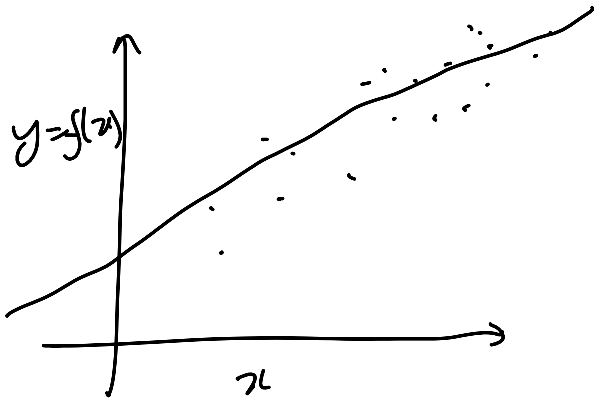
Linear Regression

input

$$y = m\,x + c$$

output parameters



Suppose you are given **Dataset** (Training Dataset)

$$D = \{(x_1,y_1),(x_2,y_2),\ldots,\underset{\uparrow}{(x_i,y_i)}\ldots(x_n,y_n)\}$$

$$x_i \xrightarrow{?} y_i$$

Guess : Model

$$\hat{y}_i(x_i; m,c) = \hat{y}_i(\underline{x}_i) = \underline{m}\,x_i + \underline{c} \qquad \forall i \in \{1,\ldots,n\}$$

Loss function : **Mean** squared Error

$$l(x_i,y_i) = |y_i - \hat{y}_i(x_i)|^2$$

(Penalty for making a wrong guess)

$$L(D) = \frac{1}{n}\sum_{(x_i,y_i)\in D} l(x_i,y_i)$$

Avg. ~~Total~~ penalty over the entire Dataset

inputs parameters

$$L(D; m,c) = \frac{1}{n}\sum_{(x_i,y_i)\in D} l(x_i,y_i; m,c)$$

① Model / Guess

② Loss | Penalty

③ Training

minimize the Loss function $\underbrace{\text{w.r.t.}}_{\text{with respect to}}$ the parameters

$$\overbrace{\underbrace{m^*, c^*}_{\text{optimal set of parameters}}}^{\text{Trained weights}} = \underset{(m,c)}{\text{minimize}} \; \mathcal{L}(D; m, c)$$

$$\longrightarrow \hat{y}_i(x_i; [m^*, c^*]) \leftarrow \text{Trained model}$$

$x \notin D$

$$\hat{y}(x; m^*, c^*)$$

$\hat{y}(x_i ; m, c) = mx + \underline{c}$ ← Affine ~~Linear~~ function $\underline{\underline{\text{in } x}}$

Is this function Linear/Affine in m and x?

Def. of Linear funct
$$\begin{cases} f(x+y) = f(x) + f(y) \\ f(\alpha x) = \alpha f(x) \end{cases}$$

$\underline{\underline{\text{No}}}$

$$z = f(x,y) = \underbrace{\overbrace{xy}}_{} \neq \text{Line}$$

× Quadratic term x^2

× Cubic term x^3

× exp(), log(), sin(), cos()

$$\hat{y}(x_i ; \{m, c\}) = \underbrace{[m \quad c]_{1\times 2}}_{\substack{\text{coeff} \\ \text{row vector}}} \underbrace{\begin{bmatrix} x \\ 1 \end{bmatrix}_{2\times 1}}_{\substack{\text{variables / inputs} \\ \text{column vector}}}$$

