1. Vector.

- $\vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}_{n\times 1}$, $\vec{v} \in \mathbb{R}^{n\times 1}$.

- Inner Product.

$\vec{x} \in \mathbb{R}^{n\times 1}$, $\vec{y} \in \mathbb{R}^{n\times 1}$.

$$\langle \vec{x}, \vec{y} \rangle \triangleq x_1 y_1 + \dots + x_n y_n = x^T y = \sum_{i=1}^{n} x_i y_i$$

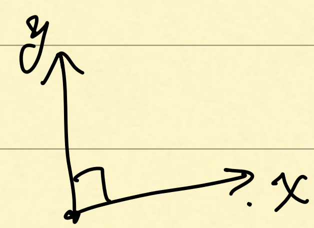
- Notations:

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + \dots + x_n y_n$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j k_j = \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j k_j\right) \leftarrow$$

$$\prod_{i=1}^{n} x_i = x_1 \cdot \dots \cdot x_n$$

- $\langle x, y \rangle = 0$. $x \perp y$



0.1. if $x \perp y$

$$\|x+y\|_2^2 = \|x\|^2 + \|y\|_2^2$$

① $\|x+y\|_2^2 = \left\| \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \right\|_2^2$

$$= (x_1+y_1)^2 + \dots + (x_n+y_n)^2 = \sum_{i=1}^{n} (x_i+y_i)^2$$

$$= \sum_{i=1}^{n} x_i^2 + 2x_iy_i + y_i^2$$

$$= \sum_{i=1}^{n} x_i^2 + 2\sum_{i=1}^{n} x_iy_i + \sum_{1} y_i^2$$

$$= \|x\|_2^2 + 2\underbrace{\langle x, y\rangle}_{=0.} + \|y\|_2^2$$

② $\|x+y\|_2^2 = \langle x+y, x+y\rangle = x^Tx + \overset{0}{2x^Ty} + y^Ty = x^Tx + y^Ty.$

- $Q \in \mathbb{R}^{n\times n}.$ $\quad Q = [\underset{\in \mathbb{R}^{n\times 1}}{q_1} | \cdots | q_n]$

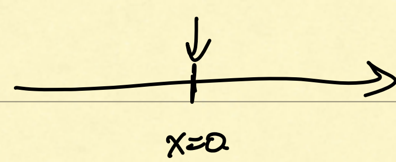
$$Q^TQ = \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} \cdot [q_1 | \cdots | q_n].$$

$$= \begin{bmatrix} q_1^Tq_1 & \cdots & q_1^Tq_n \\ \vdots & & \vdots \\ q_n^Tq_1 & \cdots & q_n^Tq_n \end{bmatrix} = I_n.$$

$$q_i^T \cdot q_i = 1, \quad q_i^T \cdot q_j = 0, \; i \neq j.$$

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

$$\mathbb{1}_{\{x=0\}} = \begin{cases} 1, & x=0 \\ 0, & \text{o.w} \end{cases}$$



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

$$tr(M) = m_{11} + \cdots + m_{nn}$$

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

· $x, y \in \mathbb{R}^{n\times 1}$, $M \in \mathbb{R}^{n\times n}$.

$$tr(\underbrace{\boxed{M \cdot y}}_{n\times 1} \cdot \underbrace{x^T}_{1\times n}) = tr(\underline{x^T \cdot M \cdot y}) = x^T \cdot M \cdot y.$$

$$\nabla_A tr(A \cdot y \cdot x^T) = \nabla_A (x^T \cdot A \cdot y) = y \cdot x^T.$$

