IC 272: DATA SCIENCE - III

LAB ASSIGNMENT – V

Data classification using Bayes classifier with Gaussian mixture model (GMM); regression using linear regression and polynomial curve fitting

Student's Name: **Mobile No:**

Roll Number: **Branch:**

PART - A

1 a.

		Prediction Outcome
True Label	100	200
	300	600

Figure 1 Bayes GMM Confusion Matrix for Q = 2

		Prediction Outcome
True Label	100	200
	300	600

Figure 2 Bayes GMM Confusion Matrix for Q = 4

		Prediction Outcome
True Label	100	200
	300	600

Figure 3 Bayes GMM Confusion Matrix for Q = 8

		Prediction Outcome
True Label	100	200
	300	600

Figure 4 Bayes GMM Confusion Matrix for Q = 16

b.

Table 1 Bayes GMM Classification Accuracy for Q = 2, 4, 8 & 16

Q	Classification Accuracy (in %)
2	
4	
8	
16	

Inferences:

1. The highest classification accuracy is obtained with Q =.
2. Infer whether increasing the value of Q increases/decreases the prediction accuracy.
3. State a suitable reason why increasing the value of Q increases/decreases the prediction accuracy.

4. As the classification accuracy increases/decreases with the increase in value of Q infer does the number of diagonal elements in the confusion matrix increase/decrease.
5. State the reason for the increase/decrease in diagonal elements.
6. As the classification accuracy increases/decreases with the increase in value of Q infer does the number of off-diagonal elements increase/decrease.
7. State the reason for increase/decrease in off-diagonal elements.
8. Inference 8 (You may add or delete the number of inferences).

   Note: Dummy values have been filled in the confusion matrices. Replace it with values obtained by you.

2

Table 2 Comparison between Classifiers based upon Classification Accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN	
2.	KNN on normalized data	
3.	Bayes using unimodal Gaussian density	
4.	Bayes using GMM	

Inferences:

1. Mention the classifiers with the highest and lowest accuracy.
2. Arrange the classifiers in ascending order of classification accuracy. Classifier a < Classifier b < Classifier c < Classifier d.
3. State the reasons behind Inference 1 and 2.
4. Any other inference (You may add or delete the number of inferences).

PART – B

1

a.

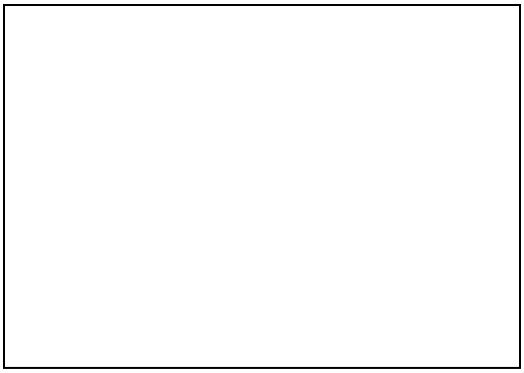

Figure 5 Univariate linear regression model: Rings vs. the chosen attribute name (replace) best fit line on the training data

Inferences:

1. The attribute with the highest correlation coefficient was used for predicting the target attribute Rings. Justify.
2. Does the best fit line fit the training data perfectly?
3. If not, why?
4. Infer upon bias and variance trade-off for the best fit line.

b.

Report the prediction accuracy on training data.

c.

Report the prediction accuracy on testing data.

Inferences:

1. Amongst training and testing accuracy, which is higher?
2. State reason for Inference 1.

d.

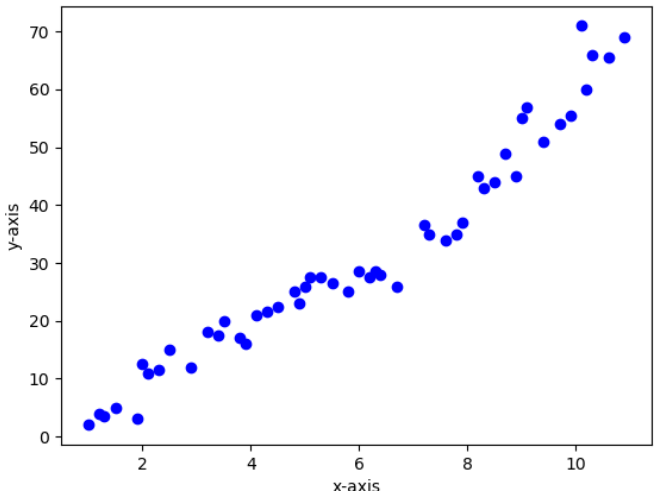

Figure 6 Univariate linear regression model: Scatter plot of predicted rings from linear regression model vs. actual rings on test data

Inferences:

1. Based upon the spread of the points, infer how accurate the predicted temperature is?
2. State the reason for Inference 1.

   Note: The above scatter plot is for illustration purposes only. Replace it with scatter plot obtained by you.

2

a.

Report the prediction accuracy on training data.

b.

Report the prediction accuracy on testing data.

Inferences:

3. Amongst training and testing accuracy, which is higher?
4. State reason for Inference 1.

c.