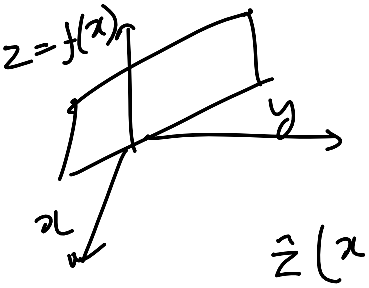
$$= \underline{w}^T \underline{x}$$

$\underline{w} = \begin{bmatrix} m \\ c \end{bmatrix}$

$\underline{x} = \begin{bmatrix} x \\ 1 \end{bmatrix}$

$\underline{w} \cdot \underline{x} = mx + c \cdot 1$

$= mx + c$



$$\hat{z}(x, y ; a, b, c) = ax + by + c$$

$$= \underbrace{[a \quad b \quad c]_{1\times 3}}_{\underline{w}^T} \underbrace{\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}_{3\times 1}}_{x}$$

$$= \underline{w}^T \underline{x}$$

Conversion of a Linear/Affine function ⟶ Dot product of two vectors

$$l(x_i, y_i;\ \underline{w}) := \Big(y_i - \underbrace{\underline{w}^T \underline{x}_i}_{\hat{y}(x_i;\,\underline{w})}\Big)^2 \qquad \underline{w} = \begin{Bmatrix} m \\ c \end{Bmatrix} \qquad \underline{x}_i = \begin{bmatrix} x_i \\ 1 \end{bmatrix}$$

$$L(D;\ \underline{w}) := \frac{1}{n} \sum_{(x_i, y_i) \in D} l(x_i, y_i)$$

$$= \frac{1}{n} \sum_i \left(y_i - \underline{w}^T \underline{x}_i\right)^2 \text{ —— ①}$$

Magnitude / norm of a vector

$$\underline{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix} \qquad |\underline{v}| = \sqrt{v_1^2 + \cdots + v_n^2} = \sqrt{\sum_i v_i^2}$$

$$\underline{l} := \begin{bmatrix} \sqrt{l(x_1, y_1)} \\ \vdots \\ \sqrt{l(x_n, y_n)} \end{bmatrix} = \begin{bmatrix} y_1 - \underline{w}^T \underline{x}_1 \\ y_2 - \underline{w}^T \underline{x}_2 \\ \vdots \\ y_n - \underline{w}^T \underline{x}_n \end{bmatrix}$$

$L(D;\ \underline{w}) = \frac{1}{n} \|\underline{l}\|_2^2$ = Same as RHS of ①
Right Hand Side
—— ②

Euclidean norm $\|v\|_2$

L_1-norm $\|v\|_1 = |v_1| + |v_2| + \cdots + |v_n|$

L_2-norm = Magnitude of vector $\|v\|_2 = \|v\|$

L_p-norm

$$\underline{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \qquad \underbrace{\|v\|_p}_{L_p\text{-norm}} = \left(|v_1|^p + |v_2|^p + \cdots + |v_n|^p\right)^{1/p}$$

L_∞-norm $= \|v\|_\infty = \cancel{\left(v_1^\infty + v_2^\infty + \dots + v_n^\infty\right)}^{1/\infty}$

$\hookrightarrow \; = \max_i |v_i|$ $\qquad 0^0 = 0$

$v_i = 0$

L_0-norm $= \|v\|_0 = \underset{1}{|v_1|^0} + \cancel{\underset{1}{|v_2|^0} + \dots} + \underset{1}{|v_n|^0}$

$\hookrightarrow$ = count the number of non zero elements in the vector

---

$$\|\underline{v}\|_2 = \sqrt{v_1^2 + \dots + v_n^2}$$

$$= \sqrt{v_1 \cdot v_1 + \dots + v_n \cdot v_n}$$

$$= \sqrt{\underline{v} \cdot \underline{v}} = \sqrt{\underline{v}^T \underline{v}}$$

Revist our loss vector and write it in vector form

$$\underline{l} = \begin{bmatrix} y_1 - \underline{W}^T \underline{x}_1 \\ y_2 - \underline{W}^T \underline{x}_2 \\ \vdots \\ y_n - \underline{W}^T \underline{x}_n \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times1}}_{\underline{y}} - \begin{bmatrix} \underline{W}^T \underline{x}_1 \\ \underline{W}^T \underline{x}_2 \\ \vdots \\ \underline{W}^T_{1\times2} \underline{x}_{n\,2\times1} \end{bmatrix}_{n\times1}$$

