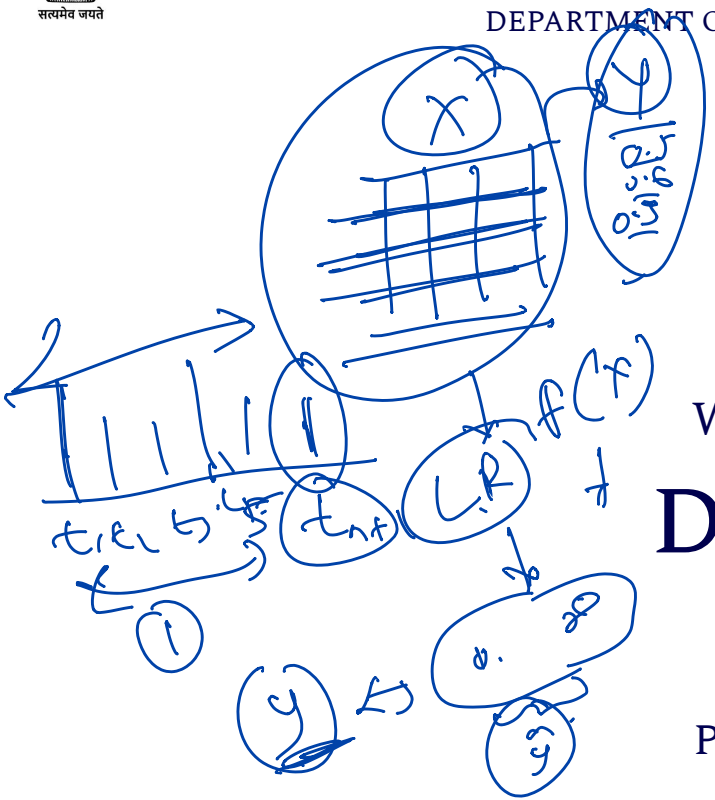
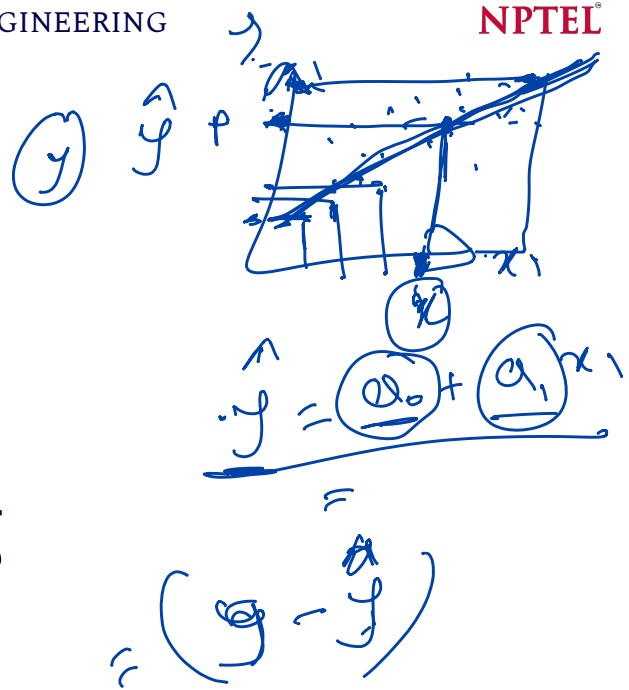




PMRF
Prime Minister's Research Fellowship

Week 8 - Live Session

Data Mining

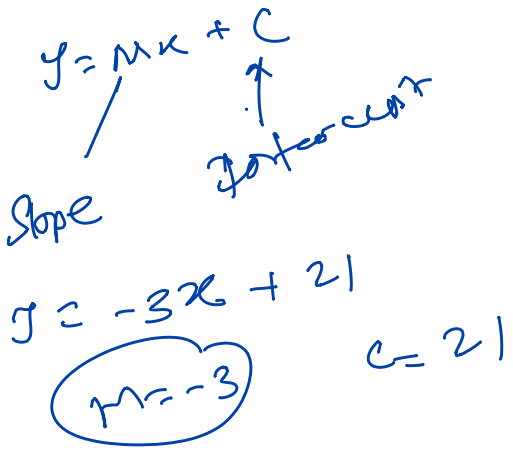
Swapnil S. Mane
mane.1@iitj.ac.in
PMRF Research Scholar

Q1. Regression is used in:

a) Predictive data mining
b) Exploratory data mining
c) Descriptive data mining
d) Explanative data mining

