$$J(\theta) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \theta_j \cdot x_j^{(i)} - y^{(i)} \right)^2$$
$$= \| X\underline{\theta} - \underline{y} \|^2$$

- vector.
  - $\underline{v} \triangleq \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$, $\underline{v} \in \mathbb{R}^{n\times 1}$, Def.
  - ① Elementory-Wise Product.

    $\underline{x} \in \mathbb{R}^n$, $\underline{y} \in \mathbb{R}^n$

    $$\underline{x} \odot \underline{y} \triangleq \begin{bmatrix} x_1 y_1 \\ \vdots \\ x_n y_n \end{bmatrix}_{n\times 1}.$$

    ② Knornecker Product (Outer Produce)

    $\underline{x} \in \mathbb{R}^n$, $\underline{y} \in \mathbb{R}^n$

    $$\underline{x} \otimes \underline{y} \triangleq \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \\ x_n y_1 & \cdots & x_n y_n \end{bmatrix}_{n\times n}.$$

    ③ Inner Product.

    $$\langle \underline{x}, \underline{y} \rangle \triangleq x_1 y_1 + \cdots + x_n y_n, \qquad \langle \underline{x}, \underline{y} \rangle_A = \underline{x}^T A \cdot \underline{y}.$$
    $$= \underline{x}^T \underline{y}$$
    $$= \sum_{i=1}^{n} x_i y_i$$

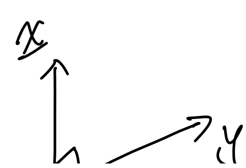
    Remark:
    - $\sum_{i=1}^{n} x_i = x_1 + \cdots + x_n$

      e.g. 1: $\sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j k_j = \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j k_j\right)$, $\qquad \sum_{i=1}^{n} x_i = 1 / 0$.

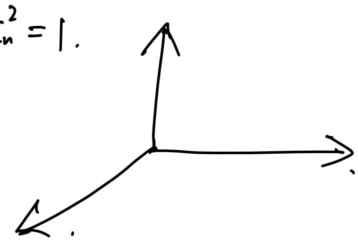
      e.g. 2: $\sum_{y} \sum_{x} Q_{X|Y}(x, y) P_X(x)$.
      $$= \sum_{x} P_X(x) \underbrace{\sum_{y} Q_{X|Y}(x,y)}_{} = Q_X(x).$$
    - $\prod_{i=1}^{n} x_i \triangleq x_1 \cdots x_n$.

- Sepcial Inner Product.

  ① Orthogonal: $\langle \underline{x}, \underline{y} \rangle = 0$, $x_1 y_1 + \cdots + x_n y_n = 0$.



② Unit Vector. $\langle \underline{x}, \underline{x} \rangle = 1$, $x_1^2 + \dots + x_n^2 = 1$.



- Orthogonal Matrix.

$Q \in \mathbb{R}^{n \times n}$, $Q = [\underbrace{\underline{q}_1}_{\in \mathbb{R}^n} | \cdots | \underline{q}_n]$.

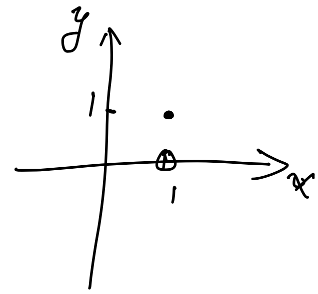
$$Q^T Q = \begin{bmatrix} \underline{q}_1^T \\ \vdots \\ \underline{q}_n^T \end{bmatrix}_{n \times 1} [\underline{q}_1 \cdots \underline{q}_n]_{1 \times n} = \begin{bmatrix} \underline{q}_1^T \underline{q}_1 & \cdots & \underline{q}_1^T \underline{q}_n \\ \vdots & & \vdots \\ \underline{q}_n^T \underline{q}_1 & \cdots & \underline{q}_n^T \underline{q}_n \end{bmatrix}_{n \times n}.$$

$$= \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

$\forall i, j \in \{1, \dots, n\}$.

$$\underline{q}_i^T \cdot \underline{q}_j = \delta_{ij} \triangleq \begin{cases} 1 & , i = j \\ 0 & , \text{o.w.} \end{cases}$$

