|  |  |  |
| --- | --- | --- |
| **Supervised Learning - Regression** | | |
| Total Marks: 30 | | |
| **Section A** | | |
| 1. | In the given equation: Y = 4 + 0.56X  What do 4 and 0.56 indicate? | 2 marks |
| 2. | Which of the following can be considered a regression problem?   1. Which runner on a track team has the fastest recorded speed 2. How the speeds of runners on a track team relate to how long they each train 3. Which runner has been on a track team the longest 4. Which runner on a track team has the fastest average speed | 2 marks |
| 3. | What is the ROC curve? How does it help in evaluating a logictic regression model? | 2 marks |
| 4. | What are the measures used to assess the effectiveness of a logistic regression? | 2 marks |
| **Section B** | | |
| 5. | What is precision and recall? Why is there a tradeoff between the two? | 3 marks |
| 6. | Why is Maximum Likelihood Estimation used instead of OLS estimation for logistic regression? | 3 marks |
| 7. | Given the following logistic model:  Predicted logit (Y) = -6.3896 + 0.0266 (X1) - 0.0208 (X2) +1.0790 (X3)  What is the predicted probability of Y = 1 given that X1=10, X2=69 and X3=5? | 3 marks |
| 8. | Why is multicollinearity not desired in a regression model? How to identify it? | 3 marks |
| 9. | Why are dummy variables used in regression? How do you create dummy variables in Pandas? | 3 marks |
| **Section C** | | |
| 10. | Use the ‘Student performance’ dataset to answer the questions given below:   * Create a new column called ‘Exam Score’ which aggregates the math, reading and writing scores. Use this new column as a dependent variable and fit a linear regression model. (4 marks) * Calculate the R2, t-values, F value and Accuracy scores. What can you say about the variables and their relation ship with ‘exam scores’ based on the calculated metrics. (3 marks) | 7 marks |