AIS11-ML
Tutorial - II

Q1) submitted earlier

Q2) Multivariate Linear Regression

Model: $Y = W^{(0)} + W^{(1)} x_1 + W^{(2)} x_2$

assuming $X = \begin{bmatrix} x_i \rightarrow \\ \vdots \\ \vdots \end{bmatrix}_{n \times 3}$ $\quad W = \begin{bmatrix} W^{(0)} \\ W^{(1)} \\ W^{(2)} \end{bmatrix}_{3 \times 1}$

n ↑ data, 3 ↑ feature

so, $\boxed{Y_{n \times 1} = XW}$ ——— ①

closed form :-

now, Loss function

$$J(W) = \sum_{i=1}^{n} \left(W_i^{(0)} + W_i^{(1)} x_1 + W^{(2)} x_2 - \hat{y} \right)^2$$

It can be represented in terms of matrix as,

$$J(W) = (XW - Y)^T (XW - Y) \quad (\because ①)$$
$$= (W^T X^T - Y^T)(XW - Y)$$
$$= W^T X^T X W - W^T X^T Y - Y^T X W + Y^T Y$$

deriving in terms of w

$$\nabla_w J(w) = \nabla_w (w^T x^T x w - w^T x^T y - y^T x w + y^T y)$$

$$= \nabla_w (w^T x^T x w) + \nabla_w (-w^T x^T y) - \nabla_w (y^T x w) + \nabla_w (y^T y)$$

$$= 2 x^T x w + \nabla_w (-w^T x^T y - y^T x w) + 0$$

(1)

$(\because \nabla_\theta \theta^T A \theta = 2A\theta$ where A is $\mathbb{R}^{n \times n}$

Here $w^T \underbrace{x^T x}_{A} w \Rightarrow 2 x^T x w)$

(Also $y^T x w$ is a scalar quantity, so

$$(y^T x w)^T = w^T x^T y = y^T x w \quad (\because [\text{scalar}]^T = \text{scalar})$$

(2)

$$= 2 x^T x w - \nabla_w (w^T x^T y + w^T x^T y) \quad (\because (2))$$

$$= 2 x^T x w - \nabla_w (2 w^T x^T y)$$

$$\boxed{= 2 x^T x w - 2 x^T y} \quad (3) \quad (\because \nabla_\theta \theta^T a = a \quad a \in \mathbb{R}^n)$$

eq (3) is $\nabla_W (J(W))$

equating it to 0.

$$\nabla_W J(W) = 0$$

$$2X^TXW - 2X^TY = 0$$

$$X^TXW = X^Ty$$

$$(X^TX)^{-1}(X^TX)W = (X^TX)^{-1}X^Ty$$

$$\boxed{W = (X^TX)^{-1}X^Ty} \quad \text{--- (4)}$$

eq (4) is closed form solution of multivariate Linear regression

⇒ Gradient descent:

for gradient descent also procedure till step (3) is same

$$\cancel{\nabla_\theta J} \quad \nabla_W J(W) = 2X^TXW - 2X^TY$$

$$\nabla_W J(W) = 2X^T(XW - Y) \quad \text{--- (5)}$$

by repeating & simultanious updating eq (5) till convergence we can find model params value by gradient descent.

Q3) ~~find~~ derive θ using MLE for

$$P(X=x) = \theta x_0^{\theta} x^{-(\theta+1)} \quad [x_0 \text{ is constant}]$$

$\Rightarrow$ Here, we define MLE function as

$$p(\theta|x) = \prod_{i=1}^{n} \theta x_0^{\theta} x_i^{-(\theta+1)}$$

$$= \theta^{n} x_0^{n\theta} \prod_{i=1}^{n} x_i^{-(\theta+1)}$$

to simplify, taking ln on both side

$$L = \ln(p(\theta|x)) = n\ln\theta + n\theta\ln x_0 - (\theta+1)\ln\left(\prod_{i=1}^{n} x_i\right)$$

to maximize likelihood, taking derivative w.r.t θ and equating to 0

$$\frac{\partial L}{\partial \theta} = \frac{n}{\theta} + n\ln x_0 - \sum_{i=1}^{n} \ln x_i = 0$$

$$\therefore \frac{n}{\theta} = \sum_{i=1}^{n} \ln x_i - n\ln x_0$$

$$\therefore \theta = \frac{n}{\sum_{i=1}^{n} \ln x_i - n\ln x_0}$$

MCQs

4. D
5. C
6. D, A
7. C