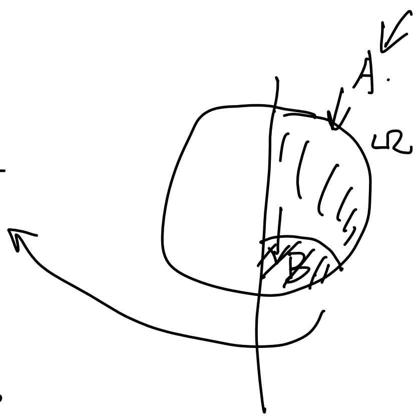
Probability.

Event 'E'.

$P\{E\} \in [0,1]$. $E = \phi$, $P\{\phi\} = 0$, $P\{\Omega\} = 1$.

- Conditional Probability:
  - A, B.

    $P\{B \mid A\} \triangleq \dfrac{P\{A \cap B\}}{P\{A\}}$.



  - r.v.s. X & Y. x, y.

    $P_{Y|X}(y|x;\theta) = P_{Y|X}(y|x)$.

    θ ↳ parameter.

  - Expectation
    - $\mathbb{E}[X] \triangleq \sum_{x \in \mathcal{X}} \underset{\text{value}}{x} \cdot \underbrace{P_X(x)}_{\text{probability}}$.
    - $\mathbb{E}[Y|X=x] \triangleq \sum_{y \in \mathcal{Y}} y P_{Y|X}(y|x) = g(x)$.
    - $\mathbb{E}_X[\mathbb{E}_{Y|X=x}[Y|X]] = \mathbb{E}_X[g(x)] = \sum_{x \in \mathcal{X}} g(x) \cdot P_X(x) = \mathbb{E}[Y]$.

  - r.v.s X, Y

    $\mathrm{Var}(X) \triangleq \mathbb{E}[(X - \mathbb{E}[X]) \cdot (X - \mathbb{E}[X])] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$.

    $\mathrm{Var}(\underline{u}) \triangleq \mathbb{E}[(X - \mathbb{E}[X]) \cdot (X - \mathbb{E}[X])^T]$

$$\mathrm{Var}\left(\begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix}\right) \neq \begin{bmatrix} \mathrm{Var}(u_1) \\ \vdots \\ \mathrm{Var}(u_d) \end{bmatrix}, \quad \begin{bmatrix} \mathrm{Var}(u_1) & & \mathrm{cov}(u_1, u_d) \\ & \ddots & \\ & & \mathrm{Var}(u_d) \end{bmatrix}$$

$\operatorname{cov}(X,Y) \triangleq \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

- independ $\longrightarrow$ $P_{XY}(x,y) = \underbrace{P_X(x)}_{}P_Y(y)$., $\quad P_{XY}(x,y) = P_{Y|X}(y|x)\cdot P_X(x)$. $\quad P_Y(y) = P_{Y|X}(y|x)$, if. $X \perp Y$.

uncorrelated. $\quad \operatorname{cov}(X,Y) = 0$.

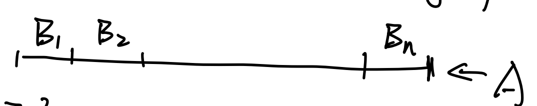
Prove : X, Y. ind $\to$ uncorrelated.

uncorr $\to$ ind.

$y = f(x)$

↳ linear

- Total Probability Rule : $\quad A = B_1 \cup \cdots \cup B_n \quad B_i \cap B_j = \phi$.



$P\{A\} = P\{A\cap B_1\} + \cdots + P\{A \cap B_n\}$.

$= \int P\{A \cap x\}\, dx$.

- Bayes Rule. $\quad$ Mnist. $\boxed{0} \sim \boxed{9}$ $\quad$ 10 classes.

0 – 9 : labels.

$P\{B_1|A\}. = \dfrac{P\{A\cap B_1\}}{P\{A\}} = \dfrac{P\{A|B_1\}\cdot P\{B_1\}}{P\{A\}}$.

Label 1. $\boxed{3}$

Label 0 $\sim$ Label 9.

rank. $\begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$

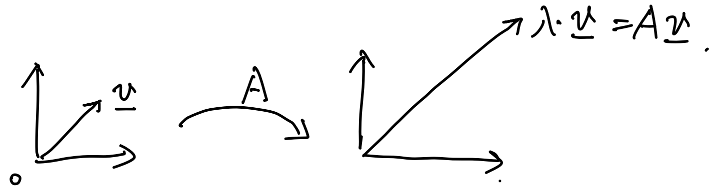
Gaussian. $\quad$ easy.

