



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: Ankit Pal Singh

Mobile No: 9149024234

Roll Number: B20181

Branch:EE

---

1 a.

	Prediction Outcome	
True Label	96	21
	38	181

Figure 1 KNN Confusion Matrix for K = 1

	Prediction Outcome	
True Label	96	21
	22	197

Figure 2 KNN Confusion Matrix for K = 3

	Prediction Outcome	
True Label	97	20
	19	200

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	82.440
3	87.202
4	88.392

#### Inferences:

1. The highest classification accuracy is obtained with K = 5 .
2. As the value of K increases the accuracy increases but this only happens in this case , otherwise increasing K doesn't mean increasing accuracy , the correct value of value of K is based on experiment.
3. Here with increasing value of K , the areas predicting each class will be more smoothed.
4. As the classification accuracy increases with the increase in value of K , the number of elements in the diagonal of confusion matrix are increasing ( which means the corrected predictions are increasing) .
5. The reason of increase in value of diagonal is (with increase in value of K , the predicted values are coming right ) .
6. As the classification accuracy increases with the increase in value of K the number of element in the off diagonal decreases ( which means the wrong predicted values decreases with increase in value of K).
7. The reason of decrease in value of off diagonal is (with increase in value of K , the predicted values are coming right ) .

2 a.

	Prediction Outcome	
True Label	110	7
	3	216

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

	Prediction Outcome	
True Label	110	7
	4	215

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

	Prediction Outcome	
True Label	109	8
	5	214

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	97.023
3	96.726
5	96.130

**Inferences:**

1. Normalization increases the Accuracy in this K nearest neighbor method.
2. The reason for the increase in accuracy is as normalization is a way of taking data that is slightly dissimilar but giving it a common state.
3. The highest classification accuracy is obtained with K = 1.
4. With increasing in value of K the accuracy decreases.
5. Since the data is normalized then with increasing in value of K , the test example have to find Euclidian distance with few number of training examples (because now the data is more scattered).
6. As the classification accuracy decreases with the increase in value of K , the number of elements in the diagonal decreases.
7. The reason of decrease in value of diagonal is (with increase in value of K , the predicted values are not coming right )
8. As the classification accuracy decreases with the increase in value of K infer , the number of values in the off diagonal increases ( as the predicted values are not coming right).

3

	Prediction Outcome	
True Label	106	11
	1	218

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 96.428 %.

## 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_Maximum	295.0	722.0
2	Y_Maximum	1662085.0	1525503.0
3	Pixels_Areas	7254.0	492.0
4	X_Perimeter	361.0	49.0
5	Y_Perimeter	211.0	44.0
6	Sum_of_Luminosity	807542.0	51893.0
7	Minimum_of_Luminosity	55.0	95.0
8	Maximum_of_Luminosity	138.0	131.0
9	Length_of_Conveyer	1385.0	1478.0
10	Steel_Plate_Thickness	40.0	100.0
11	Edges_Index	0.0	0.0
12	Empty_Index	0.0	0.0
13	Square_Index	1.0	1.0
14	Outside_X_Index	0.0	0.0
15	Edges_X_Index	1.0	1.0
16	Edges_Y_Index	1.0	1.0
17	Outside_Global_Index	0.0	1.0
18	LogOfAreas	4.0	2.0
19	Log_X_Index	2.0	1.0
20	Log_Y_Index	2.0	1.0
21	Orientation_Index	-0.0	0.0
22	Luminosity_Index	-0.0	0.0
23	SigmoidOfAreas	1.0	1.0



