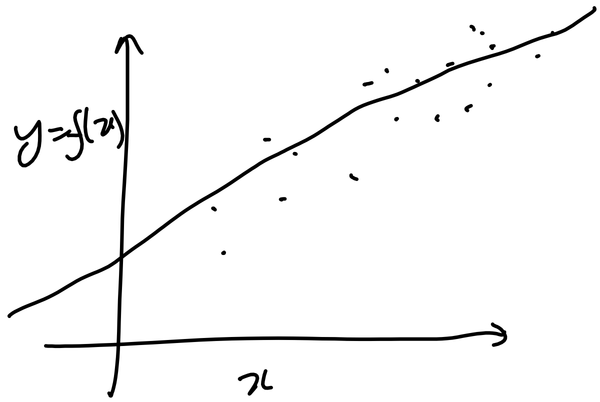
Linear Regression

input

$$y = m x + c$$

output

parameters



Suppose you are given Dataset (Training Dataset)

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

$$x_i \xrightarrow{?} y_i$$

Guess : Model

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

Loss function : Mean squared Error

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

(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

(3) Training

(2) Loss | Penalty

minimize the Loss function w.r.t. the parameters

with respect to

Trained weights

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

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

$x \notin D$

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

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

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

Is this function Linear/Affine in m and x?

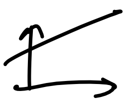
No

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

× Quadratic term x^2

× Cubic term x^3

× $\exp()$, $\log()$, $\sin()$, $\cos()$

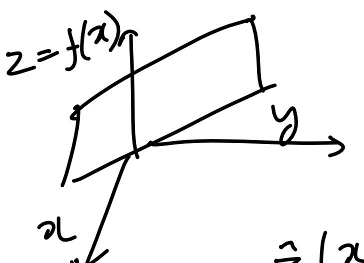


$$\hat{y}(x_i ; [m, c]) = \underbrace{[m \quad c]_{1\times 2}}_{\text{coeff row vector}} \underbrace{\begin{bmatrix} x \\ 1 \end{bmatrix}_{2\times 1}}_{\text{variables/inputs column vector}}$$

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

$$\underline{w} = \begin{bmatrix} m \\ c \end{bmatrix} \qquad \underline{x} = \begin{bmatrix} x \\ 1 \end{bmatrix}$$

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



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

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

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

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

$$l(x_i, y_i; \underline{w}) := (y_i - \underbrace{\underline{w}^T \underline{x}_i}_{\hat{y}(x_i; \underline{w})})^2$$

$$\underline{w} = \begin{Bmatrix} m \\ c \end{Bmatrix} \qquad \underline{x}_i = \begin{bmatrix} x_i \\ 1 \end{bmatrix}$$

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

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

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) ——— (2)

Right Hand Side

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

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

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

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

$v_i = 0$

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

$$\quad 1 \quad 1 \quad\quad 1$$

$\hookrightarrow$ = 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

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

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

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

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

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

$$\underline{w}^T_{1\times 2} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_n \end{bmatrix}_{2n \times 1}$$

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

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

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

$$[a]^T = [a]$$

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

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

Take transpose

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Vector calculus

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

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

scalar valued - vector function

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

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

$\nabla_{\underline{x}} f(\underline{x})$ column vector

Chain rule

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

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

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

Multi variable version $f: \underbrace{\mathbb{R} \times \mathbb{R}}_{\mathbb{R}^2} \to \mathbb{R}$, $g_1: \mathbb{R}^N \to \mathbb{R}$, $g_2: \mathbb{R}^N \to \mathbb{R}$

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

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

$\underline{g}$ 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

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

$\frac{\partial g_1}{\partial \underline{x}}$ (across the first row)

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

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

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 $\underline{x}$

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

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

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

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

$$= a_1$$

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

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

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

Memorize

$\underline{x}^T A \underline{x} \longleftarrow$ Quadratic form

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

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

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

$$= \underbrace{a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2}_{\text{scalar (homogeneous) quadradic form}}$$

$$\underset{1\times n}{\underline{x}^T}\ \underset{n\times n}{A}\ \underset{n\times 1}{\underline{x}} = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{n}x_n^2$$

$$+ a_{12}x_1x_2 + a_{ij}x_ix_j + \cdots$$

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

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

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

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

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

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

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

$$A = \begin{bmatrix} \top & \underline{a}_{1,:} & \longrightarrow \\ \underline{a}_{:,1} & & \\ | & & \end{bmatrix}$$

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

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

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

$$\frac{\partial}{\partial \underline{x}}(\underline{x}^T A \underline{x}) = \left[\underline{a}_{1,:}^T \underline{x} + \underline{a}_{:,1}^T \underline{x},\ \underline{a}_{2,:}^T \underline{x} + \underline{a}_{:,2}^T \underline{x},\ ,\ \underline{a}_{n,:}^T \underline{x} + \underline{a}_{:,2}^T \underline{x}\right]_{1\times n}$$

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

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

memorize

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

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

Memorize

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

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

$$= \begin{bmatrix} -a_1^T- \\ \vdots \\ -a_n^T- \end{bmatrix}_{n\times n} = A_{n\times n}$$

