MATHEMATICS FOR MACHINE LEARNING - I

(1)

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N \times 1}$$

Given vector y, we can calculate $y_1^2 + \dots y_n^2$ as $y^T y$

$$\underbrace{y_1^2 + \dots y_N^2}_{1 \times 1} = \underset{(1 \times N)}{y^T} \; \underset{(N \times 1)}{y}$$

(2)

$$(AB)^T = B^T A^T$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

(3) For Scalar 'S'

$$S = S^T$$

(4) Scalar 'S'

Vector $\Theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix}$

Derivative of a scalar w.r.t. vector Θ

$$\frac{\partial S}{\partial \theta} = \begin{bmatrix} \frac{\partial S}{\partial \theta_1} \\ \vdots \\ \frac{\partial S}{\partial \theta_N} \end{bmatrix}$$

(5) $\dfrac{\partial A\theta}{\partial \theta}$ when $A\theta$ is a scalar
θ is a vector
A is a matrix

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1} \qquad A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}_{1\times 2}$$

$$A\theta = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} = A_1\theta_1 + A_2\theta_2$$

$$\frac{\partial A\theta}{\partial \theta} = \begin{bmatrix} \frac{\partial}{\partial \theta_1}(A_1\theta_1 + A_2\theta_2) \\ \frac{\partial}{\partial \theta_2}(A_1\theta_1 + A_2\theta_2) \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = A^T$$

(6) $\frac{\partial}{\partial \theta}(\theta^T Z \theta)$

$\theta^T Z \theta$ is a scalar
θ is vector
Z is of the form $c^T c$

$$x = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$x^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

$$Z = x^T x = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}$$

Z has property $Z_{ij} = Z_{ji} \Rightarrow Z^T = Z$

$$Z = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$\theta^T Z \theta = \underbrace{\begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}}_{1\times 2} \underbrace{\begin{bmatrix} a & b \\ b & c \end{bmatrix}}_{2\times 2} \underbrace{\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}}_{2\times 1}$$

$$= \underbrace{\begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}}_{1\times 2} \underbrace{\begin{bmatrix} a\theta_1 + b\theta_2 \\ b\theta_1 + c\theta_2 \end{bmatrix}}_{2\times 1} = a\theta_1^2 + b\theta_1\theta_2 + b\theta_1\theta_2 + c\theta_2^2$$

$$\frac{\partial \theta^T Z \theta}{\partial \theta} = \frac{\partial}{\partial \theta}\left[a\theta_1^2 + 2b\theta_1\theta_2 + c\theta_2^2\right] = \begin{bmatrix} \frac{\partial}{\partial \theta_1}\left(a\theta_1^2 + 2\theta_1\theta_2 b + c\theta_2^2\right) \\ \frac{\partial}{\partial \theta_2}\left(a\theta_1^2 + 2\theta_1\theta_2 b + c\theta_2^2\right) \end{bmatrix}$$

$$= \begin{bmatrix} 2a\theta_1 + 2b\theta_2 \\ 2b\theta_2 + 2c\theta_2 \end{bmatrix} = 2\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$= 2z\theta$$

$$\therefore \boxed{\frac{\partial}{\partial \theta}\left(\theta^T z \theta\right) = 2z\theta = 2z^T\theta}$$