ii. A multiple linear regression model with two independent variables was fitted to a sample of n = 50 observations. The resulting sum of squares for error (SSE) was found to be 300, and the degrees of freedom for error (dfE) was 47. What is the mean squared error (MSE) for the model? **5**

iii. What is the difference between simple linear regression and multiple linear regression? **5**

Q.5 Attempt any two:

i. What is the purpose of testing the general linear hypothesis in multiple linear regression analysis? Provide an example and explain how the F-test is used to test the hypothesis. **5**

ii. Explain the difference between testing the significance of regression and testing individual regression coefficients in multiple linear regression analysis. Provide an example and explain how the t-test is used to test individual coefficients. **5**

iii. What is the special case of orthogonal columns in X in multiple linear regression analysis? Provide an example and explain why this case is useful in regression analysis **5**

Q.6 Write short note on any two:

i. Non-Linear Least Square Method **5**

ii. Parameter Estimation Methods **5**

iii. Statistical Inference **5**

******

Total No. of Questions: 6 *Total No. of Printed Pages:4*

Enrollment No.....

Faculty of Engineering
End Sem Examination May-2023
CS3ET07 Machine Learning

Programme: B.Tech. Branch/Specialisation: CSE/All

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. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. What are the assumptions of linear regression? **1**

(a) Independence of observations, normality of residuals, homoscedasticity, linearity

(b) Independence of observations, normality of residuals, heteroscedasticity, linearity

(c) Dependence of observations, normality of residuals, homoscedasticity, linearity

(d) Independence of observations, non-normality of residuals, homoscedasticity, linearity

ii. In a probabilistic model, what is the role of the likelihood function? **1**

(a) To calculate the posterior probability distribution

(b) To calculate the prior probability distribution

(c) To calculate the conditional probability distribution

(d) None of these

iii. What is the purpose of the ANOVA test in regression analysis? **1**

(a) To test whether the regression coefficients are all equal to zero.

(b) To test whether the residual variances are equal across different groups.

(c) To test whether there is a significant linear relationship between the response variable and the predictor variable.

(d) To test whether the residuals are normally distributed.

iv. A linear regression model was fitted to a dataset of 20 observations. The estimated regression coefficients were: $\beta_0 = 10$ and $\beta_1 = 2$. The residual sum of squares was found to be 100. What is the estimated variance of the error term? **1**

(a) 2.5 (b) 5 (c) 10 (d) 20

v. What is the range of values for the correlation coefficient between two independent variables in a multiple linear regression model? **1**

(a) -1 to 1 (b) 0 to 1

(c) $-\infty$ to ∞ (d) -1 to 0

vi. Which of the following is NOT an effect of multicollinearity in multiple linear regression analysis? **1**

(a) Increase in standard errors of regression coefficients.

(b) Decrease in precision of parameter estimates.

(c) Inability to determine the true contribution of independent variables to the dependent variable.

(d) Increase in the accuracy of the regression model.

vii. Which test is used to test the general linear hypothesis in multiple linear regression analysis? **1**

(a) Wald test (b) F-test

(c) Chi-square test (d) t-test

viii. When is the special case of orthogonal columns in X applicable in multiple linear regression analysis? **1**

(a) When the independent variables are highly correlated with each other

(b) When the independent variables are not correlated with each other

(c) When the dependent variable is highly correlated with one of the independent variables

(d) When the dependent variable is not correlated with any of the independent variables

ix. Which of the following methods is used to estimate regression parameters in non-linear regression models? **1**

(a) Linear least squares method

(b) Non-linear least squares method

(c) Ordinary least squares method

(d) Weighted least squares method

x. What is the purpose of transforming a non-linear model into a linear model? **1**

(a) To simplify the model

(b) To make the model easier to interpret

(c) To apply linear regression methods

(d) To improve the accuracy of the model

Q.2 Attempt any two:

i. What is the difference between supervised, unsupervised, and reinforcement learning, and what are some examples of tasks that each type of learning can be used for? **5**

ii. What are some common methods for evaluating the performance of a machine learning model, and what are some advantages and disadvantages of each approach? **5**

iii. What are some key considerations to keep in mind when implementing a machine learning model in a real-world setting, and how can these challenges be addressed? **5**

Q.3 Attempt any two:

i. What are the key assumptions of parametric machine learning methods, and how can violations of these assumptions impact the accuracy of the model's predictions? **5**

ii. A simple linear regression model was fitted to a dataset of 25 observations. The estimated regression coefficients were: $\beta_0 = 2$ and $\beta_1 = 0.6$. The residual standard error was found to be 4. What is the predicted value of y for an observation with x=10, and what is the 95% prediction interval? **5**

iii. A simple linear regression model was fitted to a dataset of 30 observations. The estimated regression coefficients were: $\beta_0 = 2$ and $\beta_1 = 0.5$. The residual standard error was found to be 3.5. What is the value of the coefficient of determination (R^2)? **5**

Q.4 Attempt any two:

i. What is multicollinearity, and how does it affect the results of multiple linear regression? **5**

