QPC: RJ18001123 AR - 18 Reg. No.

GIET MAIN CAMPUS AUTONOMOUS GUNUPUR – 765022
B. Tech Degree Examinations, June – 2021
(Sixth Semester)
BCSPC6020 / BITPC6020 – DATA ANALYTICS
(Common to CSE and I.T)

Time: 2 hrs Maximum: 50 Marks

Answer ALL Questions

The figures in the right hand margin indicate marks.

PART – A: (Multiple Choice Questions) **(1 x 10 = 10 Marks)**

Q.1. Answer ***ALL*** questions	[CO#]	[PO#]
a. Select the Inferential Statistics Techniques (i)Hypothesis Testing (ii)Confidence Interval (iii)ANOVA (iv)Range	3	3
b. Which of the following algorithms are supervised learning? (i) Linear Regression (ii) Decision Tree (iii) Clustering (iv) PCA	1	3
c. Which example(s) suits for regression techniques? (i) Cigarette consumption can be predicted based on smoking duration (ii) To classify the credit card transaction as fraud or Not fraud (iii) To classify the gender from hair length as male or female (iv) Sales can be predicted based on money spent for advertisement	2	2
d. Stepwise regression adds and removes predictors as needed for each step. It includes the _____ selection and _____ elimination regression techniques. (i) Forward (ii)Backward (iii) Feed forward (iv)Feed Backward	2	3
e. _____increases only if the new independent variable improves the model. _____always increases if add a new repressor to a model. (i) Adjusted R^2 (ii) R squared (iii)R (iv)MSE	2	2
f. Scatter plotsis used to detect_____. (i) non-linearity (ii) unequal error variances (iii)Outliers (iv)None of the above	1	3
g. Probabilities always range between _____ to _____. (i)0 (ii)1 (iii)100 (iv) 50	3	3
h. Find the correct statement about Naïve Bayes (i) Features are statistically independent of one another given the class value. (ii) Some features are dependent & Some features are independent (iii) Features are statistically dependent of one another given the class value (iv) Features are equally important	3	3
i. One of the very good methods to analyse the performance of Logistic Regression is AIC & BIC, which is similar to R-Squared in Linear Regression. Which of the following is true about AIC & BIC?	3	3

(i)Prefer a model with minimum AIC Value (ii)Prefer a model with maximum AIC Value

(iii)Prefer a model with maximum BIC Value (iv) Prefer a model with minimum BIC Value

j. Which of the following are numerical datatype 1 3

(i)Temperature (ii) Weight

(iii)Male (iv)Female

PART – B: (Short Answer Questions) **(2 x 5 = 10 Marks)**

Q.2. Answer *ALL* questions [CO#] [PO#]

a. The hourly wages of a sample of 130 system analysts are given below. mean = 60 range = 20 mode = 73 variance = 324 median = 74 .Calculated the coefficient of variation (in terms of percentage). 1 2

b. Define Filter methods & wrapper methods. 2 1

c. The given output is for ANOVA 3 2

```
Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  4114.6  4114.6  101.90  0.000
Residual Error  38  1534.4    40.4
Total           39  5649.0
```

Find the Standard Error of Estimate(SEE)

d. Calculate the value for accuracy and misclassification values for the given confusion matrix 3 2

	Predicted	
Actual	0	1
0	7	16
1	14	20

e. Apply the Max Pooling for the given Matrix 4 2

2	2	7	3
9	4	6	1
8	5	2	4
3	1	2	6

PART – C: (Long Answer Questions) **(6 x 5 = 30 Marks)**

Answer ***ANY FIVE*** questions Marks [CO#] [PO#]

3. Imagine a group of 200 applicants who took a Data Analytics test. George was among the test takers and he got 700 points (X) out of 1000. The average score was 600 (μ) and the standard deviation was 150 (σ). Now we would like to know how well George performed compared to his peers. Calculate the Z score and corresponding probability. (6) 1 2

4. Check the whether the data given in (1.c) contain outlier or not. If it contains outlier, remove the outlier and draw the boxplot else draw the box plot alone (6) 1 2

5. Reed Autos is Car Superstore, with over 500 vehicles in stock ready for same day drive away or to be delivered to your door anywhere in the India. They have a large variety of vehicles tailored to suit all budgets and needs. In the table number of TV Ads and number of Cars sold are given. Apply the simple linear regression & fine the regression equation. (6) 2 2

Number of TV Ads (x)	Number of Cars Sold (y)
1	14
3	24
2	18
1	17
3	27

6. In following result, which variable(s) is more significant and not significant and Justify your answer (based on the hypothesis statement) alpha =0.05 (6) 2 2

	coef	stderr	t	P>|t|	[0.025	0.975]
Intercept	3.5081	0.376	9.318	0.000	2.766	4.251
youtube	0.0457	0.001	32.564	0.000	0.043	0.048
facebook	0.1887	0.009	21.791	0.000	0.172	0.206
newspaper	-0.0007	0.006	-0.125	0.901	-0.012	0.011

7. Consider the following dataset given in a table that includes the information about the two predictors are X1 and X2 that could classify positively labeled data points and negatively labeled data points. Discover a simple SVM that accurately discriminates the two classes. Since the data is linearly separable, you can use a linear SVM (that is, one whose mapping function Φ() is the identity function) . First find out the three support vector and use the equation of a hyperplane y = w.x + b (w is the weight vector, x is the input and b is the bias, $\widetilde{w} = \sum_i \propto_i \widetilde{s_i}$) (6) 3 2

X1	X1	Class
1	1	Negative
2	1	Negative
1	-1	Negative
2	-1	Negative

4	0	Positive
5	1	Positive
5	-1	Positive
6	0	Positive

Write the step by step procedure build a Linear SVM.

Write the value for bias and weight vector b=? w=?

8. The rice bag weight is given in table, apply the one sample t–test. (6) 3 2

Note: Population(Hypothesized) mean is 25.

Rice_bag_weight
24.5
24.7
25.6
25
24.7
23.3
23.3

Find the t value. What is the degree of freedom for this sample?

9. Consider the following dataset given in a table that includes the information on a tasting score for a certain processed cheese. The two predictors are score for fat and salt indicating the relative presence of fat and salt in the particular cheese sample. The output variable is the cheese sample's consumer taste acceptance, where "1" indicates that a taste test panel likes the cheese and "0" that it does not like it. (6) 4 2

Initial

Fat_Score (x_1)	Salt_Score (x_2)	Acceptance (y)
0.4	0.5	1
0.3	0.8	1
0.2	0.9	1
0.1	0.1	0

Random weight weights: $w_0 = 0 w_1 = 1.2 w_2 = 0.6$ α (learning rate)=0.5

Activation function (Unit step function):

$$\hat{y} = f(y_{in}) = \begin{cases} 1 & if\, y_{in} > 0 \\ 0 & otherwise \end{cases}$$

Write the step by step procedure to train the neural network.

Write the final updated weights $w_0 = ? w_1 = ? w_2 = ?$

What is the maximum epoch required to build a trained neural network?

10.	Write about Vanishing Gradient Problem in Recurrent Neural Networks (RNN)	(6)	4	1

--- End of Paper ---