Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering

End Sem (Even) Examination May-2022

CS3ET07 Machine Learning

Programme: B.Tech. Branch/Specialisation: CSE

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

Q.1 i. A machine learning problem involves four attributes plus a class. The attributes have 3, 2, 2, and 2 possible values each. The class has three possible values. How many maximum possible different examples are there? **1**

(a) 12 (b) 24 (c) 48 (d) 72

ii. Supervised learning and unsupervised clustering both require at least one_____. **1**

(a) Hidden attribute (b) Output attribute

(c) Input attribute (d) Categorical attribute

iii. Regression modelling is a statistical framework for developing a mathematical equation that describes how: **1**

(a) One explanatory and one or more response variables are related

(b) Several explanatory and several response variables response are related

(c) One response and one or more explanatory variables are related

(d) All of these

iv. In regression analysis, the variable that is being predicted is the- **1**

(a) Response or dependent variable

(b) Independent variable

(c) Intervening variable

(d) Is usually x

v. In MLR, the square of the multiple correlation coefficient or R^2 is called the: **1**

(a) Coefficient of determination (b) Variance

(c) Covariance (d) Cross-product

P.T.O.

vi. A linear regression analysis produces the equation Y = -3.2X + 7. This indicates that: **1**
(a) A 1 unit increase in X results in a 3.2 unit decrease in Y
(b) A 1 unit decrease in X results in a 3.2 unit decrease in Y
(c) A 1 unit increase in X results in a 3.2 unit increase in Y
(d) An X value of 0 would increase Y by 7

vii. Interaction effects can be tested in multiple linear regression by using independent variables that represent: **1**
(a) Cross product between independent variables and dependent variables
(b) Cross product of independent variables
(c) Semi partial correlations squared
(d) None of these

viii. In multiple linear regression, the unique variance in the dependent variable explained by a particular independent variable is estimated by its: **1**
(a) Zero-order correlation squared
(b) Multiple correlation coefficient squared
(c) Semi-partial correlation squared
(d) None of these

ix. In _____ we use functions h that are not linear in the parameters. Often, such a function is derived from theory. In principle, there are unlimited possibilities for describing the deterministic part of the model. What should some in place of blank: **1**
(a) Linear Regression (b) Non-linear Regression
(c) Multiple Linear Regression (d) None of these

x. Nonlinear Least Squares (NLS) is a/an _____ technique that can be used to build regression models for data sets that contain nonlinear features. Models for such data sets are nonlinear in their coefficients. **1**
(a) Unsupervised (b) Orthogonal
(c) Optimization (d) None of these

Q.2 i. What are different types of Machine Learning? **4**
ii. Explain different learning models with example. **6**
OR iii. State some of the important model accuracy indicators. **6**

Q.3 i. Explain the difference between correlation and regression. **4**
ii. What are the basic assumptions of the Linear Regression Algorithm? **6**
OR iii. How do you interpret a linear regression model? **6**

Q.4 i. Explain multicollinearity and homoscedasticity. **4**
ii. What are the effects of homoscedasticity and autocorrelation in parameter estimation? Explain in detail. **6**
OR iii. Explain Least - Squares Estimation in detail. **6**

Q.5 i. What is the special case of orthogonal columns in X? **4**
ii. Construct the indicated confidence interval for the population mean using the t-distribution. Assume the population is normally distributed. c = 0.90, x = 13.1, s = 4.0, n = 5. **6**
OR iii. A poll is taken in which 329 out of 500 randomly selected voters indicated their preference for a certain candidate. **6**
(a) Find a 95% confidence interval for p.
(b) Find the margin of error for this 95% confidence interval for p.

Q.6 Write short note on any two:
i. Non-linear regression model **5**
ii. Statistical inference in non-linear regression models **5**
iii. Parameter estimation method **5**

******

Marking Scheme
CS3ET07 Machine Learning

Q.1 i. A machine learning problem involves four attributes plus a class. The attributes have 3, 2, 2, and 2 possible values each. The class has three possible values. How many maximum possible different examples are there? **1**

(d) 72

ii. Supervised learning and unsupervised clustering both require at least one_____. **1**

(a) Hidden attribute

iii. Regression modelling is a statistical framework for developing a mathematical equation that describes how: **1**

(c) One response and one or more explanatory variables are related

iv. In regression analysis, the variable that is being predicted is the- **1**

(a) Response or dependent variable

v. In MLR, the square of the multiple correlation coefficient or R^2 is called the: **1**

(a) Coefficient of determination

vi. A linear regression analysis produces the equation Y = -3.2X + 7. This indicates that: **1**

(a) A 1 unit increase in X results in a 3.2 unit decrease in Y

vii. Interaction effects can be tested in multiple linear regression by using independent variables that represent: **1**

(b) Cross product of independent variables

viii. In multiple linear regression, the unique variance in the dependent variable explained by a particular independent variable is estimated by its: **1**

(c) Semi-partial correlation squared

ix. In _____ we use functions h that are not linear in the parameters. Often, such a function is derived from theory. In principle, there are unlimited possibilities for describing the deterministic part of the model. What should some in place of blank: **1**

(b) Non-linear Regression

x. Nonlinear Least Squares (NLS) is a/an _____ technique that can be used to build regression models for data sets that contain nonlinear features. Models for such data sets are nonlinear in their coefficients. **1**

(c) Optimization

Q.2 i. Minimum 4 points types of Machine Learning **4**
1 mark for each (1 mark * 4)

ii. Learning models 3 marks **6**
Example 3 marks

OR iii. Important model accuracy indicators **6**
As per the explanation

Q.3 i. 4 points difference between correlation and regression **4**
1 mark for each (1 mark * 4)

ii. Assumptions of the Linear Regression Algorithm **6**
As per the explanation

OR iii. Interpret a linear regression model **6**
As per the explanation

Q.4 i. Multicollinearity 2 marks **4**
Homoscedasticity 2 marks

ii. Effects of homoscedasticity and autocorrelation 4 marks **6**
Explanation 2 marks

OR iii. Least - Squares Estimation **6**
As per the explanation

Q.5 i. Special case of orthogonal columns in X? **4**

ii. Construct the indicated confidence interval for the population mean using the t-distribution. **6**
As per the explanation

OR iii. (a) Find a 95% confidence interval for p. 3 marks **6**
(b) Find the margin of error for this 95% confidence interval for p. 3 marks

Q.6 Write short note on any two:

i. Non-linear regression model **5**
As per the explanation

ii. Statistical inference in non-linear regression models **5**
As per the explanation

iii. Parameter estimation method **5**
As per the explanation

******