$\underset{B_i}{\arg\max}\ P\{B_i|A\}. = \underset{B_i}{\arg\max}\ P\{A|B_i\}\, P\{B_i\}$

1. Eigenvalue Decomposition.

Def 1. (Eigenvalue). $A \in \mathbb{R}^{n \times n}$, we say that $\lambda \in \mathbb{R}$ is an eigenvalue of A, if there exists some non-zero vector $\underline{v} \in \mathbb{R}^n \setminus \{0\}$

s.t. $$A\underline{v} = \lambda \cdot \underline{v}$$



$\langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & \text{o.w} \end{cases}$

Thm 1. (Eigv. Decomposition). Let $A \in \mathbb{R}^{n \times n}$ be symmetric, i.e. $A^T = A$. There exists orthonomal vectors $\underline{v}_1, \cdots, \underline{v}_n \in \mathbb{R}^n$, real scalars $\lambda_1 \geq \cdots \geq \lambda_n$, s.t.

$V \triangleq [\underline{v}_1 | \cdots | \underline{v}_n] \in O(n)$.

$\Lambda \triangleq \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \in \mathbb{R}^{n \times n}$.

then
$$A = V \cdot \Lambda \cdot V^T = A^T$$

Remark:

① $V^T V = I$.

$(V^T V)_{ij} = \langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1 & , j = i \\ 0 & , \text{o.w.} \end{cases}$

② $A = V \cdot \Lambda \cdot V^T = [v_1 | \cdots | v_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$

$= \sum_{i=1}^{n} \lambda_i \cdot v_i v_i^T$.

2. Singular Value Decomposition (SVD).

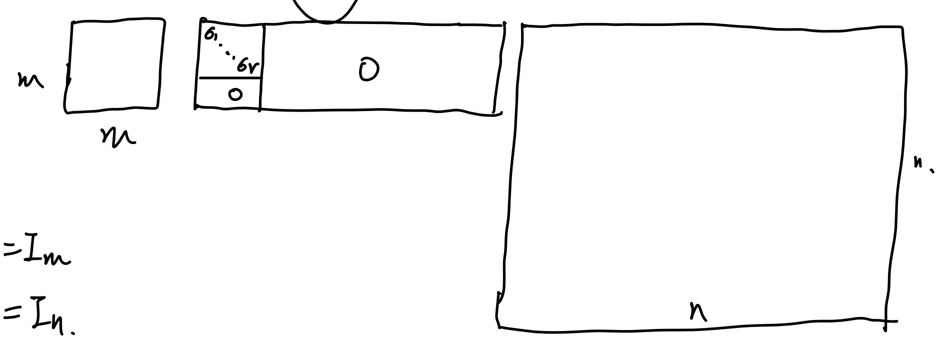
Thm2: $A \in \mathbb{R}^{m\times n}$, $\exists\ U \in \mathbb{R}^{m\times m} \in O(m)$, $V \in \mathbb{R}^{n\times n} \in O(n)$

s.t.

$$A = U\cdot\Sigma\cdot V^T,$$

where $\Sigma \triangleq \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$, $\Sigma_1 \triangleq \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$. $r = \text{rank}(A)$.

Remark: ① $A = U \cdot \Sigma \cdot V^T$



$U^TU = I_m$

$V^TV = I_n$.

② $A = U\cdot\Sigma V^T$

$$= [\underline{u}_1 | \cdots | \underline{u}_m] \cdot \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ & 0 & & 0 \end{bmatrix} \begin{bmatrix} \underline{V}_1^T \\ \vdots \\ \underline{V}_n^T \end{bmatrix}.$$

$$= \sum_{i=1}^{r} \sigma_i \cdot \underline{u}_i \cdot \underline{V}_i^T$$

③ $A\cdot\underline{V}_k = \left(\sum_{i=1}^{r} \sigma_i \cdot \underline{u}_i \cdot \underline{V}_i^T\right)\cdot \underline{V}_k$, $k \le r$

$= \sigma_1 \underline{u}_1 \cdot \underline{V}_1^T \cdot \underline{V}_k + \cdots + \sigma_r \cdot \underline{u}_r \cdot \underline{V}_r^T \cdot \underline{V}_k$

$= 0 + \cdots + \sigma_k \underline{u}_k + \cdots + 0$

$= \sigma_k \cdot u_k$.

$\langle V_i, V_k\rangle = \delta_{ik} = \begin{cases} 1, & i=k \\ 0, & \text{o.w.} \end{cases}$

④ $A\cdot A^T = U\cdot\Sigma\underbrace{(V^T\cdot V)}_{I}\Sigma^T\cdot U^T = U\cdot\Sigma\Sigma^T U^T$.

$\perp$

$$A \cdot A^T \cdot A \cdot A^T = U \cdot (\Sigma \Sigma^T)^2 U^T.$$

Power Iteration

Alg.

⑤. $\|A\|_F^2 \triangleq \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2$

$= tr(A^T A).$ (prove it!)

$= tr(V \cdot \Sigma^T \cdot \Sigma \cdot V^T).$ ($tr(AB) = tr(BA)$).

$= tr(\underbrace{V^T V}_{= I} \Sigma^T \Sigma) = tr(\Sigma^T \Sigma). = \sigma_1^2 + \ldots + \sigma_r^2$