$$A \cdot B = \begin{bmatrix} a_{11}, & \cdots, & a_{1n} \\ & & \\ a_{n1}, & \cdots, & a_{nn} \end{bmatrix} \cdot \begin{bmatrix} b_{11}, & \cdots, & b_{1n} \\ & & \\ b_{n1}, & \cdots, & b_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{1i} \cdot b_{i1}, & \cdots, & \sum_{i=1}^{n} a_{1i} \cdot b_{in} \\ \vdots & & \\ \sum_{i=1}^{n} a_{ni} \cdot b_{i1}, & \cdots, & \sum_{i=1}^{n} a_{ni} \cdot b_{in} \end{bmatrix}$$

$$tr(AB) = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ji} \cdot b_{ij}$$

$$tr(BA) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \cdot b_{ji}$$

- $x = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$, $l_2 : \|x\|_2^2 \triangleq x^T x = 1^2 + (-2)^2$

  $l_0 : \|x\|_0 = 2$

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

  $l_\infty : \|x\|_\infty = \max_i |x_i| = 2$.

II. Derivation.

$a \in \mathbb{R}$, $\vec{x} \in \mathbb{R}^n$

- $\frac{\partial a}{\partial \vec{x}} \triangleq \begin{bmatrix} \frac{\partial a}{\partial x_1} \\ \vdots \\ \frac{\partial a}{\partial x_n} \end{bmatrix}$, $\frac{\partial \vec{x}}{\partial a} \triangleq \begin{bmatrix} \frac{\partial x_1}{\partial a} \\ \vdots \\ \frac{\partial x_n}{\partial a} \end{bmatrix}$.

- $\frac{\partial (w^T x + b)}{\partial x} = w$.

$w^T x + b = \sum_{i=1}^{n} w_i x_i + b$

$$\frac{\partial (w^T x + b)}{\partial x} = \begin{bmatrix} \frac{\partial \Sigma w_i x_i + b}{\partial x_1} \\ \vdots \\ \frac{\partial \Sigma w_i x_i + b}{\partial x_n} \end{bmatrix} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$\boxed{\sum_{i=1}^{n} w_i x_i} = \sum w_i x_i$

- $\frac{\partial (w^T x)}{\partial x} = w$.

- $\frac{\partial (x^T A x)}{\partial x} = Ax + A^T x$.

$x, y \in \mathbb{R}^n$, $\quad Q \in \mathbb{R}^{n \times n}$. $\quad l : \mathbb{R}^n \to \mathbb{R}$.

$$
\begin{aligned}
l(x) &= \frac{1}{2}(Qx-y)^T(Qx-y) \\
&= \frac{1}{2}(x^TQ^T - y^T)(Qx-y) \\
&= \frac{1}{2}(x^TQ^TQx - \underline{y^TQx} - \overline{x^TQ^Ty} + y^Ty) \\
&= \frac{1}{2}x^TQ^TQx - y^TQ\cdot x + \frac{1}{2}y^Ty \\
\nabla_x l(x) &= \frac{1}{2}\cdot\frac{\partial(x^TQ^TQx)}{\partial x} - \frac{\partial(y^T\cdot Q\cdot x)}{\partial x} + 0. \\
&= \frac{1}{2}\cdot\left(Q^TQ + (Q^TQ)^T\right)\cdot x - Q^Ty \\
&= Q^TQ\cdot x - Q^Ty
\end{aligned}
$$

15:35.

III. Probability.

E_1 , E_2.

↓ head of coin ↓ tail of coin.

$\mathbb{P}(E_1) = \frac{1}{2} = \mathbb{P}(E_2) > 0.$

$\sum_{i=1}^{n} \mathbb{P}(E_i) = 1$

i) $P(E) \in [0,1]$

ii) $\sum_{E} P(E) = 1 = P(\Omega)$

When. E_1, E_2 are disjoint., i.e. $E_1 \cap E_2 = \phi$.

- $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$.

- random variable.

$X = x$

- Ex. Bernolli dist.

$x = 1$; head

$x = 0$; tail.

$\mathbb{P}(x) \to P_X(x)$.

- Events A, B. , $\mathbb{P}(A) > 0$.

$$\mathbb{P}(B|A) \triangleq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

$\to$ $\mathbb{P}(A \cap B) = \mathbb{P}(B|A) \cdot P(A)$.

