



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: Payal Sah

Mobile No: 6205720019

Roll Number: B20220

Branch:EE

---

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

IC 272: DATA SCIENCE - III  
LAB ASSIGNMENT – IV

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

	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	83.9285714
3	88.9880952
5	89.5833333

**Inferences:**

1. The highest classification accuracy is obtained with K =5.
2. Increasing the value of K increases the prediction accuracy.
3. As the value of K increases the number of neighbours also increases hence, prediction accuracy also increases with it.
4. As the classification accuracy increases with the increase in value of K the number of diagonal elements also increase.
5. With increase in value of k number of correctly predicted elements also increases.
6. As the classification accuracy increases with the increase in value of K the number of off-diagonal elements decrease and vice-versa.
7. Decrease in off-diagonal elements is due to more accurate prediction.

IC 272: DATA SCIENCE - III  
LAB ASSIGNMENT – IV

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

2 a.

	Prediction Outcome	
True Label	105	3
	7	221

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

	Prediction Outcome	
True Label	106	2
	4	224

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

	Prediction Outcome	
True Label	105	3
	3	225

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	96.13095
3	97.02385
5	96.72619

**Inferences:**

1. Data normalization increases classification accuracy.
2. Increase in classification accuracy after data normalization is due to cleaning of data with the help of normalization.
3. The highest classification accuracy is obtained with K =3.
4. Increasing the value of K increases the prediction accuracy.
5. As the value of K increases the number of neighbours also increases hence, prediction accuracy also increases with it.
6. As the classification accuracy increases with the increase in value of K the number of diagonal elements also increase.
7. Increase in classification accuracy indicates more accurate prediction, so there is increase in diagonal elements.
8. As the classification accuracy increases with the increase in value of the number of off-diagonal elements decrease.
9. With increase in classification accuracy true prediction increases, hence number of off-diagonal elements decreases.

3

	Prediction Outcome	
True Label	96	12
	2	226

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 95.833 %

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

### Inferences:

1. Bayes Classifier accuracy = 95.833% which is less than the previous applied classifiers. This is because, when solving a problem which 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.
2. The diagonal elements of the covariance matrix denote the variance of the attribute with itself, that is, how much the data is spread out from the mean. 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. For the given data, the attributes genergy and energy have the maximum covariance along the diagonal (for both the classes). This means that they have these attributes maximum spread and dispersion in their values. The attributes shift and ghazard have the minimum covariance which means that they have the least dispersion.
3. 3. The off-diagonal elements indicate the covariance between the two attributes-how the attributes vary with respect to each other. Larger the value of covariance between 2 attributes, greater is the joint variability of the two variables. 2 attributes with maximum covariance and joint variability: For class 0: maxenergy and energy, for class 1: genergy and energy. 2 attributes with minimum covariance and joint variability: For class 0: shift and ghazard, for class 1: shift and ghazard.

### 4

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

S. No.	Classifier	Accuracy (in %)
1.	KNN	89.5833
2.	KNN on normalized data	97.0238
3.	Bayes	95.8333

### Inferences:

1. Highest accuracy-97.02% and lowest accuracy-89.6%.
2. KNN on normalized data > Bayes > KNN.
3. KNN on normalized data has highest precision accuracy for that case the scale of each attribute is same and Bayes also increases the accuracy as compare to original KNN