Derivatives / Gradients used for least squares estimation for linear regression

(1) $f(\vec{w}) = \vec{a}^T \vec{w}$, where $\vec{w} = [w_1 \ldots\ldots w_n]^T$
and
$\vec{a} = [a_1 \ldots\ldots a_n]^T$

$$\nabla_{\vec{w}} f = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{bmatrix}$$

Now, $f(\vec{w}) = \vec{a}^T \vec{w} = a_1 w_1 + a_2 w_2 + \cdots + a_n w_n$

$$\therefore \nabla_{\vec{w}} f = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \vec{a}$$

$$\boxed{\therefore \nabla_{\vec{w}} f = \vec{a} \quad \text{when } f(\vec{w}) = \vec{a}^T \vec{w}}$$

(2) $\vec{w}^T A \vec{w}$:

$$f(\vec{w}) = \vec{w}^T A \vec{w} \quad , \quad \text{where } \vec{w} \text{ is an } n \times 1 \text{ vector}$$
$$\vec{A} \text{ is an } n \times n \text{ matrix}$$

Now,

$$f(\vec{w}) = \vec{w}^T A \vec{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} \times \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & & & \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \times \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}}_{1 \times n} \times \underbrace{\begin{bmatrix} A_{11}w_1 + A_{12}w_2 + \cdots + A_{1n}w_n \\ A_{21}w_1 + A_{22}w_2 + \cdots + A_{2n}w_n \\ \vdots \\ A_{n1}w_1 + A_{n2}w_2 + \cdots + A_{nn}w_n \end{bmatrix}}_{n \times 1}$$

$$= (A_{11}w_1 + A_{12}w_2 + \cdots + A_{1n}w_n)w_1$$
$$+ (A_{21}w_1 + A_{22}w_2 + \cdots + A_{2n}w_n)w_2$$
$$\vdots$$
$$+ (A_{n1}w_1 + A_{n2}w_2 + \cdots + A_{nn}w_n)w_n$$

Now,

$$\nabla_{\vec{w}} f = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{bmatrix}$$

$$\therefore \nabla_{\vec{w}} f = \begin{bmatrix} (A_{11}w_1 + A_{12}w_2 + \dots + A_{1n}w_n) + A_{11}w_1 + A_{21}w_2 + \dots + A_{n1}w_1 \\ (A_{21}w_1 + A_{22}w_2 + \dots + A_{2n}w_n) + A_{12}w_1 + A_{22}w_2 + \dots + A_{n2}w_n \\ \vdots \\ (A_{n1}w_1 + A_{n2}w_2 + \dots + A_{nn}w_n) + A_{1n}w_1 + A_{2n}w_2 + \dots + A_{nn}w_n \end{bmatrix}$$

$$= \begin{bmatrix} A_{11}w_1 + A_{12}w_2 + \dots + A_{1n}w_n \\ \vdots \\ A_{n1}w_1 + A_{n2}w_2 + \dots + A_{nn}w_n \end{bmatrix} + \begin{bmatrix} A_{11}w_1 + A_{21}w_2 + \dots + A_{n1}w_n \\ \vdots \\ A_{1n}w_1 + A_{2n}w_2 + \dots + A_{nn}w_n \end{bmatrix}$$

$$= A\vec{w} + A^T\vec{w}$$

$$= (A + A^T)\vec{w}$$

$$\boxed{\therefore \nabla_{\vec{w}} f = (A + A^T)\vec{w} \quad \text{when } f(\vec{w}) = \vec{w}^T A w}$$

Note: when A is symmetric, $\nabla_{\vec{w}} f = 2A\vec{w}$