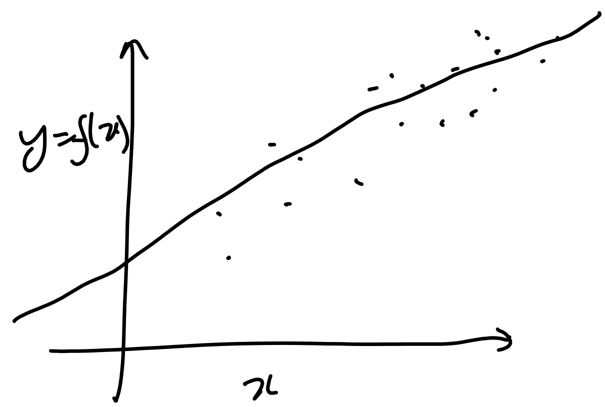
Linear Regression

input

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

output ← y; parameters ← m, c



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

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

Loss function: Mean squared Error

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

(Penalty for making a wrong guess)

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

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

inputs parameters

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

① Model / Guess

② Loss | Penalty

③ Training

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 $\underline{\underline{\text{in } x}}$

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?

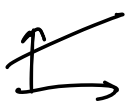
$\underline{\underline{\text{No}}}$

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

× Quadratic term x^2

× Cubic term x^3

× exp( ), log( ), sin( ), cos( )



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

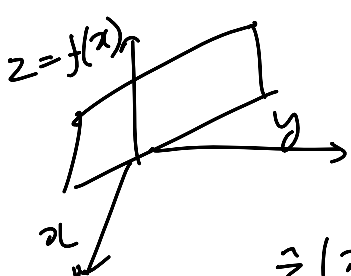
coeff row vector

variables / inputs 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

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

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

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

Magnitude / norm of a vector

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

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

$$L(D; \underline{w}) = \frac{1}{n} \|\underline{\ell}\|_2^2 = \text{Same as RHS of } (1)$$

Right Hand Side

——— (2)

Euclidean norm $\|v\|_2$

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

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

L_p - norm

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

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

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

$v_i = 0$

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

$\quad 1 \quad 1 \quad 1$

$\hookrightarrow$ = 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

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

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

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

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

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

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

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

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

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

$[a]^T = [a]$

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

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

Take transpose

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

scalar valued - vector function

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

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