- Indictor Function: $\mathbb{1}\{x = 0\} \triangleq \begin{cases} 1 & , x = 0 \\ 0 & , \text{o.w.} \end{cases}$



$\{x^{(1)}, \dots x^{(n)}\}$, $\{0, 1\}$.

$$\sum_{i=1}^{n} \mathbb{1}\{x^{(i)} = 0\}$$

- Norm: $\underline{x} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$,

l_2: $\|\underline{x}\|_2^2 \triangleq \underline{x}^T \underline{x} = 1^2 + (-2)^2 + 0^2 = 5$

l_1: $\|\underline{x}\|_1 = 1 + |-2| = 3$

l_0: $\|\underline{x}\|_0 = 2$.

l_∞: $\|\underline{x}\|_\infty = \max_i |x_i| = 2$.

- Prove. $\|\underline{x}+\underline{y}\|^2 = \|\underline{x}\|^2 + \|\underline{y}\|^2 \Leftrightarrow \underline{x} \perp \underline{y}$.

$$x_1 y_1 + \dots + x_n y_n = 0.$$

$$x^T y = 0.$$

- Trace. $M \in \mathbb{R}^{n \times n}$, $M = \begin{bmatrix} m_{11} & \cdots & m_{1n} \\ \vdots & & \\ m_{n1} & \cdots & m_{nn} \end{bmatrix}$.

$$tr(M) \triangleq \sum_{i=1}^{n} m_{ii}$$

$$tr(M \cdot \underline{y} \cdot \underline{x}^T) = tr(\underset{1\times n}{\underline{x}^T} \cdot M \cdot \underset{n \times 1}{\underline{y}}) = \underline{x}^T \cdot M \cdot \underline{y} = \langle \underline{x}, \underline{y} \rangle_M .$$

$$tr(AB) = tr(BA).$$

$$tr(ABC) = tr(CAB) = tr(BCA).$$

$$J(\theta) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \theta_j \cdot x_j^{(i)} - y^{(i)} \right)^2$$

$$= \|X\underline{\theta} - \underline{y}\|^2$$

$$\underline{\theta} \triangleq \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$x_j^{(i)}$, $y^{(i)}$

$$x^{(i)} \triangleq \begin{bmatrix} x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}.$$

$$\sum_{j=1}^{n} \theta_j \cdot x_j^{(i)} = \underline{\theta}^T \cdot x^{(i)}$$

$$J(\theta) = \sum_{1}^{m} \left((x^{(i)})^T \cdot \underline{\theta} - y^{(i)} \right)^2$$

$$= \|X\underline{\theta} - \underline{y}\|_2^2.$$

$$\|\underline{x}\|_2^2 \triangleq x_1^2 + \dots + x_n^2.$$

$$X \triangleq \begin{bmatrix} - & x^{(1)T} & - \\ & \vdots & \\ - & x^{(m)T} & - \end{bmatrix}, \quad \underline{y} \triangleq \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix} \in \mathbb{R}^{m \times 1}.$$

Calculus

$w \cdot x \cdot y \to$ vector

$A \to$ matrix

$b \to$ scalar.

$y = x \quad x = 0 \quad \frac{dy}{dx} = 1 \Rightarrow \frac{\partial y}{\partial x}$

$y = f(X)$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1} \qquad \frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}_{n\times 1}$$

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{bmatrix}_{n\times n} \qquad \frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \vdots & & \vdots \\ \frac{\partial y}{\partial x_{n1}} & \cdots & \frac{\partial y}{\partial x_{nn}} \end{bmatrix}_{n\times n}$$

ex 1. $\frac{\partial (w^T x + b)}{\partial x}$ $\qquad w = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}_{n\times 1} \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$ $\quad b$: scalar.