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

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

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

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

Always take derivatives of column vectors

Linear

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

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

~~cubic form~~

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

Tensor

Quadratic

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

Linear Regression

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

$$\frac{\partial}{\partial \underline{w}}\left\{\frac{1}{n}\Big(\overbrace{\underline{y}^T\underline{y}}^{\text{const.}} - \overbrace{2\underbrace{\underline{y}^T_{1\times 2}X_{n\times 2}}\underline{w}}^{\text{Linear}} + \overbrace{\underline{w}^T X^T X\underline{w}}^{\text{Quadratic form of } w}\Big)\right\} = \underline{0}^T_{1\times 2} \qquad ②$$

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

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

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

④ — $\dfrac{\partial}{\partial \underline{x}} \underline{w}^T \underbrace{\underset{2\times n}{X^T} \underset{n\times 2}{X}}_{2\times 2} \underline{w} = 2\underline{w}^T X^T X$

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

when A is symmetric

is $X^T X$ symmetric? Yes

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

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

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

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

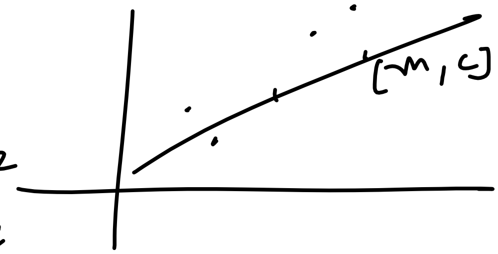
$$\underline{w}^T (X^T X) = \underline{y}^T X$$

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

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

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

1. Dataset
2. Model $\hat{y}(x) = mx + c$
3. Loss $\ell(x,y) = \|y_i - \hat{y}(x_i)\|_2^2$



4. Training $\underset{\underline{w}\ [m, c]}{\text{minimize}} \sum_{(x_i, y_i) \in D} \ell(x_i, y_i) = L(D; \underline{w})$

$$\frac{\partial}{\partial \underline{w}} L(D; \underline{w}) = \underline{0}^T$$

$$\begin{bmatrix} m \\ c \end{bmatrix} = \boxed{\underline{w}^* = \underbrace{(X^T X)^{-1} X^T}_{\text{Pseudo Inverse of } X} \underline{y}}, \quad X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_n & 1 \end{bmatrix}$$

$X^\dagger \leftarrow$ dagger $\qquad \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

$$X^\dagger = (X^T X)^{-1} X^T \qquad \dagger$$

Proof

$$L(D; \underline{w}) = \|\underline{y} - X\underline{w}\|_2^2 \longrightarrow \text{minimize}$$

$$\underset{n\times 1}{\underline{y}} \approx \underset{n\times 2}{X}\,\underset{2\times 1}{\underline{w}} \longrightarrow \underline{w} = X^\dagger \underline{y} = \underbrace{(\underset{2\times n}{X^T}\,\underset{n\times 2}{X})}_{2\times 2}^{-1} X^T \underline{y}$$

np.linalg.solve$(A, \underline{b})$

$$A\underline{x} = b \Rightarrow x = A^{-1}\underline{b}$$

np.linalg.lstsq$(X, \underline{y}) \Rightarrow \underline{w} = X^\dagger \underline{y}$

$f(\underline{x} + \underline{y}) = f(\underline{x}) + f(\underline{y})$ $\quad f(\alpha \underline{x}) = \alpha f(\underline{x})$

$$\boxed{f(\alpha \underline{x} + \beta \underline{y}) = \alpha f(\underline{x}) + \beta f(\underline{y})}$$

$$f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 3x_1 + 2x_2$$

$$= \begin{bmatrix} 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

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

$$L.H.S = f(\alpha \underline{x} + \beta \underline{y})$$

$$= \underline{c}^T (\alpha \underline{x} + \beta \underline{y})$$

$$= \alpha \underline{c}^T \underline{x} + \beta \underline{c}^T \underline{y}$$

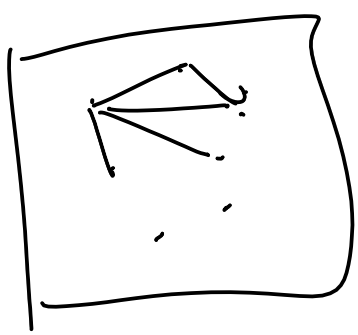
$$= \alpha f(\underline{x}) + \beta f(\underline{y}) = R.HS$$

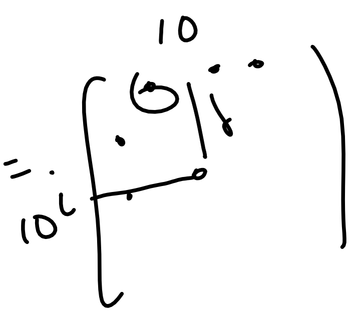
$$
\begin{array}{cc}
 & \xrightarrow{\text{feature dim}} \\
X = \underset{\text{sample size}}{\big\downarrow} & \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & & \\ x_n & y_n & z_n \end{bmatrix}
\end{array}
$$

sample feature

$$X.\text{shape} = (\underset{\text{sample}}{n}, \underset{\text{feature}}{3})$$

Assign meanings to dimensions rather than visualizing higher dim arrays



DIST = 

$$\|x_i - x_j\|_2^2$$