

• Term project:

- project proposal
- term project report
- online/offline project presentation

- vector X

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

$$X = [x_1, x_2, \dots, x_n]^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Matrix X

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

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_m^T \end{bmatrix} \in \mathbb{R}^{m \times n}$$

x_i : \Rightarrow i^{th} row of X
(x_i column)

x_{ij} \Rightarrow j^{th} col of X

- Tensor > 2 axis

element-wise product

$$C_{ij} = A_{ij} B_{ij}$$

- matrix product

$$C_{ij} = \sum_k A_{ik} B_{kj}$$

$$\begin{bmatrix} \blacksquare \end{bmatrix} = \begin{bmatrix} \text{row } i \end{bmatrix} \cdot \begin{bmatrix} \text{col } j \end{bmatrix}$$

- Inverse

$$A^{-1}A = I$$

- norm

vector norm L_p $p \geq 1$

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p}$$

e.g. $p=1$

$$\|x\|_1 = \sum_i |x_i|$$

$$\|x\|_2 = \sqrt{\sum_i x_i^2}$$

$$\|x\|_\infty = \max_i |x_i|$$

not valid but used in ML
e.g. $x = [0, 1, 0, 3, 2, 0]^T$

$$\|x\|_0 = 3$$

e.g. 2 dimension vector

$$L_1 \text{ norm} = 1$$

$$\|x\|_1 = 1$$

$$x_1 + x_2 = 1$$



$$\begin{matrix} x_1 > 0 \\ x_2 > 0 \\ x_1 < 0 \\ x_2 < 0 \end{matrix}$$

$$L_2 \text{ norm} = 1$$



FACT

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2 \leq n \|x\|_\infty$$

Proof

$$\|x\|_\infty = \max_i |x_i| = \max_i \sqrt{|x_i|^2} = \sqrt{\frac{1}{n} \sum_i |x_i|^2} = \|x\|_2$$

• Matrix Norm

$$\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2} \quad \text{L2 norm matrix}$$

(Frobenius norm)

λ = largest eigenvalue

L_2 norm matrix

$$\|A\|_2 = \sqrt{\lambda_{\max}(A)}$$

• vector inner product

$$\langle x, y \rangle = \|x\|_2 \|y\|_2 \cos \theta$$

if $\|x\|_2 = 1$

$$\langle x, y \rangle = \|x\|_2 \cos \theta \quad (\text{projection on } y \text{ space})$$

• Holder's inequality

$$|\langle x, y \rangle| \leq \|x\|_p \|y\|_q \quad \text{s.t. } \frac{1}{p} + \frac{1}{q} = 1$$

e.g. $p=2, q=2$
 $p=1, q=\infty$

• Linear combination

$$Ax$$

• Linear dependence, span

orthogonal / orthonormal: all column vector are in "inner-product = 0" relations:

orthogonal: $x^T y = 0$

$$\|x\|_2 = \|y\|_2 = 1$$

orthogonal matrix: $A^T A = I$
 $A^T = A^{-1}$

• eigenvalue decomposition

$$A = Q \Lambda Q^T$$

(square matrix)

Q, Q^T are orthogonal

Λ is diagonal w/ λ values

• quadratic form:

$$x^T A x = x^T Q \Lambda Q^T x = (\Lambda x)^T (\Lambda x)$$

$$= \sum_{i=1}^n \lambda_i (\lambda_i^T x)^2 \leq \lambda_i \sum_{i=1}^n (\lambda_i^T x)^2 = \lambda_i \|x\|_2^2$$

(computational approach: QR decomposition)

• Positive definite

$$\forall x, x^T A x > 0$$

$$SPD$$

$$\forall x, x^T A x \geq 0$$

quadratic form w/ PD or SPD:

$$0 \leq x^T A x \leq \lambda_1 \|x\|_2^2$$

\hookrightarrow helpful sometimes

• SVD (singular vector decomposition)

don't hv to be square matrix

$$A = U D V^T \rightarrow \begin{bmatrix} \cdot & \cdot & \cdot \end{bmatrix} \text{ diagonal}$$

$$U U^T = I$$

$$V V^T = I$$

could be used in more cases

• Derivative

- chain rule

$$f(x) = h(g(x))$$

$$\frac{df}{dx} = \frac{dh}{dg} \frac{dg}{dx}$$

- $f(x) \in \mathbb{R}^1$

$$x \in \mathbb{R}^n$$

$$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

• trace =

$$\sum_{i=1}^n A_{ii} \quad (\text{tr}(A))$$

• Jacobian

- $f(x) \in \mathbb{R}^{n \times 1}$

$$x \in \mathbb{R}^1$$

$$\frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_2}{\partial x} & \dots & \frac{\partial f_n}{\partial x} \end{bmatrix}$$

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i$$

$$\text{e.g. } f = a^T x b$$

$$\frac{\partial f}{\partial x} = ?$$

$$df = (da^T) x b + a^T (dx) b + a^T x (db)$$

$$df = \text{tr}(df)$$

$$df = a^T (dx) b$$

$$= \text{tr}(a^T (dx) b)$$

$$= \text{tr}(b a^T (dx))$$

$$= \text{tr}[(b a^T)^T dx]$$

$$\therefore \frac{df}{dx} = a b^T$$

• Probability

- random variable

Discrete: PMF (mass-function)

PDF (density-function)

- marginal probability

- conditional probability

- independence

- expectation variance covariance

- different kinds distribution

- Bayes Rules