$$= \begin{bmatrix} \frac{\partial (w^T x + b)}{\partial x_1} \\ \vdots \\ \frac{\partial (w^T x + b)}{\partial x_n} \end{bmatrix}_{n\times 1} \qquad \frac{\partial (w^T x + b)}{\partial x_1} = w_1$$

$$= \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = w$$

ex 2. $\frac{\partial (x^T A x + b)}{\partial x}$ $\qquad x^T A x \to$ scalar $\quad b \to$ scalar

$$= \begin{bmatrix} \frac{\partial (x^T A x + b)}{\partial x_1} \\ \vdots \\ \frac{\partial (x^T A x + b)}{\partial x_n} \end{bmatrix}_{n\times 1}$$

$$\frac{\partial (x^T A x + b)}{\partial x_1} = \frac{\partial (\sum_{j=1} \sum_{i=1} x_i a_{ij} x_j)}{\partial x_1} = \sum_{i=1}^{n} x_i a_{i1} + \sum_{j=1}^{n} x_j a_{1j}$$

$$x^T A = [x_1 \cdots\cdots x_n]_{1\times n} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}_{n\times n}$$

$$= \left[\sum_{i=1}^{n} x_i a_{i1}, \cdots\cdots, \sum_{i=1}^{n} x_i a_{in} \right]_{1\times n}$$

$$(x^T A) \cdot x = [\cdots\cdots]_{1\times n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} x_i a_{ij} x_j .$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i a_{i1} + \sum_{j=1}^{n} x_j a_{1j} \\ \vdots \\ \sum_{i=1}^{n} x_i a_{in} + \sum_{j=1}^{n} x_j a_{nj} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i a_{i1} \\ \vdots \\ \sum_{i=1}^{n} x_i a_{in} \end{bmatrix}_{n\times 1} + \begin{bmatrix} \sum_{j=1}^{n} x_j a_{1j} \\ \vdots \\ \sum_{j=1}^{n} x_j a_{nj} \end{bmatrix}_{n\times 1}$$

$$= \underset{n\times n\ \ n\times 1}{A x} + \underset{n\times n\ \ n\times 1}{A^T x}$$

$$\frac{\partial (x^T A^{-1} y)}{\partial A} =$$

$$X = \begin{bmatrix} x_{11} & \cdots\cdots & x_{1m} \\ \vdots & & \\ x_{n1} & \cdots\cdots & x_{nm} \end{bmatrix}_{n\times m} \qquad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \qquad \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}_{m\times 1}$$

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{m} \theta_j x_j^{(i)} - y^{(i)} \right)^2 = (X\theta - Y)^T (X\theta - Y)$$

$$\frac{\partial (X\theta - Y)^T (X\theta - Y)}{\partial \theta} = \frac{\partial (\theta^T X^T - Y^T)(X\theta - Y)}{\partial \theta}$$

$$= \frac{\partial (\theta^T X^T X \theta - Y^T X \theta - \theta^T X^T Y + Y^T Y)}{\partial \theta}$$

$$= \frac{\partial \theta^T X^T X \theta}{\partial \theta} - \frac{\partial Y^T X \theta}{\partial \theta} - \frac{\partial \theta^T X^T Y}{\partial \theta} + \cancel{\frac{\partial Y^T Y}{\partial \theta}}$$

$$= \quad \boxed{A = X^T X} \qquad \boxed{X^T Y = w} \qquad \boxed{X^T Y} = w$$

$$= \frac{\partial \theta^T A \theta}{\partial \theta} - \frac{\partial w^T \theta}{\partial \theta} - \frac{\partial \left((X^T Y)^T \theta\right)^T}{\partial \theta}$$

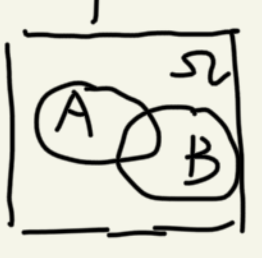
$$= \frac{\partial \theta^T A \theta}{\partial \theta} - \frac{\partial (w^T \theta)}{\partial \theta} - \frac{\partial (w^T \theta)^T}{\partial \theta} \Rightarrow \frac{\partial (w^T \theta)^T}{\partial (w^T \theta)} \cdot \frac{\partial w^T \theta}{\partial \theta}$$

$$= 2X^T X \theta - 2X^T Y$$

Probability $\leftarrow$ Measure Theory

$\rightarrow$ Concepts & Notations

- Random Phenomenon (diff outcomes in same setting)
- Sample Space 「Ω」 discrete / continuous
- Random Events 「$A\ B \cdots E$」 sets
  - o language: A: Today will rain (statement)
  - o $A \subseteq \Omega$



