AVD613 - Machine Learning for Signal Processing

Submitted by : Vatsalya Gupta (SC19B098)

Problem - 2 :

We are given the dataset, hypothesis and cost function.

Dataset

i	$x^{(i)}$	$y^{(i)}$
1	1	3
2	-1	-2
3	2	4

Hypothesis : $h_{\theta_1}(x) = \theta_1 x$

Cost function :

$$J(\theta_1) = \frac{1}{2} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

We want to find $\left[\arg\min_{\theta_1} J(\theta_1)\right]$

$$\frac{d}{d\theta_1} J(\theta_1) = 0 \Rightarrow \frac{1}{2} \sum_{i=1}^{m} 2\left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x^{(i)} = 0$$

We know that, $h_\theta(x^{(i)}) = \theta_1 x^{(i)}$

$$\Rightarrow \left(h_\theta(x^{(i)}) - y^{(i)}\right) x^{(i)} = 0$$

$$\Rightarrow \left(\theta_1 x^{(i)} - y^{(i)}\right) x^{(i)} = 0$$

Now, $m = 3$

$$\Rightarrow \theta_1 - 3 + \theta_1 - 2 + 4\theta_1 - 8 = 0$$

$$\Rightarrow \boxed{\theta_1 = \frac{13}{6}}$$

Hence, this θ_1 minimises $J(\theta_1)$.

Problem - 3 :

From problem - 2 ,

$$\sum_{i=1}^{m} (\theta_1 x^{(i)} - y^{(i)}) x^{(i)} = 0$$

$$\Rightarrow \sum_{i=1}^{m} (\theta_1 (x^{(i)})^2 - y^{(i)} x^{(i)}) = 0$$

$$\Rightarrow \boxed{\theta_1 = \frac{\sum_{i=1}^{m} y^{(i)} x^{(i)}}{\sum_{i=1}^{m} (x^{(i)})^2}}$$

Substituting the given data for $m = 3$,

$$\theta_1 = \frac{1(3) - 1(-2) + 4(2)}{1 + 1 + 4} = \frac{13}{6}$$

This confirms the value of θ_1 obtained in problem - 2.

Problem - 4 :

We are given the hypothesis and cost function as

$$h_\theta(x) = \theta_0 + \theta_1 x \quad \text{and} \quad J(\theta_0, \theta_1) = \sum_{i=1}^{m} [h_\theta(x^{(i)}) - y^{(i)}]^2$$

We want to minimise J over θ_0 & θ_1. So,

$$\frac{\partial J}{\partial \theta_0} = 0 \quad \text{and} \quad \frac{\partial J}{\partial \theta_1} = 0$$

$$\Rightarrow \frac{\partial J}{\partial \theta_0} = 2 \sum_{i=1}^{m} [h_\theta(x^{(i)}) - y^{(i)}] \frac{\partial h_\theta(x^{(i)})}{\partial \theta_0} = 0$$

$$\Rightarrow \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)}) = 0$$

$$\Rightarrow m\theta_0 + \theta_1 \sum_{i=1}^{m} x^{(i)} = \sum_{i=1}^{m} y^{(i)} \quad \text{——— (1)}$$

Now,

$$\frac{\partial J}{\partial \theta_1} = 0 \Rightarrow \sum_{i=1}^{m} (\theta_0 + \theta_1 x^{(i)} - y^{(i)}) x^{(i)} = 0$$

$$\Rightarrow \theta_0 \sum_{i=1}^{m} x^{(i)} + \theta_1 \sum_{i=1}^{m} (x^{(i)})^2 = \sum_{i=1}^{m} y^{(i)} x^{(i)} \quad \text{——— (2)}$$

Substitute value of θ_1 from (2) into (1),

$$\theta_1 = \frac{\sum_{i=1}^{m} y^{(i)} - m\theta_0}{\sum_{i=1}^{m} x^{(i)}} \quad \text{——— (3)}$$

$$\Rightarrow \theta_0 \sum_{i=1}^{m} x^{(i)} + \sum_{i=1}^{m} (x^{(i)})^2 \left[\frac{\sum_{i=1}^{m} y^{(i)} - m\theta_0}{\sum_{i=1}^{m} x^{(i)}} \right] = \sum_{i=1}^{m} y^{(i)} x^{(i)}$$

$$\Rightarrow \boxed{\theta_0 = \frac{\sum_{i=1}^{m} x^{(i)} \sum_{i=1}^{m} y^{(i)} x^{(i)} - \sum_{i=1}^{m} (x^{(i)})^2 \sum_{i=1}^{m} y^{(i)}}{\left(\sum_{i=1}^{m} x^{(i)}\right)^2 - m \sum_{i=1}^{m} (x^{(i)})^2}}$$