## 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
1		X_Maximu	Y_Maximu	Pixels_Are	X_Perimet	Y_Perimet	Sum_of_L	Minimum	Maximum	Length_of	Steel_Plat	Edges_Ind	Empty_In	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	63525.6	-9.4E+07	-385745	-18239.2	-9468.3	-4.2E+07	4505.14	2734.17	2535.96	209.385	32.2782	-11.1922	4.80594	-1.74619	25.7757	31.8496	33.0013	-97.0068	-60.523	-40.5102	32.0118	23.0782	-35.8983
3	Y_Maximu	-9.4E+07	2.4E+12	-3.7E+08	-1.8E+07	-1.7E+07	-5.6E+10	-7050234	-9205552	-9284928	-323758	-67104.4	17091.7	-78652.2	4539.63	-14281	-49590.3	-113977	233881	159608	76399.6	-129742	-69558.3	105631
4	Pixels_Are	-385745	-3.7E+08	2.8E+07	1363117	831873	3.3E+09	-142919	-14600.9	24505.4	-160.23	-520.991	341.847	474.706	244.253	-877.557	-679.37	21.295	2857.1	1514.76	1648.31	217.6	-230.446	613.37
5	X_Perimet	-18239.2	-1.8E+07	1363117	73074.5	44474.3	1.6E+08	-6579.71	-437.696	1827.95	0.10401	-24.8704	21.6197	31.1661	12.0002	-50.7771	-35.1256	13.9293	137.091	71.769	85.1562	21.3399	-9.0211	29.6859
6	Y_Perimet	-9468.3	-1.7E+07	831873	44474.3	27620.2	1E+08	-3835.09	-83.1602	1327.01	-5.66014	-13.7369	13.1118	21.0308	6.94005	-31.5517	-20.3094	13.5656	80.1725	40.5528	51.797	17.4743	-4.05004	16.9014
7	Sum_of_L	-4.2E+07	-5.6E+10	3.3E+09	1.6E+08	1E+08	4E+11	-1.6E+07	-1119251	2937983	-39805.9	-57947.4	40746.6	63307.9	27786.7	-106311	-77558.2	14203.8	325216	169104	192903	37655.5	-20351.7	67932.3
8	Minimum	4505.14	-7050234	-142919	-6579.71	-3835.09	-1.6E+07	1593.94	520.936	-130.138	-3.29256	4.83744	-2.08234	1.4108	-1.70221	4.25147	5.23363	5.02803	-24.1522	-14.1694	-11.5714	4.02566	5.12806	-7.46646
9	Maximum	2734.17	-9205552	-14600.9	-437.696	-83.1602	-1119251	520.936	423.305	-30.2199	-7.98	2.60708	-0.58488	2.32229	-0.4657	0.30472	2.15115	4.38427	-7.69569	-5.52523	-2.47242	4.51266	3.40163	-3.31066
10	Length_of	2535.96	-9284928	24505.4	1827.95	1327.01	2937983	-130.138	-30.2199	2534.46	41.3279	1.06952	0.37546	3.62209	-0.29954	-2.17186	0.34203	4.56817	0.58665	-1.09392	2.0947	4.83508	-0.70319	0.01739
11	Steel_Plat	209.385	-323758	-160.23	0.10401	-5.66014	-39805.9	-3.29256	-7.98	41.3279	6.8942	-0.02552	-0.02019	-0.00512	0.0075	0.01975	0.0432	0.06744	-0.05202	-0.04237	-0.01525	0.05675	-0.05927	0.01731
12	Edges_Ind	32.2782	-67104.4	-520.991	-24.8704	-13.7369	-57947.4	4.83744	2.60708	1.06952	-0.02552	0.03641	-0.01242	0.00725	-0.00708	0.01947	0.02851	0.02905	-0.09831	-0.06357	-0.0438	0.02871	0.02234	-0.03352
