QPC: RN19BTECH585 AR 19 Reg. No

GIET UNIVERSITY, GUNUPUR – 765022

B. Tech (Seventh Semester – Regular) Examinations, November – 2022

BPECS7011 / BPECT7011 – STATISTICAL MACHINE LEARNING

(CSE & CST)

Time: 3 hrs Maximum: 70 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. The k-means algorithm is a	3	4
(i) Supervised learning algorithm (ii) Unsupervised learning algorithm (iii) Semi-supervised learning algorithm (iv) Weakly supervised learning algorithm		
b. When the number of features increase	2	2
(i) Computation time increases (ii) Model becomes complex (iii) Learning accuracy decreases (iv) All of the above		
c. We can define this probability as p(A|B) = p(A,B)/p(B) if p(B) > 0	3	3
(i) Conditional probability (ii) Marginal probability (iii) Bayes probability (iv) Normal probability		
d. Predicting whether a tumour is malignant or benign is an example of?	3	
(i) Unsupervised Learning (ii) Supervised Regression Problem (iii) Supervised Classification Problem (iv) Categorical Attribute		
e. Which machine learning models are trained to make a series of decision based on the rewards and feedback they receive for their actions ?	1	4
(i) Supervised (ii) Unsupervised (iii) Reinforcement (iv) All the above		
f. What is the most significant phase in a genetic algorithm?	4	4
(i) Selection (ii) Mutation (iii) Crossover (iv) Fitness function		
g. A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is Which of the following is true about SVM?	4	2
(i) It is useful only in high-dimensional spaces (ii) It requires less memory (iii) SVM does not perform well when we have a large data set (iv) SVM performs well when we have a large data set		

h. Logistic regression is a _____ regression technique that is used to model data having a _____ outcome — 3 — 2

(i) Linear , Binary (ii) Linear , numeric

(iii) Non-linear ,binary (v) Non-linear ,numeric

i. When you find many noises in data, which of the following options would you consider in kNN? — 3 — 4

(i) Increase the value of k (ii) Decrease the value of k

(iii) Noise does not depend on k (iv) k = 0

j. Lasso can be interpreted as least-squares linear regression where — 2 — 2

(i) weights are regularized with the l1 norm (ii) weights are regularized with the l2 norm

(iii) the solution algorithm is simpler (iv) None of these

PART – B: (Short Answer Questions) **(2 x 10 = 20 Marks)**

Q.2. Answer ***ALL*** questions	[CO#]	[PO#]
a. Explain the difference between bias and variance trade-offs,	1	2
b. Outline the role of Grid search in Machine Learning?	1	1
c. Define F-Test with a suitable example	2	2
d. When to stop tuning machine learning models.	1	2
e. How stochastic gradient descent works ?	4	3
f. Consider learning a classifier in a situation with 1000 features total. 50 of them are truly informative about class. Another 50 features are direct copies of the first 50 features. The final 900 features are not informative. Assume there is enough data to reliably assess how useful features are, and the feature selection methods are using good threshold How many features will be selected by mutual information filtering? How many features will be selected by a wrapper method?	4	3
g. Interpret the statement “To improve the Naive Bayes Text classifier using Laplace Smoothing?	3	2
h. What do you mean by the curse of dimensionality in machine learning?	3	2
i. Write down the steps in machine learning model development and deployment	1	6
j. What is ANOVA? Explain	2	2

PART – C: (Long Answer Questions) **(10 x 4 = 40 Marks)**

Answer ***ALL*** questions	Marks	[CO#]	[PO#]
3. a. How Does Linear Regression Work on the below-mentioned dataset? How is the error calculated in a linear regression model? Explain briefly the below-mentioned dataset.	5	2	3

X	2	4	6	8	3
Y	3	5	7	4	2

b. Write short note on : (i) F-Test and (ii) Chie square test and goodness of fit . — 5 — 2 — 2

(OR)

c. Find the regression line of Y on X and vice versa 5 2 2

X	1	3	4	8	9	11	14
Y	1	2	4	5	7	8	9

d. Find the coefficient correlation for the following data 5 3 2

X	14	16	17	18	19	20	21	23	23
Y	84	78	70	75	66	67	62	58	60

4. a. Exemplifying briefly the about Ridge and Lass regression. How do you calculate the penalty? 5 3 2

b. Explain briefly the Random Forest algorithm with a suitable example 5 3 2

(OR)

c. Compare Classification with a regression with an example. 5 3 2

d. Why is Random forest is better than logistic regression? Explain briefly 5 3 3

5. a. List and explain the various activation functions. Also, explain their suitability with respect to applications. 3 4 2

b. Apply the K-NN classification algorithm of the below dataset and predict the type of fruit or Food type to which Tomato(sweet =6 ,crunch=4) 7 3 3

Ingredients	Sweet	chrunch	Food type
Grape	8	5	FRUIT
Green bean	3	7	Vegetable
Nuts	3	6	Protein
Orange	7	3	FRUIT

(OR)

c. Explain briefly about the K-NN algorithm and generate the 3 neighbours of the following 5 3 3

Dataset. Check the class label of (Customer=>' JOHN' , AGE=>37,#Credit card=2 and Class=?)

Customer	Age	Income	No. of credit cards	Class
George	35	35K	3	No
Meorge	22	50K	2	Yes
Keorge	63	200k	1	No
Leorge	59	59	170k	No
				?

d.	Explain briefly the Naive Bayes classifier with a suitable example of the mobile dataset. I like this mobile It's a good mobile working nicely like it Reading on the mobile is bad But works good. Overall nice Nice looking but working slow Overall mobile is bad	5	3	4
6. a.	State the mathematical formulation of the SVM problem. Give an outline of the method for solving the problem.	5	4	
b.	Compare Feature Extraction and Feature Selection techniques. Explain how dimensionality can be reduced using a subset selection procedure.	5	3	
	(OR)			
c.	Summarizing the concept of back-propagation algorithm	3	4	2
d.	To assess the significance of possible variations in performance in a certain test between the convent schools of a city, a common test was given to a number of students taken at random from the fifth class of the 3 schools, and results were given below	7	3	5

A	B	C
9	13	14
11	12	13
13	10	17
9	5	17
8	5	9

Apply one-way ANOVA to analyze the variance of the given data

--- End of Paper ---