- o Random Variables 「X Y Z」

  $X = 1 / 0 \rightarrow$ functions

- Probability Measure $\mathbb{P}$

  「$P(A)$. $P(X=1)$. $P_X(x)$ / $f_X(x)$. $P_i = P(X=x_i)$」

  e.g. frequency $P(E) = \frac{1}{2} \approx \frac{4998}{10000}$

  e.g. classic method: toss a coin twice

  $\Omega = \{HH . HT . TH . TT\}$ H: head T: tail

  $\hookrightarrow$ sample / elementary outcome

  A = different outcomes = $\{HT . TH\}$

  $\mathbb{P}(A) = \frac{2}{4} = \frac{1}{2}$

  - o Properties of $\mathbb{P}$
    1. $P(E) \geq 0$
    2. $P(\Omega) = P(E_1 \cup E_2 \cdots \cup E_k) = 1$
    3. if $E_1 \cap E_2 = \Phi$: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

    $\triangle$ Partition of Ω $E_i \cap E_j = \Phi$ $\Omega = \bigcup_k E_k$

E_1	E_2	Ω
E_3	E_4	

$\rightarrow$ Conditional Prob.

- $P(B|A) \triangleq \frac{P(A \cap B)}{P(A)}$ $\leftarrow P(A \cap B) = P(B|A) \cdot P(A)$

  $= P(A|B) \cdot P(B)$

- Total prob rule: given a partition of $\Omega = \bigcup_i B_i$.

  $P(A) = \sum_i P(B_i) \cdot P(A|B_i)$

- Bayes

  $P(B_i|A) = \frac{P(A \cap B)}{\sum_i P(B_i) \cdot P(A|B_i)} = \frac{P(A|B_i) \cdot P(B_i)}{\sum_i P(B_i) \cdot P(A|B_i)}$

$\rightarrow$ Random Variables 「X Y Z $\leftarrow$ x y z」

- $\Omega = \{w\}$
- $X(w) \in \mathbb{R}$ function
- $A \overset{?}{\longleftrightarrow} X$

  $A \triangleq \{a < X \leq b\} = \{w \mid a < X(w) \leq b\} \subseteq \Omega$

- Prob Density Function (pdf)

  「$p(x) \leftarrow P_X(x) = P(X=x)$. $f_X(x)$ / $f(x)$」

  $P(a \leq X \leq b) = \int_a^b p(x)\,dx$

- Distrition e.g. Gaussian dist.

  $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \rightarrow f_X(x; \mu, \sigma)$

$\rightarrow$ Expectation: $(X . \mathbb{P})$

$E[X] \triangleq \sum_x x\, P_X(x) \triangleq \int x\, P_X(x)\,dx$

$Y = g(x)$ $E[g(x)] = \int g(x) \cdot P_X(x)\,dx$

e.g. $E[X^2] = \sum_x x^2 \cdot P_X(x) = \frac{1}{2} \times 0^2 + \frac{1}{4} \times 1^2 + \frac{1}{4} \times 2^2$

X	0	1	2
P	1/2	1/4	1/4

$E[Y|X=x] \triangleq \sum_y y\, P_{Y|X}(y|x) \leftarrow g(x)$

e.g. $E[E[Y|X]] = E[g(X)]$ $\sum_x P_{XY}(x,y) = P_Y(y)$

$= \sum_x g(x) \cdot P_X(x)$

$= \sum_x \sum_y y\, P_{Y|X}(y|x) \cdot P_X(x) = P_{XY}(x,y)$

$= \sum_y y \cdot P_Y(y) = E[Y]$

$\rightarrow$ Variance:

$Var[X] \triangleq E[(X - E[X])^2]$

$= \int (x - E[X])^2 \cdot P_X(x)\,dx$

$= \int (x^2 - 2E[X] \cdot x + (E[X])^2) \cdot P_X(x)\,dx$

$= E[X^2] - (E[X])^2$

$\rightarrow$ Covariance:

constants

$Cov[X.Y] \triangleq E[(X - E[X])(Y - E[Y])]$

$= E[XY] - E[X]E[Y]$ ②

$\rightarrow$ Independent: $f_{XY}(x.y) = f_X(x) \cdot f_Y(y)$ ①

① + ② $\Rightarrow X \perp\!\!\!\perp Y$ $Cov[X.Y] = 0$