Marking Scheme

CS3ET07 [T]-Machine Learning

Q.1 i) What are the assumptions of linear regression? **1**

a) Independence of observations, normality of residuals, homoscedasticity, linearity

b) Independence of observations, normality of residuals, heteroscedasticity, linearity

c) Dependence of observations, normality of residuals, homoscedasticity, linearity

d) Independence of observations, non-normality of residuals, homoscedasticity, linearity

ii) In a probabilistic model, what is the role of the likelihood function? **1**

a) To calculate the posterior probability distribution

b) To calculate the prior probability distribution

c) To calculate the conditional probability distribution

d) None of the above

iii) What is the purpose of the ANOVA test in regression analysis? **1**

a) To test whether the regression coefficients are all equal to zero

b) To test whether the residual variances are equal across different groups

c) To test whether there is a significant linear relationship between the response variable and the predictor variable

d) To test whether the residuals are normally distributed

iv) A linear regression model was fitted to a dataset of 20 observations. The estimated regression coefficients were: $\beta_0 = 10$ and $\beta_1 = 2$. The residual sum of squares was found to be 100. What is the estimated variance of the error term? **1**

a) 2.5

b) 5

c) 10

d) 20

v) What is the range of values for the correlation coefficient between two independent variables in a multiple linear regression model? **1**

a) -1 to 1

b) 0 to 1

c) $-\infty$ to ∞

d) -1 to 0

vi) Which of the following is NOT an effect of multicollinearity in multiple linear regression analysis? **1**

a) Increase in standard errors of regression coefficients

b) Decrease in precision of parameter estimates

c) Inability to determine the true contribution of independent variables to the dependent variable

d) Increase in the accuracy of the regression model

vii) Which test is used to test the general linear hypothesis in multiple linear regression analysis? **1**

a) Wald test

b) F-test

c) Chi-square test

d) t-test

viii) When is the special case of orthogonal columns in X applicable in multiple linear regression analysis? **1**

a) When the independent variables are highly correlated with each other

b) When the independent variables are not correlated with each other

c) When the dependent variable is highly correlated with one of the independent variables

d) When the dependent variable is not correlated with any of the independent variables

ix) Which of the following methods is used to estimate regression parameters in non-linear regression models? **1**

a) Linear least squares method

b) Non-linear least squares method

c) Ordinary least squares method

d) Weighted least squares method

x) What is the purpose of transforming a non-linear model into a linear model? **1**

a) To simplify the model
b) To make the model easier to interpret
c) To apply linear regression methods
d) To improve the accuracy of the model

Q.2 Attempt any two:

i. What is the difference between supervised, unsupervised, and reinforcement learning, and what are some examples of tasks that each type of learning can be used for? **5**

ii. What are some common methods for evaluating the performance of a machine learning model, and what are some advantages and disadvantages of each approach? **5**

iii. What are some key considerations to keep in mind when implementing a machine learning model in a real-world setting, and how can these challenges be addressed? **5**

Q.3 Attempt any two:

i. What are the key assumptions of parametric machine learning methods, and how can violations of these assumptions impact the accuracy of the model's predictions? **5**

ii. A simple linear regression model was fitted to a dataset of 25 observations. The estimated regression coefficients were: $\beta_0 = 2$ and $\beta_1 = 0.6$. The residual standard error was found to be 4. What is the predicted value of y for an observation with x=10, and what is the 95% prediction interval? **5**

Predicted value = 8; Prediction interval = 4=-prof $\beta_0 + \beta_1 x$

iii. A simple linear regression model was fitted to a dataset of 30 observations. The estimated regression coefficients were: $\beta_0 = 2$ and $\beta_1 = 0.5$. The residual standard error was found to be 3.5. What is the value of the coefficient of determination (R^2)? **5**

Answer: 0.25

Q.4 Attempt any two:

i. A multiple linear regression model with two independent variables was fitted to a sample of n = 50 observations. The resulting sum of squares for error (SSE) was found to be 300, and the degrees of freedom for error (dfE) was 47. What is the mean squared error (MSE) for the model? **5**

1.

Solution: MSE = SSE / dfE = 300 / 47 = 6.383.

ii. What is multicollinearity, and how does it affect the results of multiple linear regression? **5**

iii. What is the difference between simple linear regression and multiple linear regression? **5**

Q.5 Attempt any two:

i. What is the purpose of testing the general linear hypothesis in multiple linear regression analysis? Provide an example and explain how the F-test is used to test the hypothesis. **5**

.

ii. Explain the difference between testing the significance of regression and testing individual regression coefficients in multiple linear regression analysis. Provide an example and explain how the t-test is used to test individual coefficients. **5**

iii. What is the special case of orthogonal columns in X in multiple linear regression analysis? Provide an example and explain why this case is useful in regression analysis **5**

Q.6		Write short note on any two:	
	i.	Non-Linear Least Square Method	**5**
	ii.	Parameter Estimation Methods	**5**
	iii.	Statistical Inference	**5**

******