13	Empty_In	-11.1922	17091.7	341.847	21.6197	13.1118	40746.6	-2.08234	-0.58488	0.37546	-0.02019	-0.01242	0.01672	0.00138	0.00632	-0.01646	-0.01543	-0.00178	0.04859	0.03333	0.03103	-0.00266	-0.00545	0.01692
14	Square_In	4.80594	-78652.2	474.706	31.1661	21.0308	63307.9	1.4108	2.32229	3.62209	-0.00512	0.00725	0.00138	0.06724	-0.0057	-0.03662	0.00192	0.07254	-0.0054	-0.0272	0.02187	0.07328	0.01729	-0.01466
15	Outside_X	-1.74619	4539.63	244.253	12.0002	6.94005	27786.7	-1.70221	-0.4657	-0.29954	0.0075	-0.00708	0.00632	-0.0057	0.00606	-0.00136	-0.0084	-0.01148	0.03246	0.02444	0.0155	-0.01166	-0.00477	0.00844
16	Edges_X_I	25.7757	-14281	-877.557	-50.7771	-31.5517	-106311	4.25147	0.30472	-2.17186	0.01975	0.01947	-0.01646	-0.03662	-0.00136	0.06201	0.02735	-0.03614	-0.10078	-0.04031	-0.07097	-0.0429	0.00541	-0.02836
17	Edges_Y_I	31.8496	-49590.3	-679.37	-35.1256	-20.3094	-77558.2	5.23363	2.15115	0.34203	0.0432	0.02851	-0.01543	0.00192	-0.0084	0.02735	0.03417	0.02734	-0.10764	-0.0681	-0.05171	0.02396	0.0191	-0.03383
18	Outside_C	33.0013	-113977	21.295	13.9293	13.5656	14203.8	5.02803	4.38427	4.56817	0.06744	0.02905	-0.00178	0.07254	-0.01148	-0.03614	0.02734	0.21132	-0.07453	-0.0822	0.00686	0.13837	0.03471	-0.03305
19	LogOfArea	-97.0068	233881	2857.1	137.091	80.1725	325216	-24.1522	-7.69569	0.58665	-0.05202	-0.09831	0.04859	-0.0054	0.03246	-0.10078	-0.10764	-0.07453	0.46256	0.27006	0.23018	-0.06073	-0.07754	0.13901
20	Log_X_Inc	-60.523	159608	1514.76	71.769	40.5528	169104	-14.1694	-5.52523	-1.09392	-0.04237	-0.06357	0.03333	-0.0272	0.02444	-0.04031	-0.0681	-0.0822	0.27006	0.17553	0.12329	-0.07647	-0.05207	0.08497
21	Log_Y_Inc	-40.5102	76399.6	1648.31	85.1562	51.797	192903	-11.5714	-2.47242	2.0947	-0.01525	-0.0438	0.03103	0.02187	0.0155	-0.07097	-0.05171	0.00686	0.23018	0.12329	0.13246	0.01276	-0.02867	0.06619
22	Orientatio	32.0118	-129742	217.6	21.3399	17.4743	37655.5	4.02566	4.51266	4.83508	0.05675	0.02871	-0.00266	0.07328	-0.01166	-0.0429	0.02396	0.13837	-0.06073	-0.07647	0.01276	0.1321	0.03481	-0.02956
23	Luminosit	23.0782	-69558.3	-230.446	-9.0211	-4.05004	-20351.7	5.12806	3.40163	-0.70319	-0.05927	0.02234	-0.00545	0.01729	-0.00477	0.00541	0.0191	0.03471	-0.07754	-0.05207	-0.02867	0.03481	0.03212	-0.03071
24	SigmoidOf	-35.8983	105631	613.37	29.6859	16.9014	67932.3	-7.46646	-3.31066	0.01739	0.01731	-0.03352	0.01692	-0.01466	0.00844	-0.02836	-0.03383	-0.03305	0.13901	0.08497	0.06619	-0.02956	-0.03071	0.05396