$p(x,y) = p(x|y) \cdot p(y)$

$$\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \mathbb{P}(A_1, A_2, A_3 | A_4) \cdot \mathbb{P}(A_4).$$

- $P(y \mid x ; \theta) = P(y \mid x)$

function of θ.

- Expectation.

$x \in \mathcal{X} = \{1, \cdots, n\}$.

$\mathbb{E}[X] \triangleq \sum_{i=1}^{n} i \cdot P_X(i)$. or $\mathbb{E}[X] \triangleq \sum_{x} x \cdot P_X(x)$.

- $\mathbb{E}[X^2] = \sum_{x} x^2 \cdot P_X(x)$. $\rightarrow$ $\mathbb{E}[f(X)] = \sum_{x} f(x) \cdot P_X(x)$.

e.g. $X = 0, 1$.

$P_X(0) = 1-p$, $P_X(1) = p$, $p \in (0,1)$.

$\mathbb{E}[X] = 0 \cdot P_X(0) + 1 \cdot P_X(1) = p$

- $\mathbb{E}[Y \mid X = x] \triangleq \sum_{y} y \cdot P_{Y|X}(y|x)$.

e.g. $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$.

$$\mathbb{E}[Y \mid X = x]$$

$\hookrightarrow$ it is a function of x.

$$\mathbb{E}[\,\mathbb{E}[X|Y]\,] = \mathbb{E}[X].$$

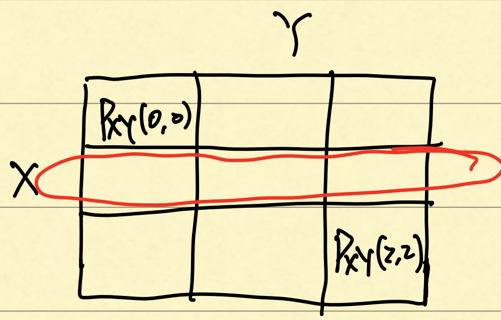
① $\mathbb{E}[X|Y=y] = \sum_x x \cdot P_{X|Y}(x|y)$.

$$\mathbb{E}\Big[\underbrace{\sum_x x \cdot P_{X|Y}(x|y)}_{g(y)}\Big] = \sum_y g(y) \cdot P_Y(y).$$

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

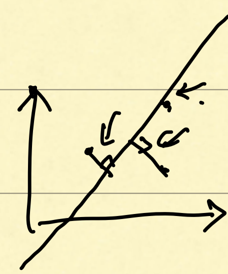
$$= \sum_{x,y} x \cdot P_{XY}(x,y).$$

$$= \sum_x x \cdot \Big(\underline{\sum_y P_{XY}(x,y)}\Big) = \sum_x x \cdot P_X(x) = \mathbb{E}[X].$$



· Covariance. X, Y.

$$\mathrm{Cov}(X,Y) \triangleq \mathbb{E}\big[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\big]$$



X :	0,	1,	2.
$P_X(x)$:	$\frac{1}{2}$.	$\frac{1}{6}$,	$\frac{1}{3}$.

$$\mathbb{E}[X] = 0 \times \frac{1}{2} + 1 \times \frac{1}{6} + 2 \times \frac{1}{3} = \frac{1}{6} + \frac{4}{6} = \frac{5}{6}$$

$$\mathrm{Var}(X) = \left(0 - \frac{5}{6}\right)^2 \cdot \frac{1}{2} + \left(1 - \frac{5}{6}\right)^2 \cdot \frac{1}{6} + \left(2 - \frac{5}{6}\right)^2 \cdot \frac{1}{3}.$$

$$\mathrm{Cov}(X,Y) \triangleq \mathbb{E}\big[(X-\mathbb{E}[X])\cdot(Y-\mathbb{E}[Y])\big]$$

$$= \mathbb{E}\big[X\cdot Y - \mathbb{E}[X]\cdot Y - X\cdot\mathbb{E}[Y] + \mathbb{E}[X]\cdot\mathbb{E}[Y]\big].$$

$$= \mathbb{E}[XY] - \mathbb{E}[X]\cdot\mathbb{E}[Y] \quad \Leftarrow$$

• Independent.