$(\underline{x}^T \underline{w})^T = \underline{w}^T \underline{x}$

$\underline{x} \cdot \underline{w} = \underline{w} \cdot \underline{x}$

$x_i = \begin{bmatrix} x \\ 1 \end{bmatrix}_{2\times1}$

$$\begin{bmatrix} \underline{x}_1^T \underline{w} \\ \underline{x}_2^T \underline{w} \\ \vdots \\ \underline{x}_n^T \underline{w} \end{bmatrix} = \begin{bmatrix} \underline{x}^T_{1\,1\times2} \\ \underline{x}_{2\,1\times2} \\ \vdots \\ \underline{x}^T_{n\,1\times2} \end{bmatrix}_{n\times2} \underline{W}_{2\times1}$$

$$\underline{W}^T_{1\times2} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix}_{2n\times1}$$

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

$\underline{x}^T \underline{w} = \underline{w}^T \underline{x}$

$(\underline{x}^T_{1\times2} \underline{w}_{2\times1})^T = \left[\underline{w}^T (\underline{x}^T)^T\right] = \underline{w}^T \underline{x}$

$[a]^T = [a]$

$$\underbrace{\begin{bmatrix} - \underline{a}_1^T - \\ \vdots \\ - \underline{a}_n^T - \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} | & & | \\ \underline{b}_1 & \cdots & \underline{b}_n \\ | & & | \end{bmatrix}}_{B} = \begin{bmatrix} \underline{a}_1^T \underline{b}_1 & \cdots & \underline{a}_1^T \underline{b}_n \\ & \underline{a}_i^T \underline{b}_j & \\ \underline{a}_n^T \underline{b}_1 & \cdots & \underline{a}_n^T \underline{b}_n \end{bmatrix}$$

$$(AB)^T = B^T A^T \qquad (AB)_{i,j} = \underline{a}_i^T \underline{b}_j$$

Take transpose

$$((AB)^T)_{i,j} = \boxed{\underline{a}_j^T \underline{b}_i}$$

$$\overset{B^T}{\begin{bmatrix} \underline{b}_1^T \\ \vdots \\ \underline{b}_n^T \end{bmatrix}} \overset{A^T}{\begin{bmatrix} \vdots & & \vdots \\ \underline{a}_1 & \cdots & \underline{a}_n \\ \vdots & & \vdots \end{bmatrix}} = \begin{bmatrix} \underline{b}_1^T \underline{a}_1 & \cdots & \underline{b}_1^T \underline{a}_n \\ & \boxed{\underline{b}_i^T \underline{a}_j} & \\ \underline{b}_n^T \underline{a}_1 & \cdots & \underline{b}_n^T \underline{a}_n \end{bmatrix}$$

$$\underline{a}_j = [a_{j1} \cdots a_{jn}]^T$$
$$\underline{b}_i = [b_{i1} \cdots b_{in}]$$

$$\underline{a}_j^T \underline{b}_i = a_{j1} b_{i1} + a_{j2} b_{i2} + \cdots a_{jn} b_{in}$$

$$(ABC)^T = C^T B^T A^T$$

$$\underline{l} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\underline{y}} - \underbrace{\begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \vdots \\ \underline{x}_n^T \end{bmatrix}}_{X_{n \times 2}} \underline{w} = \underline{y} - X\underline{w}$$

$$L(D;\underline{w}) = \frac{1}{n}\|\underline{e}\|_2^2 = \frac{1}{n}\|\underline{y} - X\underline{w}\|_2^2$$

Use identity: $\|v\|_2^2 = \underline{v}\cdot\underline{v} = v^T v$

$$L(D;w) = \frac{1}{n}(\underline{y} - X\underline{w})^T(\underline{y} - X\underline{w})$$

$$= \frac{1}{n}(\underline{y}^T - \underline{w}^T X^T)(\underline{y} - X\underline{w})$$

$$= \frac{1}{n}\Big(\underbrace{\underset{1\times n}{\underline{y}^T}\,\underset{n\times 1}{y}}_{1\times 1} - \underbrace{\underset{1\times 2}{w^T}\,\overset{2\times n}{X^T}\,\underset{n\times 1}{\underline{y}}}_{1\times 1} - \underbrace{\underset{1\times n}{\underline{y}^T}\,\overset{n\times 2}{X}\,\underset{2\times 1}{\underline{w}}}_{1\times 1} + \underbrace{\underset{1\times 2}{\underline{w}^T}\,\underset{2\times n}{X^T}\,\underset{n\times 2}{X}\,\underset{2\times 1}{\underline{w}}}_{1\times 1}\Big)$$