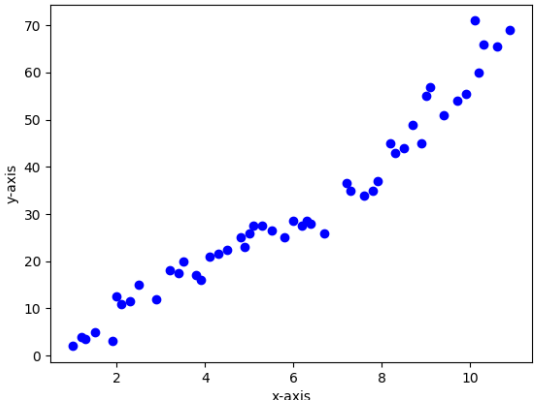

Figure 7 Multivariate linear regression model: Scatter plot of predicted rings from linear regression model vs. actual rings on test data

Inferences:

1. Based upon the spread of the points, infer how accurate the predicted temperature is?
2. State the reason for Inference 1.
3. Compare and contrast the performance of univariate linear with multivariate linear regression.
   Note: The above scatter plot is for illustration purposes only. Replace it with scatter plot obtained by you.

3

a.

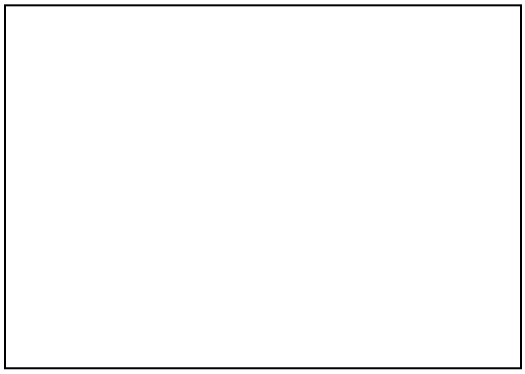

Figure 8 Univariate non-linear regression model: RMSE vs. different values of degree of polynomial (p = 2, 3, 4, 5) on the training data

Inferences:

1. Infer whether RMSE value decreases/ increases with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).
2. Is the increase/decrease uniform or after a certain p-value the increase/decrease becomes gradual?
3. State the reason for Inference 1 and 2.
4. From the RMSE value, infer which degree curve will approximate the data best.
5. Infer based upon bias and variance trade-off with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).

b.

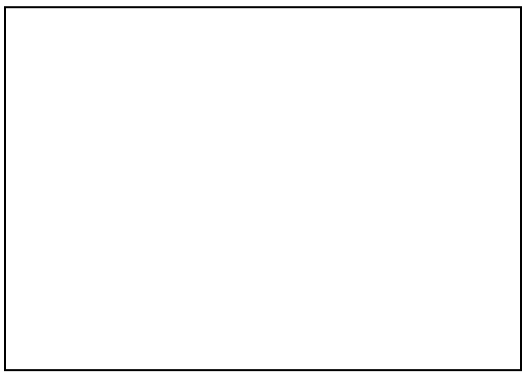

Figure 9 Univariate non-linear regression model: RMSE vs. different values of degree of polynomial (p = 2, 3, 4, 5) on the test data

Inferences:

1. Infer whether RMSE value decreases/ increases with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).
2. Is the increase/decrease uniform or after a certain p-value the increase/decrease becomes gradual.
3. State the reason for Inference 1 and 2.
4. From the RMSE value, infer which degree curve will approximate the data best.
5. Infer based upon bias and variance trade-off with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).

c.

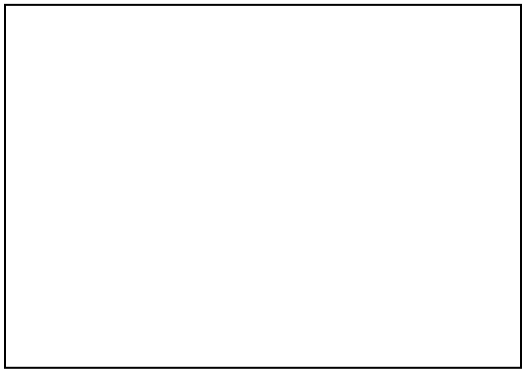

Figure 10 Univariate non-linear regression model: Rings vs. chosen attribute(replace) best fit curve using best fit model on the training data

Inferences:

1. State the p-value corresponding to the best fit model.
2. State the reason behind inference 1.
3. Infer based upon bias and variance trade-off with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).

d.

