- **Term project :**
  - project proposal
  - term project report
  - online/offline project presentation

- vector X

  $X \in R^n$

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

- Matrix X

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

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

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

  $x_{:j}$ $\Rightarrow$ j^{th} col of X

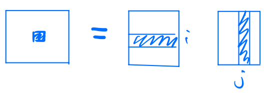
- Tensor >2 axis

  element-wise product

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

- matrix product

  $C_{ij} = \sum_k A_{i,k} B_{k,j}$



- Inverse

  $A^{-1} A = I$

- norm

  vector norm L_p $\quad p \geq 1$

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

  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 — $\|x\|_0$ = refer to # of non zero element

  e.g. $x = [0, 1, 0, 3, 2, 0]^T$

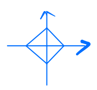
  $\|x\|_0 = 3$

  e.g. 2 dimension vector

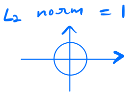
  L_1 norm $= 1$

  $\|x\|_1 = 1$

  $x_1 + x_2 = 1$ $\quad x_1 > 0,\ x_2 > 0$; $\ x_1 > 0,\ x_2 < 0$; $\ \vdots$



  L_2 norm $= 1$



FACT

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

Proof

$$\|x\|_\infty = \max_i |x_i| = \max_i \sqrt{|x_i|^2} \leq \sqrt{\sum_{i=1}^n |x_i|^2} = \|x\|_2$$

- Matrix Norm

  $\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2} \neq L_2$ norm matrix

  (Frobenius norm)

  σ = largest eigen value

  L_2 norm matrix $\neq$ $\|A\|_2 = \sigma_{max}(A)$

  $$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

- vector inner product

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

  if $\|y\|_2 = 1$

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

- Holder's inequality

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

  e.g. $p=2, q=2$

  $p=1, q=\infty$

- remark: norm will be frequently used for regularization

- Linear combination

  Ax

- Linear dependence, span

- orthogonal / orthonormal :
  - all column vector are in "inner-product $=0$" relations :
  - othogonal : $x^T y = 0$

    $\|x\|_2 = \|y\|_2 = 1$
  - othogonal matrix : $A^T A = I$

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

- eigenvalue decomposition

  $A = Q \Lambda Q^T$ $\quad$ (square matrix)

  Q, Q^T are orthogonal

  Λ is diagonal w/ λ-values

- quadratic form :

  $$\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \\ & & & \lambda_n \end{bmatrix}$$

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

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

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

  $\leq \lambda_i \|x\|_2^2$

  (computational approach : QR decomposition)

  bounded by λ_i & $\|x\|_2^2$

- Positive definite

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

  SPD

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

  $\therefore$ quadratic form w/ PD or SPD :

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

  $\rightarrow$ helpful sometimes

- SVD (singular vector decomposition)

  don't hv to be square matrix

  $A = U D V^T$ $\quad D \rightarrow [\ddots]$ diagonal

  $U U^T = I$

  $V V^T = I$

  could be used in more cases

- Derivative
  - chain rule

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

    $\frac{df}{dx} = \frac{dh}{dg}\frac{dg}{dx}$
  - $f(x) \in R^1$

    $x \in R^n$

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

- trace =

  $\sum_{i=j}^{n} A_{ij} \quad (tr(A))$

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

    $x \in R^1$

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

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

  e.g. $f = a^T x b$

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

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

  ($da^T = 0$, $db = 0$)

  $df = tr(df)$

  $$\begin{aligned} df &= a^T (dx) b \\ &= tr(a^T (dx) b) \\ &= tr(b a^T (dx)) \\ &= tr[(ab^T)^T dx] \end{aligned}$$

  $\therefore \frac{df}{dx} = ab^T$

- Probability
  - random variable

    Discrete : PMF (mass-function)

    PDF (density-function)
  - marginal probability
  - conditional probability
  - independence
  - expectation variance covariance
  - different kinda distribution
  - Bayes Rules