$\underline{y} \in \mathbb{R}^{n\times 1}$

$X \in \mathbb{R}^{n\times 2}$

$\underline{w} \in \mathbb{R}^{2\times 1}$

$$(\underline{y}^T X\underline{w})^T = \underline{w}^T X^T(\underline{y}^T)^T = \underline{w}^T X^T \underline{y}$$

$$L(D;w) = \frac{1}{n}(\underline{y}^T\underline{y} - 2\underline{y}^T X\underline{w} + \underline{w}^T X^T X\underline{w})$$

$$\underline{w}^* = \underset{\underline{w}}{\text{minimize}}\ L(D:\underline{w})$$

$$\left.\frac{\partial}{\partial\underline{w}} L(D;w)\right|_{w^*} = 0 \quad \Big\}\ \text{Solve for } \underline{w}^*$$

Vector calculus

Definition: Partial derivative of a function $f(\underline{x})$

$$f(\underline{x}): \mathbb{R}^n \longrightarrow \mathbb{R}$$

scalar valued - vector function

$$f(\underline{x}) = f(x_1, x_2, \ldots, x_n)$$

$$\frac{\partial}{\partial \underline{x}} f(\underline{x}) = \underbrace{\left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]_{1\times n}}$$

$\nabla_{\underline{x}} f(\underline{x})^T \longrightarrow$

column vector

Chain rule $\quad f : \mathbb{R} \to \mathbb{R}, \quad g : \mathbb{R}^n \to \mathbb{R}$

$$\frac{\partial}{\partial \underline{x}} f(g(\underline{x})) = \frac{\partial f}{\partial g}\left[\frac{\partial g}{\partial x_1} \quad \cdots \quad \frac{\partial g}{\partial x_n}\right]_{1\times n}$$

$$= \frac{\partial f}{\partial g}\,\frac{\partial g}{\partial \underline{x}}_{1\times n}$$

Multi variable version $\quad f : \underbrace{\mathbb{R} \times \mathbb{R}}_{\mathbb{R}^2} \to \mathbb{R}, \quad g_1 : \mathbb{R}^n \to \mathbb{R}$

$\qquad g_2 : \mathbb{R}^n \to \mathbb{R}$

$$\boxed{\frac{\partial}{\partial \underline{x}} f(g_1(\underline{x}), g_2(\underline{x})) = \underbrace{\frac{\partial f}{\partial g_1}}_{\text{scalar}}\underbrace{\frac{\partial g_1}{\partial \underline{x}}}_{1\times n} + \underbrace{\frac{\partial f}{\partial g_2}}_{\text{scalar}}\underbrace{\frac{\partial g_2}{\partial \underline{x}}}_{1\times n}}$$

$$\underline{g}(\underline{x}) = \begin{bmatrix} g_1(\underline{x}) \\ g_2(\underline{x}) \end{bmatrix}$$

$\underline{g}$ is a vector-valued vector func

$\underline{g} : \mathbb{R}^n \to \mathbb{R}^2$

$$\frac{\partial}{\partial \underline{x}} f(\underline{g}(\underline{x})) = \underbrace{\frac{\partial f}{\partial g_1}}_{1\times 1}\cdot\underbrace{\frac{\partial g_1}{\partial \underline{x}}}_{1\times n} + \cdot\underbrace{\frac{\partial f}{\partial g_2}}_{1\times 1}\underbrace{\frac{\partial g_2}{\partial \underline{x}}}_{1\times n}$$