#### Covariance matrix of 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
1		X_Maximu	Y_Maximu	Pixels_Are	X_Perimet	Y_Perimet	Sum_of_Li	Minimum	Maximum	Length_of	Steel_Plati	Edges_Ind	Empty_In	Square_In	Outside_X	Edges_X_I	Edges_Y_I	Outside_G	LogOfArea	Log_X_Inc	Log_Y_Inc	Orientatio	Luminosity
2	X_Maximu	259853.9	1.54E+08	-27426.5	-229.487	-2357.86	-2395001	-1207.68	-931.68	12611.36	-2436.96	8.382259	-3.57527	13.93903	0.381382	4.290362	-5.99641	-3.27726	-17.3693	0.606504	-20.1731	-19.502	-9.22165
3	Y_Maximu	1.54E+08	3.4E+12	55765350	5262775	-2823397	5.57E+09	-3149961	-509368	2678979	-3.8E+07	25560.44	-24519.6	19086.79	5590.926	71868.09	-13190.9	-83937.7	-13845.8	40660.76	-74057.9	-122854	-18833
4	Pixels_Are	-27426.5	55765350	1211324	62390.89	50346.4	1.24E+08	-9334.28	1374.996	-11922.5	5400.401	-26.9276	31.77679	-79.9993	21.38498	-78.8544	-40.2151	32.97901	418.0359	193.8794	252.4393	66.26756	-28.731
5	X_Perimet	-229.487	5262775	62390.89	4396.513	2985.714	6545245	-389.127	137.2422	-832.938	184.0284	-0.65555	3.49717	-6.40678	1.68097	-4.97609	-5.81406	-1.89442	26.3182	15.78253	13.76645	-1.14186	-1.14227
6	Y_Perimet	-2357.86	-2823397	50346.4	2985.714	3325.97	5148406	-468.934	-98.1263	-784.566	452.1501	-1.77074	2.984791	-6.72615	0.581885	-8.42231	-0.83591	6.825987	24.33245	7.518977	19.53359	11.88386	-1.85477
7	Sum_of_Li	-2395001	5.57E+09	1.24E+08	6545245	5148406	1.3E+10	-766220	323436.9	-1449008	453767	-2126.73	3483.123	-8367.35	2305.232	-7850.01	-4523.69	2716.129	42844.97	20477.93	25499.3	5810.551	-1579.9
8	Minimum	-1207.68	-3149961	-9334.28	-389.127	-468.934	-766220	829.4385	399.1023	-873.118	-201.135	1.460803	0.658852	0.724213	-0.07725	0.622366	-0.91479	-2.73962	-5.31643	-1.21295	-3.60124	-3.01797	4.223342
9	Maximum	-931.68	-509368	1374.996	137.2422	-98.1263	323436.9	399.1023	502.1509	-318.585	-215.417	0.802966	-0.03329	-0.35147	0.154719	1.096162	-1.1229	-2.3629	-1.32978	1.091224	-2.4034	-3.51317	3.148479
10	Length_of	12611.36	2678979	-11922.5	-832.938	-784.566	-1449008	-873.118	-318.585	23019.42	1519.033	-1.50811	-3.88023	2.230176	-0.45717	5.271178	0.518928	-1.11454	-5.73128	-0.80507	-8.01892	-5.69344	-4.14052
11	Steel_Plati	-2436.96	-3.8E+07	5400.401	184.0284	452.1501	453767	-201.135	-215.417	1519.033	5276.647	-1.76833	1.05515	-1.10763	-0.1603	-2.69282	1.727065	5.52655	2.687897	-1.3728	4.751397	6.926703	-1.74168
12	Edges_Ind	8.382259	25560.44	-26.9276	-0.65555	-1.77074	-2126.73	1.460803	0.802966	-1.50811	-1.76833	0.089811	-0.00128	0.011586	0.000189	0.005792	-0.00429	-0.01825	-0.01289	0.00258	-0.01576	-0.02418	0.007364
13	Empty_In	-3.57527	-24519.6	31.77679	3.49717	2.984791	3483.123	0.658852	-0.03329	-3.88023	1.05515	-0.00128	0.020889	-0.00309	0.000891	-0.01282	-0.00958	-0.00831	0.025825	0.020479	0.022935	-0.00236	0.002007
14	Square_In	13.93903	19086.79	-79.9993	-6.40678	-6.72615	-8367.35	0.724213	-0.35147	2.230176	-1.10763	0.011586	-0.00309	0.081369	-0.00195	0.021758	0.013203	-0.01504	-0.04884	-0.01675	-0.03398	-0.02333	0.001487