Q2. In the regression equation Y = 21 - 3X, the slope is

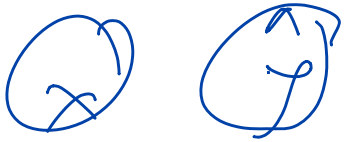
a) 21
b) -21
c) 3
d) -3



Q3. The output of a regression algorithm is usually a:

a) Real variable
b) Integer variable
c) Character variable
d) String variable

Q4. Regression finds out the model parameters which produces the least square error between -



a) input value and output value
b) input value and target value
c) output value and target value
d) model parameters and output value

Q5. The linear regression model $y = a_0 + a_1 x$ is applied to the data in the table shown below. What is the value of the sum squared error function $S(a_0, a_1)$, when $a_0 = 1$, $a_1 = 2$?

a) 0.0
b) 27
c) 13.5
d) 54

$$S(a_0, a_1) = \sum_{i=1}^{n} (y - (a_0 + a_1 x))^2$$

x	y
1	1
2	1
4	6
3	2

① $= (1 - (1+2))^2 = (1-3)^2 = 4$ +

② $+ (1 - (1+4))^2 = (1-5)^2 = 16$ +

③ $+ (6 - (1+8))^2 = (6-9)^2 = 9$ +

④ $+ (2 - (1+6))^2 = (2-7)^2 = 25$

$\Rightarrow (4 + 16 + 9 + 25) = 54$

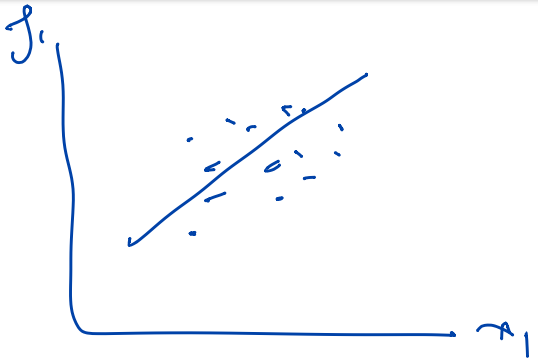
Q6. Consider x_1, x_2 to be the independent variables and y the dependent variable, which of the following represents a linear regression model?

a) $y = a_0 + a_1/x_1 + a_2/x_2$

b) $y = a_0 + a_1 x_1 + a_2 x_2$

c) $y = a_0 + a_1 x_1 + a_2 x^2$

d) $y = a_0 + a_1 x_1^2 + a_2 x_2$



$y = a_0 + a_1 x_1$

$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$

Q7. Find all the eigenvalues of the following matrix A.

a) 1,3
b) 2,3
c) 1,2,3
d) Eigenvalues cannot be found.

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 3 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Q8. In the figures below the training instances for classification problems are described by dots. The blue dotted lines indicate the actual functions and the red lines indicate the regression model. Which of the following statement is correct?

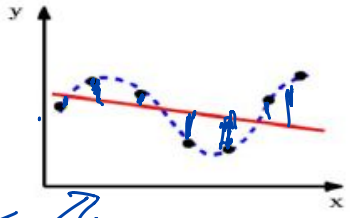


Figure 1

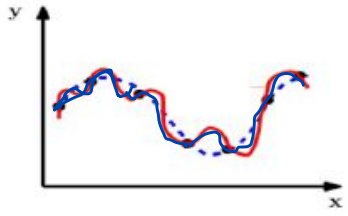


Figure 2

a) Figure 1 represents overfitting and Figure 2 represents underfitting

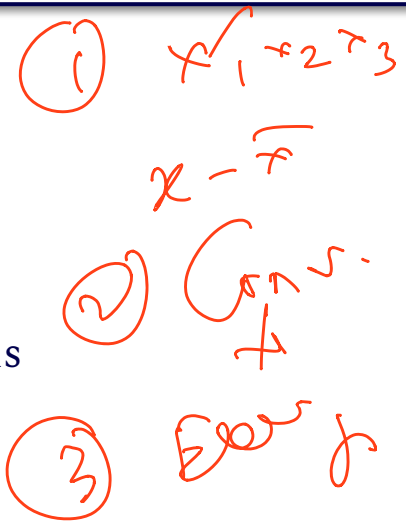
b) Figure 1 represents underrfitting and Figure 2 represents overfitting

c) Both Figure 1 and Figure 2 represents underfitting

d) Both Figure 1 and Figure 2 represents overfitting

Q9. In principal component analysis, the projected lower dimensional space corresponds to –

a) Subset of the original co-ordinate axis
b) Eigenvectors of the data covariance matrix
c) Eigenvectors of the data distance matrix
d) Orthogonal vectors to the original co-ordinate axis



Q10. A time series prediction problem is often best solved using?

a) Multivariate regression
b) Autoregression
c) Logistic regression
d) Sinusoidal regression