X., $P_X(x)$., $f_X(x)$., pdf of X.

$$f_{XY}(x,y) = f_X(x) f_Y(y). \qquad , X,Y$$

①. $X \perp\!\!\!\perp Y \Rightarrow \mathrm{Cov}(X,Y)=0.$

$$\Downarrow$$

$f_{XY}(x,y) = f_X(x)\cdot f_Y(y).$ $\quad \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0.$

$$\mathbb{E}[XY] = \mathbb{E}[X]\cdot\mathbb{E}[Y].$$

$$\sum_{x,y} x\cdot y \, P_{XY}(x,y) = \sum_x \sum_y x\cdot y \, P_X(x) P_Y(y).$$

$$X \perp\!\!\!\perp Y \Longleftrightarrow \forall f.g \quad \mathrm{Cov}(f(X).g(Y))=0.$$

$\because$ $\mathrm{Cov}(f(X).g(Y))=0.$

$\therefore$ $\mathbb{E}[f(X)\cdot g(Y)] = \mathbb{E}[f(X)]\cdot\mathbb{E}[g(Y)].$

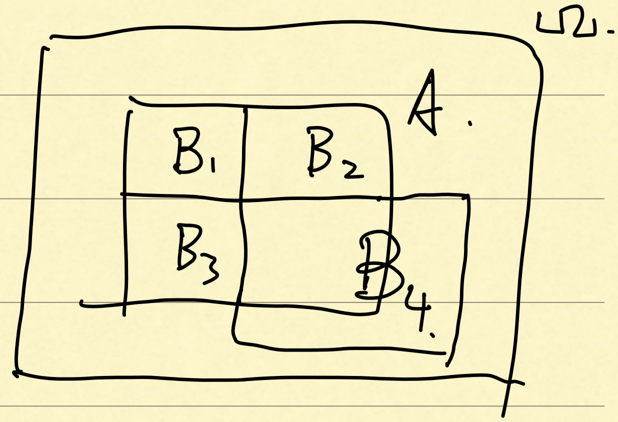
$$\sum_{x,y} f(x)\cdot g(y)\cdot P_{XY}(x,y) = \sum_{x,y} f(x)\cdot g(y)\cdot P_X(x)\cdot P_Y(y). \quad \checkmark$$

Set. $f(x) \triangleq \begin{cases} 1, & X=0 \\ 0, & o.w \end{cases}$, $g(Y) \triangleq \begin{cases} 1, & Y=0 \\ 0, & o.w \end{cases}$

Want: $\forall x', y'$, $P_{XY}(x', y') = P_X(x') P_Y(y')$.

$\rightarrow P_{XY}(0,0) = P_X(0) \cdot P_Y(0)$.

- $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$.
- $P(A) = \sum_{i=1}^{4} P(B_i) \cdot P(A|B_i)$.



- Data.

$$P(B_1|A) = \frac{P(A,B_1)}{P(A)}$$

label.

$$= \frac{P(A|B_1) \cdot P(B_1)}{P(A)}$$

$$= \frac{P(A|B_1) \cdot P(B_1)}{\sum_{1}^{k} P(A|B_k) \cdot P(B_k)}$$

- Bernoulli dist.

$X=0$, $1-p$. $\quad E[X] = p$.

$X=1$, p $\quad Var(X) = p(1-p)$.

$X \sim Bern(p)$.

· Gaussian Dist.

$X \sim \mathcal{N}(\mu, \sigma^2)$. $\mu, \sigma \in \mathbb{R}$

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$