$$= \underbrace{\begin{bmatrix} \underbrace{\frac{\partial f}{\partial g_1}}_{1\times 1} & \underbrace{\frac{\partial f}{\partial g_2}}_{1\times 1} \end{bmatrix}_{1\times 2}} \begin{bmatrix} \underbrace{\frac{\partial g_1}{\partial \underline{x}}}_{1\times n} \\ \underbrace{\frac{\partial g_2}{\partial \underline{x}}}_{1\times n} \end{bmatrix}_{2\times n}$$

$$\boxed{\frac{\partial}{\partial \underline{x}} f(\underline{g}(\underline{x})) = \underbrace{\frac{\partial f}{\partial \underline{g}}}_{1\times 2} \underbrace{\frac{\partial \underline{g}}{\partial \underline{x}}}_{(\text{new})\, 2\times n}}$$

Jacobian is the vector derivative of a vector-valued vector function

$$\frac{\partial \underline{g}}{\partial \underline{x}} = \begin{array}{c} \xrightarrow{\frac{\partial g_1}{\partial \underline{x}}} \\ \downarrow \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots\cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \cdots\cdots & \frac{\partial g_2}{\partial x_n} \\ \cdot & & \end{bmatrix}_{2\times n} \\ \frac{\partial g_n}{\partial x}. \end{array} = \text{Jacobian of } \underline{g}(\underline{x}) \text{ w.r.t. } \underline{x}$$

$$\frac{\partial}{\partial \underline{x}} \underline{f}\left(\underline{g}_1(\underline{h}(\underline{x})), \underline{g}_2(\underline{h}(\underline{x}))\right)$$

$$= \frac{\partial f}{\partial \underline{g}_1} \frac{\partial \underline{g}_1}{\partial \underline{h}} \frac{\partial \underline{h}}{\partial \underline{x}} + \frac{\partial f}{\partial \underline{g}_2} \frac{\partial \underline{g}_2}{\partial \underline{h}} \frac{\partial \underline{h}}{\partial \underline{x}}$$

Vector calculus

① $\frac{\partial}{\partial \underline{x}} \underbrace{\underline{a}^T \underline{x}}_{\text{Linear}} = ?$ if $\underline{a}$ is constant w.r.t $\underline{x}$

② $\frac{\partial}{\partial \underline{x}} \underline{x}^T A \underline{x} = ?$ if A is a constant matrix w.r.t $\underline{x}$

$$f(\underline{x}) \qquad [\partial f/\partial x_1 \; \ldots\ldots \; \partial f/\partial x_n]_{1\times n}$$

$$\frac{\partial}{\partial \underline{x}}(\underline{a}^T\underline{x}) = \left[\frac{\partial(\underline{a}^T\underline{x})}{\partial x_1}, \;\ldots\ldots\; \frac{\partial(\underline{a}^T\underline{x})}{\partial x_n}\right]_{1\times n}$$

$$x = \begin{bmatrix} x_1 \\ x_n \end{bmatrix}$$

$$\underbrace{\frac{\partial}{\partial x_1}\quad \underline{a}^T\underline{x}} = \overset{a_1}{\frac{\partial}{\partial x_1} a_1 x_1} + \overset{0}{\frac{\partial}{\partial x_1} a_2 x_2} + \ldots + \overset{0}{\frac{\partial}{\partial x_1} a_n x_n}$$

$$= a_1$$

$$\frac{\partial}{\partial x_2}\underline{a}^T\underline{x} = a_2$$

$$f(\underline{x}): \mathbb{R}^n \longrightarrow \mathbb{R}$$

$$\boxed{\frac{\partial}{\partial \underline{x}}\left(\underline{a}^T\underline{x}\right)_{n} = [a_1, \; a_2 \;\ldots\ldots, a_n] = \underline{a}^T}$$ Memorize