15	Outside_X	0.381382	5590.926	21.38498	1.68097	0.581885	2305.232	-0.07725	0.154719	-0.45717	-0.1603	0.000189	0.000891	-0.00195	0.001034	0.001078	-0.00362	-0.00419	0.007586	0.007691	0.000829	-0.00611	-6.06E-05
16	Edges_X_I	4.290362	71868.09	-78.8544	-4.97609	-8.42231	-7850.01	0.622366	1.096162	5.271178	-2.69282	0.005792	-0.01282	0.021758	0.001078	0.064472	-0.01116	-0.06207	-0.0696	0.007418	-0.08645	-0.09729	0.005673
17	Edges_Y_I	-5.99641	-13190.9	-40.2151	-5.81406	-0.83591	-4523.69	-0.91479	-1.1229	0.518928	1.727065	-0.00429	-0.00958	0.013203	-0.00362	-0.01116	0.044842	0.061284	-0.01834	-0.0497	0.023545	0.07895	-0.00653
18	Outside_G	-3.27726	-83937.7	32.97901	-1.89442	6.825987	2716.129	-2.73962	-2.3629	-1.11454	5.52655	-0.01825	-0.00831	-0.01504	-0.00419	-0.06207	0.061284	0.230278	0.047536	-0.07065	0.110535	0.225571	-0.017
19	LogOfArea	-17.3693	-13845.8	418.0359	26.3182	24.33245	42844.97	-5.31643	-1.32978	-5.73128	2.687897	-0.01289	0.025825	-0.04884	0.007586	-0.0696	-0.01834	0.047536	0.258917	0.102338	0.178986	0.082689	-0.02272
20	Log_X_Inc	0.606504	40660.76	193.8794	15.78253	7.518977	20477.93	-1.21295	1.091224	-0.80507	-1.3728	0.00258	0.020479	-0.01675	0.007691	0.007418	-0.0497	-0.07065	0.102338	0.104339	0.016798	-0.09037	-0.001
21	Log_Y_Inc	-20.1731	-74057.9	252.4393	13.76645	19.53359	25499.3	-3.60124	-2.4034	-0.81892	4.751397	-0.01576	0.022935	-0.03398	0.000829	-0.08645	0.023545	0.110535	0.178986	0.16798	0.181678	0.169697	-0.0202
22	Orientatio	-19.502	-122854	66.26756	-1.14186	11.88386	5810.551	-3.01797	-3.51317	-5.69344	6.926703	-0.02418	-0.00236	-0.02433	-0.00611	-0.09729	0.07895	0.225571	0.082689	-0.09037	0.169697	0.292764	-0.02167
23	Luminosity	-9.22165	-18833	-28.731	-1.14227	-1.85477	-1579.9	4.223342	3.148479	-4.14052	-1.74168	0.007364	0.002007	0.001487	-6.06E-05	0.005673	-0.00653	-0.017	-0.02272	-0.001	-0.0202	-0.02167	0.027847
24	SigmoidOf	-13.1262	-41669.2	172.6131	13.00211	12.17599	17804.05	-2.08887	-0.91256	-5.10768	1.680304	-0.00338	0.023984	-0.02688	0.003576	-0.04687	-0.01575	0.018737	0.147387	0.061641	0.105576	0.043959	-0.01052



**Inferences:**

1. The classification accuracy obtained from Bayes Classifier is 96.428 % , and it is more than KNN (without normalized data) but less than KNN ( with normalized data).
2. The values at diagonal of covariance matrix are maximum than all those at off diagonal places , as the diagonal elements of the covariance matrix contain the variances of the variables and the off diagonal elements contain the covariances between all possible pairs of variables.

4

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

S. No.	Classifier	Accuracy (in %)
1.	KNN	88.392
2.	KNN on normalized data	97.023
3.	Bayes	96.428

**Inferences:**

1. KNN model has highest accuracy of 97.023 when normalized data is given and KNN model has lowest accuracy of 88.392 when data given is unnormalized.
2. KNN ( without normalized data) < Bayes classifier < KNN ( with normalized data).