Question-2

Linear Regression

(a) Pseduo-inverse Explain what is pseduo-inverse of a matrix. Write the expression for pseduo-inverse to find Solutions to :

i. Underdetermined System of equations
ii. Over-determined System of equations.

A) Pseduo inverse (A^+) of a matrix A is generalization of inverse of a matrix.
It is the way to calculate the inverse for rectangular matrices.
Pseduo inverse gives 'best-fit' (least squares) solution to a system of linear equations that lacks unique solution.
Realtime data is not always square, but inverse exists only for square matrices generally.
Also, realtime data is not always consistent.
In these cases we can still find inverse of the matrix using pseduo inverse (A^+).

(i) pseduo inverse for underdetermined system of equations.

The system of equations in which number of equations (m) > no. of unknowns (n) is said to be underdetermined system.

$$A_{m\times n}\, x_{n\times 1} = b_{m\times 1} \qquad m>n$$

$$\boxed{A^+ = A^T \cdot (A A^T)^{-1}} \qquad \text{rank} = m.$$

(ii) pseduoinverse for overdetermined system of equations.

The system of equations in which no. of equations less than no. of unknowns is said to be overdetermined system

$$A_{m\times n}\, x_{n\times 1} = b_{m\times 1} \qquad m<n$$

$$\boxed{A^+ = (A^T A)^{-1} A^T}$$

Date :
Page no.

(b) Solve the following system of linear equations.

$$x_1 + 3x_2 = 17$$
$$5x_1 + 7x_2 = 19$$
$$11x_1 + 13x_2 = 23.$$

Sol:- No. of equations = $m = 3$

No. of unknowns = $n = 2$ (x_1, x_2)

$\because$ $m > n$ overdetermined system.

→ The above equations can be written in the form of matrix $Ax = b$.

Unknowns, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

$$\begin{bmatrix} 1 & 3 \\ 5 & 7 \\ 11 & 13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 17 \\ 19 \\ 23 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 3 \\ 5 & 7 \\ 11 & 13 \end{bmatrix} \quad b = \begin{bmatrix} 17 \\ 19 \\ 23 \end{bmatrix}.$$

→ Finding rank of matrix.

$$\begin{vmatrix} 1 & 3 \\ 5 & 7 \end{vmatrix}_{2\times2} = 7 - 15 = -8 \neq 0.$$

Rank$(A) = 2$ (Non-zero determinant of order 2×2)

→ pseduo-inverse of this matrix is given by $\boxed{A^+ = (A^TA)^{-1} \cdot A^T}$

$$A^T = \begin{bmatrix} 1 & 5 & 11 \\ 3 & 7 & 13 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 3 \\ 5 & 7 \\ 11 & 13 \end{bmatrix}.$$

$$A^TA = \begin{bmatrix} 147 & 181 \\ 181 & 227 \end{bmatrix}$$

→ Finding inverse of A^TA.

$$(A^TA)^{-1} = \frac{\operatorname{adj}(A^TA)}{|A^TA|}$$

$$(A^TA)^{-1} = \frac{1}{608}\begin{bmatrix} 227 & -181 \\ -181 & 147 \end{bmatrix}.$$

multiplying with matrix A^T

$$(A^TA)^{-1}.\, A^T = \frac{1}{608}\begin{bmatrix} 227 & -181 \\ -181 & 147 \end{bmatrix}\begin{bmatrix} 1 & 5 & 11 \\ 3 & 7 & 13 \end{bmatrix}.$$

$$A^{+} = \frac{1}{608}\begin{bmatrix} -316 & -132 & 144 \\ 260 & 124 & -80 \end{bmatrix}$$

$$A^{+} = \begin{bmatrix} -0.5197 & -0.2171 & 0.2368 \\ 0.4276 & 0.2039 & -0.1315 \end{bmatrix}$$

→ We have $Ax = b$

$$\tilde{x} = A^{+} b$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -0.5197 & -0.2171 & 0.2368 \\ 0.4276 & 0.2039 & -0.1315 \end{bmatrix}\begin{bmatrix} 17 \\ 19 \\ 23 \end{bmatrix}.$$

$$= \begin{bmatrix} -8.8349 - 4.124 + 5.446 \\ 7.269 + 3.874 - 3.025 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -7.5134 \\ 8.1165 \end{bmatrix}.$$

Date :
Page no.

(C) (i) Write the closed form expression (using normal equations) to solve a linear regression problem.

Sol:- For a linear Regression model, cost function is given by.