---

$$\underline{x}^T A \underline{x} \leftarrow \text{Quadratic form}$$

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\underline{x}^T A \underline{x} = [x_1 \;\; x_2] \underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$$

$$= [x_1 \;\; x_2]_{1\times 2}\begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}_{2\times 1} =$$

$$= \underbrace{a_{11}\overset{\checkmark}{x_1^2} + a_{12}\overset{\checkmark}{x_1}x_2 + a_{21}x_2x_1 + a_{22}\overset{\checkmark}{x_2^2}}_{\text{scalar (homogeneous) quadradic form}}$$

$$\underset{1\times n}{\underline{x}^T}\,\underset{n\times n}{A}\,\underset{n\times 1}{\underline{x}} = a_{11}x_1^2 + a_{22}x_2^2 + \dots + a_n x_n^2 + a_{12}x_1x_2 + a_{ij}x_ix_j + \dots$$

$$\underline{x}^T A \underline{x} = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} x_i x_j a_{ij} \quad \text{———} ①$$

$$\frac{\partial}{\partial x}\underline{x}^T A \underline{x} = \left[\frac{\partial (x^T A x)}{\partial x_1} \quad \cdots\cdots \quad \frac{\partial (\underline{x}^T A \underline{x})}{\partial x_n}\right]$$

$$\frac{\partial (\underline{x}^T A \underline{x})}{\partial x_1} = \underbrace{\frac{\partial}{\partial x_1}\sum_{i=1}^{n} a_{ii}x_i^2}_{} + \underbrace{\frac{\partial}{\partial x_1}\sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} a_{ij}x_ix_j}_{x_i = x_1 \qquad x_j = x_1}$$

$$= 2a_{11}x + \sum_{j=1,\, j\neq 1}^{n} a_{1j}x_j + \sum_{i=1,\, i\neq 1}^{n} a_{i1}x_i$$

$$= \sum_{j=1}^{n} a_{1j}x_j + \sum_{i=1}^{n} a_{i1}x_i$$

$$= \sum_{i=1}^{n} a_{1i}x_i + \sum_{i=1}^{n} a_{i1}x_i$$

$$= \sum_{i=1}^{n} (a_{1i} + a_{i1})\, x_i$$

$$A = \begin{bmatrix} \top & a_{1,:} & \text{———} \\ a_{:,1} & & \\ | & & \end{bmatrix}$$

$\underline{a}_{1,:}$ is the first row of A as a column vector

$\underline{a}_{:,1}$ is the first column

$$\frac{\partial}{\partial x_1}(\underline{x}^T A \underline{x}) = \underline{a}_{1,:}^T \underline{x} + \underline{a}_{:,1}^T \underline{x}$$

$$\frac{\partial}{\partial \underline{x}}(\underline{x}^T A \underline{x}) = \begin{bmatrix} \underline{a}_{1,:}^T \underline{x} + a_{:,1}^T \underline{x}, & \underline{a}_{2,:}^T \underline{x} + \underline{a}_{:,2}^T \underline{x}, & & , \underline{a}_{n:}^T \underline{x} + \underline{a}_{:2}^T \underline{x} \end{bmatrix}_{1\times n}$$

$$= \underset{1\times n}{\underline{x}^T} \underbrace{\begin{bmatrix} | & | & & | \\ \underline{a}_{1,:}, & \underline{a}_{2,:} & --- & \underline{a}_{n,:} \\ | & | & & | \end{bmatrix}_{n\times n}}_{A^T} + \underline{x}^T \underbrace{\begin{bmatrix} | & | & & | \\ \underline{a}_{:,1}, & \underline{a}_{:,2} & \cdots\cdots & \underline{a}_{:,n} \\ | & | & & | \end{bmatrix}_{n\times n}}_{A}$$

$$\boxed{\frac{\partial(\underline{x}^T A \underline{x})}{\partial \underline{x}} = \underline{x}^T(A^T + A)}$$ memorize

