Name: Anooshka Bajaj

1 a.

		Predicted Values	
		1	0
True Label	0	675	48
	1	47	6

Figure 1 KNN Confusion Matrix for K = 1

		Predicted Values	
		1	0
True Label	0	708	15
	1	51	2

Figure 2 KNN Confusion Matrix for K = 3

		Predicted Values	
		1	0
True Label	0	716	7
	1	52	1

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	87.7577
3	91.4948
5	92.3969

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 increase the accuracy of the model because when the number of neighbors are increased, more features are extracted from individual classes and more features are compared with the test data hence giving better results.
4. Diagonal elements represent the true positives. As the classification accuracy increases with the increase in value of K, the number of diagonal elements increase i.e. accuracy is improved.
5. When the accuracy increases with increase in value of K, the number of diagonal elements also increases because the accuracy of the model is directly proportional to the sum of number of Diagonal elements (total true positives).

6. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements decrease.
7. The off diagonal elements either represent the False Positives and the False Negatives. Thus, with increase in accuracy the count of these elements decreases.

2 a.

		Prediction Outcome	
		0	1
True Label	0	678	45
	1	49	4

Figure 6 KNN Confusion Matrix for K = 1 post data normalization

		Prediction Outcome	
		0	1
True Label	0	709	14
	1	50	3

Figure 7 KNN Confusion Matrix for K = 2 post data normalization

		Prediction Outcome	
		0	1
True Label	0	718	5
	1	52	1

Figure 8 KNN Confusion Matrix for K = 3 post data normalization

b.

Table 2 KNN Classification Accuracy for K = 1,3 and 5 post data normalization

K (Number of Neighbors)	Classification Accuracy (in %)
1	87.8865
3	91.7525
5	92.6546

Inferences:

1. Data normalization increases classification accuracy.
2. Increase in classification accuracy after data normalization because one attribute value not overshadowing the other
3. The highest classification accuracy is obtained with K =.5
4. Increasing the value of K increases the prediction accuracy.
5. Increasing the value of K increase the accuracy of the model because when the number of neighbors are increased, more features are extracted from individual classes and more features are compared with the test data hence giving better results.
6. As the classification accuracy increases with the increase in value of K, does the number of diagonal elements increase.
7. When the accuracy increases with increase in value of K, the number of diagonal elements also increases because the accuracy of the model is directly proportional to the sum of number of Diagonal elements (total true positives).
8. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements decrease.
9. The off diagonal elements either represent the False Positives and the False Negatives. Thus, with increase in accuracy the count of these elements decreases.

3

		Prediction Outcome	
		0	1
True Label	0	675	48
	1	38	15

Figure 11 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 88.9175%.

Table 3 Mean for Class 0

S. No.	Attribute Name	Mean
1.	seismic	1.3329
2.	seismoacoustic	1.4098
3.	shift	1.3737
4.	genergy	76427.581
5.	gpuls	502.933
6.	gdenergy	12.9284
7.	gdpuls	4.4092
8.	ghazard	1.1076
9.	energy	4726.256
10.	maxenergy	4107.096

Table 4 Mean for Class 1

S. No.	Attribute Name	Mean
1.	seismic	1.4957
2.	seismoacoustic	1.4444
3.	shift	1.1025
4.	genergy	189497.1
5.	gpuls	939.9230
6.	gdenergy	15.5726
7.	gdpuls	9.7435
8.	ghazard	1.0854
9.	energy	8809.829
10.	maxenergy	6850.854

Inferences:

1. The accuracy of Bayes Classifier is 88.9175%. This accuracy is low as compared to the KNN model of classification. This is because Bayes classifier depends upon the prior probability of classes which is not equal for all the classes and the model is biased towards a particular class.
2. Infer from covariance matrix the nature of values along the diagonal. State the reason.
3. Infer from off-diagonal elements the covariance between attributes. Write 2 pair of attributes with maximum and 2 pair of attributes with minimum covariance.

4

Table 7 Comparison between Classifier based upon Classification Accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN (Neighbors = 5)	92.3969
2.	KNN Normalized (Neighbors = 5)	92.6546
3.	Bayes Classifier	88.9175

Inferences:

1. Highest accuracy: KNN after normalization of data
   Lowest accuracy: Bayes Classifier
2. Bayes Classifier < KNN < KNN after normalization of data
3. The prediction accuracies are different for different classification models. For KNN with increasing value of K, accuracy increases. Bayes classifier is less accurate as it depends upon the prior probability of each class and which may vary. KNN after normalization of data gives the best result.

5 a.

		Prediction Outcome	
		0	1
True Label	0	711	12
	1	49	4

Figure 2 Bayes GMM Confusion Matrix for Q = 2

		Prediction Outcome	
		0	1
True Label	0	615	108
	1	30	23

Figure 2 Bayes GMM Confusion Matrix for Q = 4

		Prediction Outcome	
		0	1
True Label	0	719	4
	1	52	1

Figure 3 Bayes GMM Confusion Matrix for Q = 8

		Prediction Outcome	
		0	1
True Label	0	717	6
	1	53	0

Figure 4 Bayes GMM Confusion Matrix for Q = 16

b.

Table 2 Bayes GMM Classification Accuracy for Q = 2, 4, 8 & 16

Q	Classification Accuracy (in %)
2	**92.139**
4	**82.216**
8	**92.783**
16	**92.396**

Inferences:

1. The highest classification accuracy is obtained with Q = 8
2. As we go from Q = 2 to Q = 4, accuracy decreases. Then as Q = 8 we get the highest accuracy and at Q = 16 it almost remains same (decreases slightly).
3. By increasing Q each datapoint can be correctly associated with its correct cluster and thus the prediction accuracy increases.
4. As the classification accuracy increases with the increase in value of Q, the number of diagonal elements in Confusion matrix increase.
5. Diagonal elements represent the true positives and true negatives. As the classification accuracy increases with the increase in value of K, the number of diagonal elements increase i.e., accuracy is improved.
6. As the classification accuracy increases with the increase in value of Q, the number of off-diagonal elements decrease.
7. The off-diagonal elements represent the False Positives and the False Negatives. Thus, with increase in accuracy the count of these elements decreases.

6

Table 2 Comparison between Classifiers based upon Classification Accuracy

S. No.	Classifier	Accuracy (in %)
4.	KNN (Neighbors=5)	92.396
5.	KNN on normalized data (Neighbors=5)	92.654
6.	Bayes using unimodal Gaussian density	88.917
7.	Bayes using GMM	92.783

Inferences:

1. Highest accuracy: Bayes using GMM
   Lowest accuracy: Bayes using unimodal Gaussian density.
2. Classification Accuracy:
   Bayes Classifier using unimodal Gaussian density < KNN < KNN after normalization of data < Bayes using GMM
3. The above inferences imply that the data can be modeled best by using GMM. When unimodal Gaussian distribution is used, the accuracy is reduced. This shows that the data in each class has more than one clusters. KNN on Normalized data provides better results than using KNN without normalizing the data.