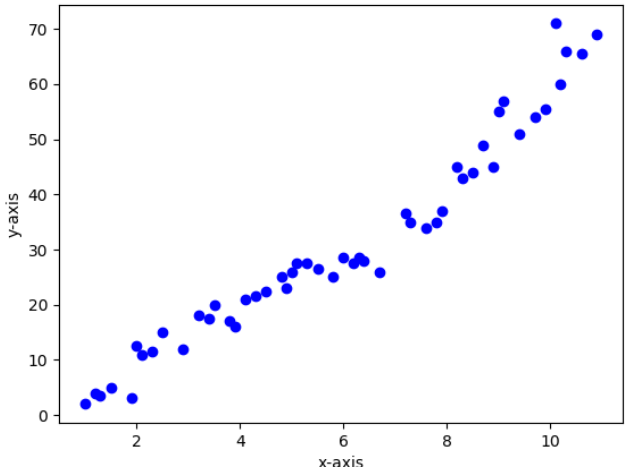

Figure 11 Univariate non-linear regression model: Scatter plot of predicted rings vs. actual rings on test data

Inferences:

1. Based upon the spread of the points, infer how accurate the predicted temperature is?
2. State the reason for Inference 1.
3. Compare and contrast univariate linear, multivariate linear and non-linear regression model based upon the accuracy of predicted temperature value and spread of data points in Scatter Plot
4. State the reason for Inference 3.
5. Inference based upon bias and variance trade-off between linear and non-linear regression models.
   Note: The above scatter plot is for illustration purposes only. Replace it with scatter plot obtained by you.

4

a.

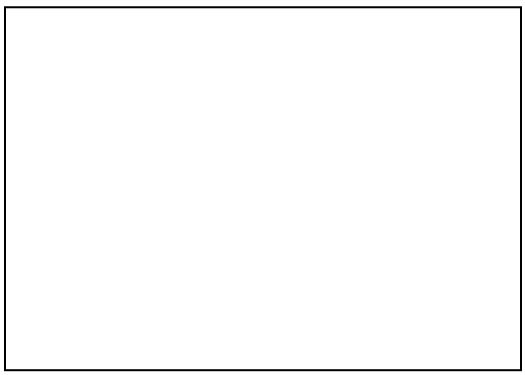

Figure 12 Multivariate non-linear regression model: RMSE vs. different values of degree of polynomial (p = 2, 3, 4, 5) on the training data

Inferences:

1. Infer whether RMSE value decreases/ increases with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).
2. Is the increase/decrease uniform or after a certain p-value the increase/decrease becomes gradual?
3. State the reason for Inference 1 and 2.
4. From the RMSE value, infer which degree curve will approximate the data best.
5. Infer based upon bias and variance trade-off with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).

b.

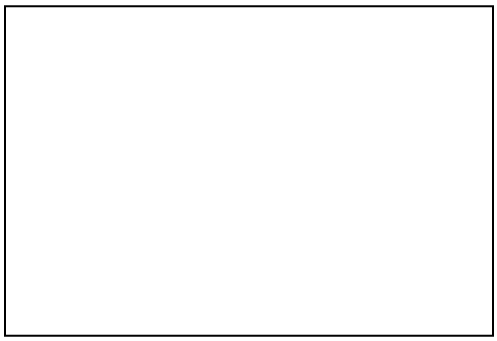

Figure 13 Multivariate non-linear regression model: RMSE vs. different values of degree of polynomial (p = 2, 3, 4, 5) on the test data

Inferences:

1. Infer whether RMSE value decreases/ increases with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).
2. Is the increase/decrease uniform or after a certain p-value the increase/decrease becomes gradual.
3. State the reason for Inference 1 and 2.
4. From the RMSE value, infer which degree curve will approximate the data best.
5. Infer based upon bias and variance trade-off with respect to the increase in the degree of the polynomial (p = 2, 3, 4, 5).

c.

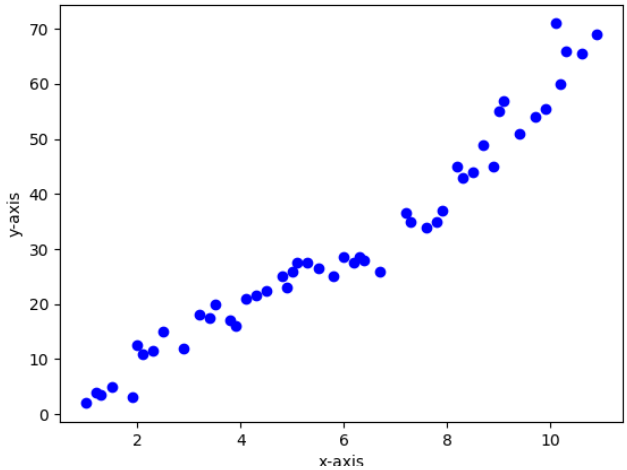

Figure 14 Multivariate non-linear regression model: Scatter plot of predicted rings vs. actual rings on test data

Inferences:

1. Based upon the spread of the points, infer how accurate the predicted temperature is?
2. State the reason for Inference 1.
3. Compare and contrast univariate linear, multivariate linear, univariate non-linear and multivariate non-linear regression model based upon the accuracy of predicted temperature value and spread of data points in Scatter Plot
4. State the reason for Inference 3.
5. Inference based upon bias and variance trade-off between linear and non-linear regression models.
   Note: The above scatter plot is for illustration purposes only. Replace it with scatter plot obtained by you.

Guidelines for Report (Delete this while you submit the report):

- **The plot/graph/figure/table should be centre justified with sequence number and caption.**
- **Inferences should be written as a numbered list.**
- **Use specific and technical terms to write inferences.**
- **Values observed/calculated should be rounded off to three decimal places.**
- **The quantities which have units should be written with units.**
- **Please fit a confusion matrix/ table in one page only.**