To find θ_1, put θ_0 in (3),

$$\theta_1 = \frac{\sum_{i=1}^{m} y^{(i)} \left(\sum_{i=1}^{m} x^{(i)}\right)^2 - \sum_{i=1}^{m} y^{(i)} m \sum_{i=1}^{m} (x^{(i)})^2 + m \sum_{i=1}^{m} y^{(i)} \sum_{i=1}^{m} (x^{(i)})^2 - m \sum_{i=1}^{m} x^{(i)} \sum_{i=1}^{m} y^{(i)} x^{(i)}}{\sum_{i=1}^{m} x^{(i)} \left[\left(\sum_{i=1}^{m} x^{(i)}\right)^2 - m \sum_{i=1}^{m} (x^{(i)})^2 \right]}$$

$$\Rightarrow \boxed{\theta_1 = \frac{\sum_{i=1}^{m} y^{(i)} \sum_{i=1}^{m} x^{(i)} - m \sum_{i=1}^{m} y^{(i)} x^{(i)}}{\left(\sum_{i=1}^{m} x^{(i)}\right)^2 - m \sum_{i=1}^{m} (x^{(i)})^2}}$$

* (4a) Given $m = 2$, substituting,

$$\theta_1 = \frac{(y^{(1)} + y^{(2)})(x^{(1)} + x^{(2)}) - 2(x^{(1)} y^{(1)} + x^{(2)} y^{(2)})}{(x^{(1)} + x^{(2)})^2 - 2((x^{(1)})^2 + (x^{(2)})^2)}$$

$$\Rightarrow \theta_1 = \frac{(x^{(1)} - x^{(2)})(y^{(1)} - y^{(2)})}{(x^{(1)} - x^{(2)})^2} \Rightarrow \boxed{\theta_1 = \frac{y^{(1)} - y^{(2)}}{x^{(2)} - x^{(1)}}}$$

This is the expression for slope b/w 2 points.

* (4b) Now, for θ_0,

$$\theta_1 = \frac{\sum_{i=1}^{m} y^{(i)} - m\theta_0}{\sum_{i=1}^{m} x^{(i)}}$$

$$\Rightarrow \theta_0 = -\frac{(x^{(1)} + x^{(2)})}{2} \left[\frac{y^{(2)} - y^{(1)}}{x^{(2)} - x^{(1)}}\right] + \frac{y^{(2)} + y^{(1)}}{2}$$

$$\Rightarrow \boxed{\theta_0 = \frac{x^{(2)} y^{(1)} - x^{(1)} y^{(2)}}{x^{(2)} - x^{(1)}}}$$

* The number of additions & multiplications required for this method become very large and computationally expensive as m becomes large. Hence, gradient descent can give the solution much faster without much computations.

Problem - 1

(1) $\begin{bmatrix} 3 & -1 \\ 2 & 5 \\ -2 & 2 \end{bmatrix} \cdot \begin{bmatrix} u & a \\ v & b \end{bmatrix} = \begin{bmatrix} 3u-v & 3a-b \\ 2u+5v & 2a+5b \\ -2u+2v & -2a+2b \end{bmatrix}$

(2) Given, $A \in \mathbb{R}^{2\times 2}$ and $B \in \mathbb{R}^{2\times 4}$

The matrix product $A \cdot B$ exists since the number of columns of A is equal to number of rows of B.

$\Rightarrow \boxed{A \cdot B \in \mathbb{R}^{2\times 4}}$, size $= 2 \times 4$

(3) Given, $A \in \mathbb{R}^{3\times 5}$ and $B \in \mathbb{R}^{4\times 1}$

The matrix product $A \cdot B$ does not exist since number of columns of A is not equal to number of rows of B.

(4) $A \in \mathbb{R}^{3\times 2}$ and $y \in \mathbb{R}^3$ ie., $\mathbb{R}^{3\times 1}$

$y^T A$ will be of size $[\cdot]_{1\times 3} \times [\cdot]_{3\times 2} = [\cdot]_{1\times 2}$

$\Rightarrow y^T A$ is a row vector with size 1×2.

(5) $A \in \mathbb{R}^{3\times 2}$ and $x \in \mathbb{R}^2$ ie., $\mathbb{R}^{2\times 1}$

$Ax \Rightarrow [\cdot]_{3\times 2} \times [\cdot]_{2\times 1} = [\cdot]_{3\times 1}$

$\Rightarrow Ax$ is a column vector with size 3×1.

(6) Given $(Bx+y)^T A^T = 0$ ⑥

$A, B \in \mathbb{R}^{n\times n}$ and A, B are invertible ; $x, y \in \mathbb{R}^n$

$$(Bx+y)^T A^T = (x^T B^T A^T + y^T A^T) \in \mathbb{R}^{1\times n}$$
$$= 0$$

$$\Rightarrow x^T B^T A^T = -y^T A^T$$

multiply $(A^T)^{-1}$ on both sides,

$$x^T B^T A^T (A^T)^{-1} = -y^T A^T (A^T)^{-1}$$

$$\Rightarrow x^T B^T = -y^T$$

multiply $(B^T)^{-1}$ on both sides,

$$x^T B^T (B^T)^{-1} = -y^T (B^T)^{-1}$$

$$\Rightarrow x^T = -y^T (B^{-1})^T \quad \{\because (B^{-1})^T = (B^T)^{-1}\}$$

$$\Rightarrow x^T = (-B^{-1}y)^T$$

$$\Rightarrow (x^T)^T = ((-B^{-1}y)^T)^T \quad \{\because (x^T)^T = x\}$$

$$\Rightarrow \boxed{x = -B^{-1}y}$$