Special case $A^T = A$ (A is a symmetric matrix)

$$\boxed{\frac{\partial}{\partial x}(\underline{x}^T A \underline{x}) = 2\underline{x}^T A}$$ Memorize

$$\frac{\partial}{\partial \underline{x}} \underline{a}^T \underline{x} = a^T$$

$$\frac{\partial}{\partial \underline{x}} \underbrace{A\underline{x}}_{n\times 1} = \frac{\partial}{\partial \underline{x}} \underbrace{\begin{bmatrix} -\underline{a}_1^T- \\ -\underline{a}_2^T- \\ | \\ -\underline{a}_n^T- \end{bmatrix}_{n\times n}}_{A} \underline{x}_{n\times 1} = \frac{\partial}{\partial \underline{x}} \begin{bmatrix} \underline{a}_1^T \underline{x} \\ \vdots \\ \underline{a}_n^T \underline{x} \end{bmatrix}_{n\times 1} = \begin{bmatrix} \frac{\partial \underline{a}^T \underline{x}}{\partial \underline{x}} \\ \vdots \\ \frac{\partial \underline{a}_n^T x}{\partial \underline{x}} \end{bmatrix}_{n\times n}$$

$$= \begin{bmatrix} -\underline{a}_1^T- \\ \vdots \\ -a_n^T- \end{bmatrix}_{n\times n} = A_{n\times n}$$

$$\underline{x}^T\underline{a} \overset{?}{=} \underline{a}^T\underline{x} \overset{?}{=} \underline{x}\cdot\underline{a} \overset{?}{=} \underline{a}\cdot\underline{x} .$$

$$\boxed{\frac{\partial}{\partial \underline{x}} A\underline{x} = A .}$$ Memorize

$$\frac{\partial}{\partial \underline{x}} \underbrace{a^T}_{A}x = \underbrace{a^T}_{A}$$

$$\frac{\partial}{\partial \underline{x}} \underline{x}^T\underline{a} = \underline{a}^T$$

$\downarrow$

Always take derivatives of <u>column vectors</u>

Linear

① $\frac{\partial}{\partial \underline{x}} \underline{a}^T\underline{x} = \underline{a}^T$

$\frac{\partial}{\partial \underline{x}} A\underline{x} = A$

Quadratic

② $\frac{\partial}{\partial \underline{x}} \underline{x}^TA\underline{x} = x^T(A^T + A)$

~~Cubic form~~

$\cancel{(\underline{x}^T A\ \underline{x})\underline{x}} = ?$

$\uparrow$ Tensor

Linear Regression

$$\frac{\partial}{\partial \underline{w}_{2\times1}} L(D;\underline{w})\Big|_{w^*} = \underline{0}^T_{1\times2}$$

$$\frac{\partial}{\partial \underline{w}}\left\{\frac{1}{n}\left(\overbrace{\underline{y}^T\underline{y}}^{\text{const.}} - \overbrace{2\underbrace{\underline{y}^T X_{n\times2}}_{1\times2}\underline{w}}^{\text{Linear}} + \overbrace{\underline{w}^TX^TX\underline{w}}^{\text{Quadratic form of } w}\right)\right\} = \underline{0}^T_{1\times2} \quad \text{—— ②}$$

① This equation is Quadratic in $\underline{w}$

③ — $\frac{\partial}{\partial \underline{w}} -2\underbrace{\underline{y}^T_{1\times n} X_{n\times 2}}_{1\times 2} \underline{w}_{2\times 1} = -2\underline{y}^T X$

$\frac{\partial}{\partial \underline{w}} \underline{a}^T \underline{w} = \underline{a}^T$

④ — $\frac{\partial}{\partial \underline{x}} \underline{w}^T \underbrace{X^T_{2\times n} X_{n\times 2}}_{2\times 2} \underline{w} = 2\underline{w}^T X^T X$

$\frac{\partial}{\partial \underline{w}} \underline{w}^T A \underline{w}$

$= 2\underline{w}^T A$

when

A is

symmetric

is X^TX symmetric? Yes

$$(X^TX)^T = X^T(X^T)^T = X^TX$$

$$A^T = A \qquad (X^TX)^T = X^T(X^T)^T = X^TX$$

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

$$\frac{1}{n}\left(-2\underline{y}^TX + 2\underline{w}^TX^TX\right) = 0^T$$

$$\underline{w}^T(X^TX) = \underline{y}^TX$$

Right multiply by $(X^TX)^{-1}$

$$\underline{w}^T = \underline{y}^TX(X^TX)^{-1}$$

$$\boxed{\underline{w} = (X^TX)^{-1}X^T\underline{y}}$$