34.55241658	27.97262199	6318.13488	-5.12180304	-0.931531	-7.824302724	-0.0425	0.0424888	3.137833784	-0.0091	0.02358	-0.0528	0.011227	-0.0654	-0.0232	0.04151	0.26397	0.11113	0.17299	0.06952865	-0.01913	0.14396
Log_X_Index	-17.69682655	-2.328622991	67484.52168	67488.25554	313.5723351	21.88046254	9.792724565	33003.21435	-1.1252859	1.14022351	-2.401521936	0.00926	-0.009256	-1.323991809	0.0035	0.01979	-0.0184	0.01098	0.01067	-0.0547	-0.0743	0.11113	0.13191	0.01578	-0.0994845	0.000483	0.06186
Log_Y_Index	-24.53986805	-22.97467847	-64025.94289	-64008.22416	325.8134608	16.02187314	20.68831366	33381.37968	-3.45678551	-1.9769766	-8.589345266	-0.06	0.05995859	4.973512072	-0.0128	0.02008	-0.0363	0.001231	-0.0849	0.02315	0.10804	0.17299	0.01578	0.17393	0.16410323	-0.01757	0.10039
Orientation_Index	-7.922598673	-19.57774969	-129745.7598	-129732.1326	43.70334905	-3.73802206	11.04903391	3664.594154	-2.90654723	-3.2126129	-5.07110393	-0.0777	0.07771097	7.374493001	-0.0216	-0.0042	-0.0242	-0.00822	-0.0996	0.08367	0.22564	0.06953	-0.0995	0.1641	0.29593986	-0.02062	0.03893
Luminosity_Index	-9.224806651	-9.431951641	-16609.74067	-16611.29051	-30.48207785	-1.04301705	-1.6364724	-1983.358862	3.966338811	2.972325626	-4.927378756	-0.0188	0.01879117	-1.746035577	0.00638	0.00261	0.00078	-0.00014	0.00436	-0.0076	-0.0161	-0.0191	0.00048	-0.0176	-0.0206199	0.026311	-0.0086
SigmoidOfAreas	-22.96515837	-16.81293666	-23239.50842	-23230.0017	210.5371973	14.65662215	12.69680272	21875.15614	-1.992605104	-0.6764549	-5.630427402	-0.0368	0.03682815	1.865068786	-0.0017	0.02223	-0.0284	0.00439	-0.0445	-0.0157	0.01807	0.14396	0.06186	0.10039	0.03893198	-0.00861	0.10045

Inferences:

The accuracy from Bayes classifier came is 93.99%.

This is because, when solving a problem Bayes directly focusses on finding similarity between observations, K-NN does better because of its inherent nature to optimize locally.

2. The diagonal elements of the covariance matrix denote the variance of the attribute with itself, that is, how much is the data spread across the median. From looking at the diagonal elements, we can infer the dispersion of the attribute and have an idea about the range of values in the attribute.

3. The off-diagonal elements indicate the covariance between the two attributes-how the attributes vary with respect to each other.

2 attributes with maximum covariance are Y_maximum and Sum_of_Luminosity for both classes.

4

Table 4 Comparison between classifiers based upon classification accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN	**0.896**
2.	KNN on normalized data	**0.970**
3.	Bayes	0.93992

Inferences:

1. KNN on normalized data- highest KNN- lowest

KNN
Bayes
KNN on normalized data

2. Ascending order

Figure 8: Covariance matrix for class 0

Reason=

KNN performs better when data is normalized because, the attributes on a bigger scale can no longer overpower and influence the results in their favor. This happens because Euclidean Distance is the total absolute distance along various axes and doesn't consider for the different ranges. The Bayes classifier directly focusses on finding similarity between observations, K-NN does better because of its inherent nature to optimize locally. Also in the above example which involves just 2 clusters, KNN will give more accurate predictions than Bayes..