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2$$

$$X = \begin{bmatrix} -(x^{(1)})^T- \\ -(x^{(2)})^T- \\ \vdots \\ -(x^{(m)})^T- \end{bmatrix} \qquad X\theta = \begin{bmatrix} -(x^{(1)})^T- \\ -(x^{(2)})^T- \\ \vdots \\ -(x^{(m)})^T- \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} (x^{(1)})^T\theta \\ (x^{(2)})^T\theta \\ \vdots \\ (x^{(m)})^T\theta \end{bmatrix} = \underbrace{\begin{bmatrix} h_0(x^{(1)}) \\ h_1(x^{(2)}) \\ \vdots \\ h_\theta(x^{(m)}) \end{bmatrix}}_{X\theta}$$

predicted value of output.

$$\vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} = \text{Actual output.}$$

→ Cost function can be written as,

$$J(\theta) = \frac{1}{2}(X\theta - y)^T(X\theta - y).$$

$$\nabla_\theta J(\theta) = \frac{1}{2}\nabla_\theta (X\theta - y)^T(X\theta - y)$$

$$= \frac{1}{2}\nabla_\theta\left[\theta^T X^T - y^T\right]\left[X\theta - y\right]$$

$$= \frac{1}{2}\nabla_\theta\left[\theta^T X^T X\theta - \theta^T X^T y - y^T X\theta + y^T y\right]$$

$$= \frac{1}{2}\left[X^T X\theta + X^T X\theta - X^T y - X^T y\right].$$

$$\nabla_\theta J(\theta) = X^T X\theta - X^T y = 0.$$

$$X^T X\theta = X^T y. \qquad \text{"Normal equation".}$$

$$\boxed{\theta = (X^T X)^{-1} X^T y}$$

→ We can directly find the parameters by normal equation in single step.

Date :
Page no.

(C) (ii) Why do we prefer iterative methods like Gradient descent rather than using closed form solutions to solve a Linear Regression problem.

Sol:- → Gradient descent basically finds optimal parameters in small steps. This can be applied to any optimization problem.

→ For closed-form solution there is high Computational Complexity

→ Computational complexity for gradient descent is less compared to closed form solution.

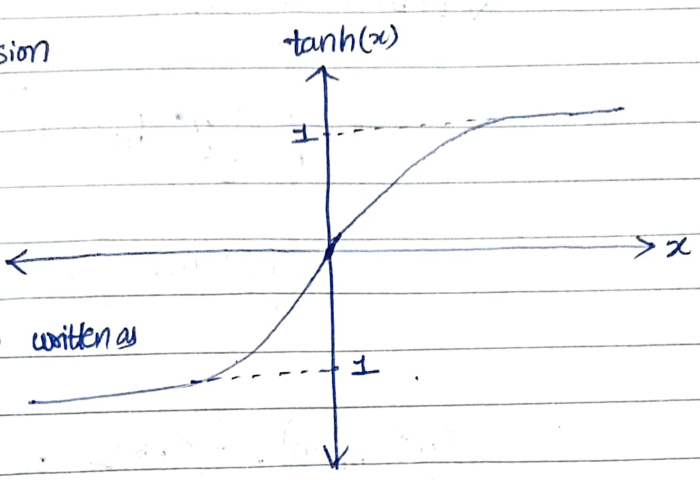
→ For the large data set the matrix dimensions increases and complexity increases in the calculations

$$\theta = (X^T X)^{-1} X^T . y$$

We need to perform lots of multiplications to reach the optimized parameters.

3.(c) Derive an expression for gradient descent update rule for logistic regression using 'tanh' function as decision boundary in place of 'Sigmoid' function.

Sol:- Logistic regression



$h_b(x) = \tanh(x)$

tanh(x) belongs to exponential family can be written as

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

$$h_b(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$$

From the definition of Gradient descent

$$b_{new} \leftarrow b_{old} - \eta \nabla L(b)$$

η – learning rate.

$L(b)$ = squared error

$$L(b) = [y^* - h_b(u)]^2.$$

$$h_b(u) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}.$$

$$\frac{\partial h_b(u)}{\partial x} = \frac{(e^{x} + e^{-x})(e^{x} + e^{-x}) - (e^{x} - e^{-x})(e^{x} - e^{-x})}{(e^{x} + e^{-x})^2}$$

$$= 1 - \left[\frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}\right]^2 \qquad \Big|\ \because \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}.$$

Date :
Page no.

$$\frac{\partial h_b(u)}{\partial x} = 1 - \tanh^2(x).$$

$$\nabla L(b) = \frac{\partial}{\partial b}\left[L(b)\right]$$

$$= \frac{\partial}{\partial b}\left[y^* - h_b(u)\right]^2$$

$$= 2\left[y^* - h_b(u)\right] \cdot \frac{\partial}{\partial b}\left[y^* - h_b(u)\right] \cdot \frac{\partial u}{\partial b}.$$

$u = xb.$

$$= 2\left[y^* - h_b(u)\right] \cdot \left[-\frac{\partial h_b(u)}{\partial b}\right] \cdot x$$

Substitute $h_b(u) = \tanh(x)$, $\frac{\partial h_b(u)}{\partial b} = 1 - \tanh^2(x)$

$$= 2\left[y^* - \tanh(x)\right] \cdot \left[-(1 - \tanh^2(x))\right] \cdot x$$

$$= 2x\left(y^* - \tanh(x)\right)\left(\tanh^2(x) - 1\right)$$

$$\nabla L(b) = 2x\left[y^* - \tanh(x)\right]\left[\tanh^2(x) - 1\right].$$

$$\boxed{b_{new} \leftarrow b_{old} - \eta\, 2x\left[y^* - \tanh(x)\right]\left[\tanh^2(x) - 1\right]}$$

Gradient descent update rule when function is 'tanh' in place of Sigmoid.