

CS182 Introduction to Machine Learning

Recitation 2

2025.3.5

Outline

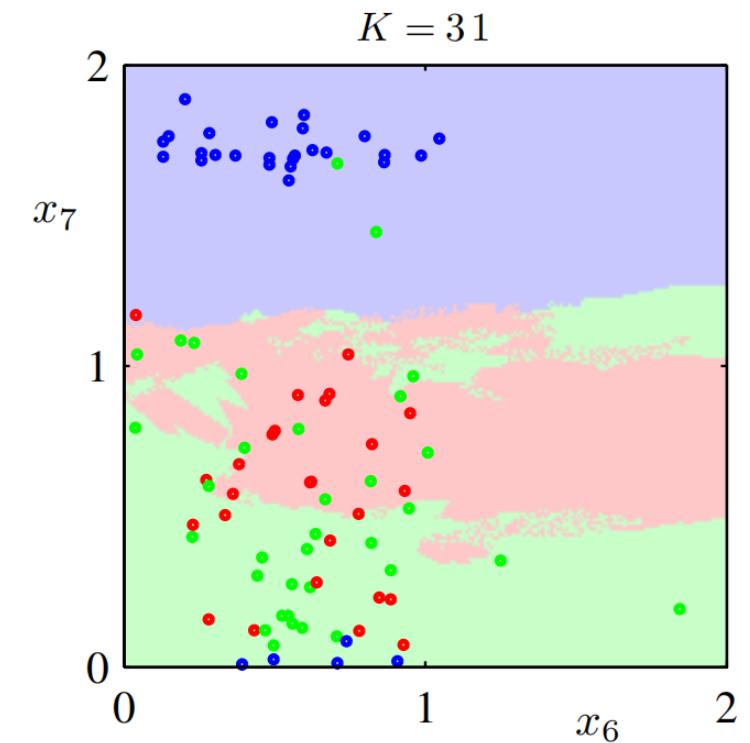
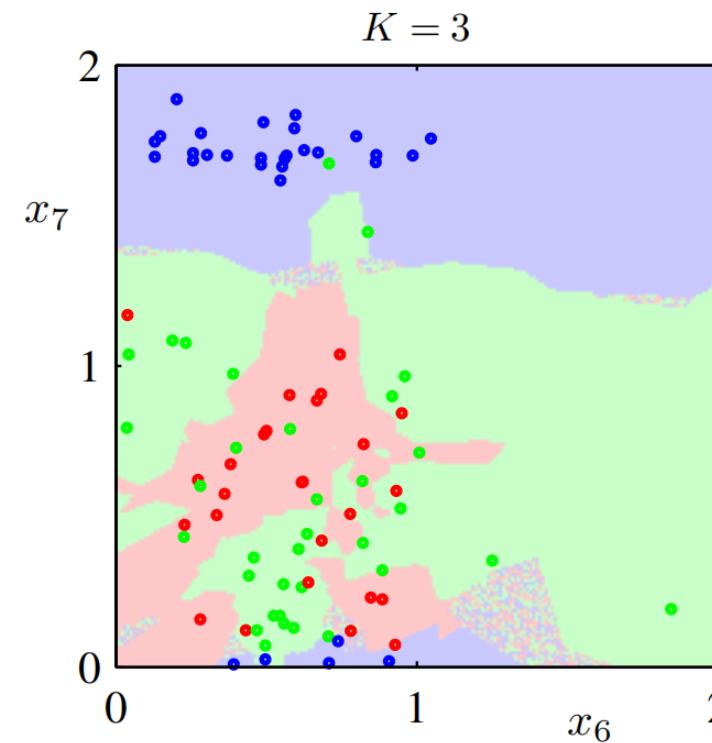
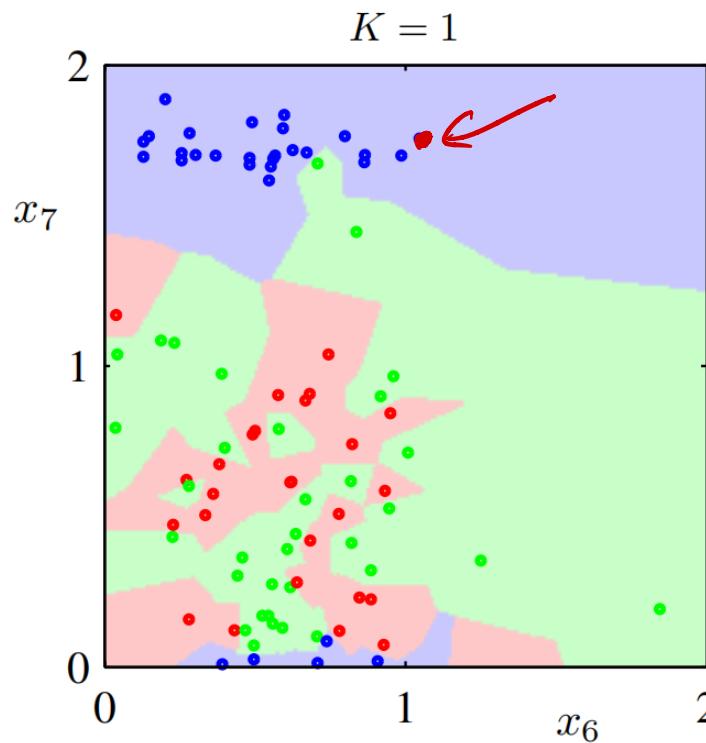
- KNN
- Curse of Dimensionality
- Review(Preview): Linear Algebra

$$\{x_i\}_{i=1}^N$$

$$x_i \in \mathbb{R}^d$$

$$d=2 \uparrow \uparrow$$

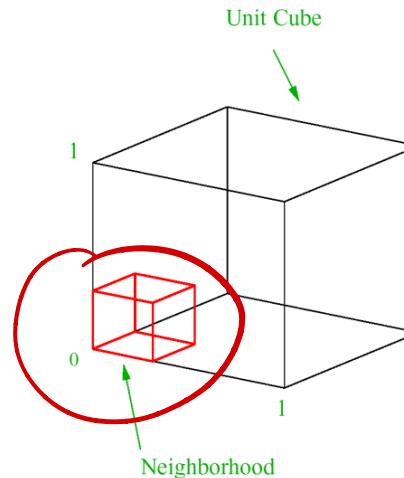
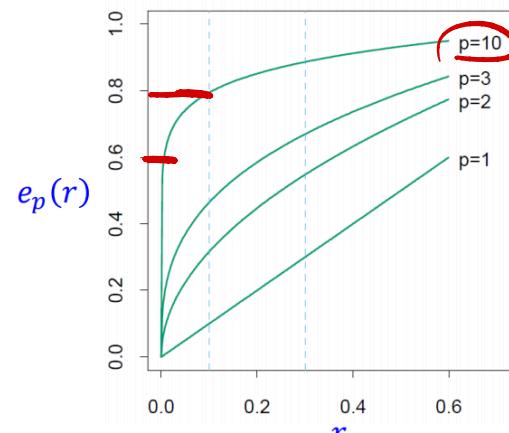
KNN (K-Nearest Neighbors)



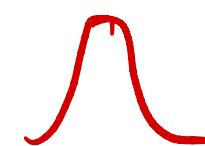
Curse of Dimensionality (维度诅咒)

Local neighborhoods become increasingly global, as the p increases

p : 维度数, e : 边长, $r = \frac{e^p}{1}$, $e_p(r) = r^{\frac{1}{p}}$



$$r = \frac{e^P}{l^P}$$



$$\frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}$$

Reducing r reduces the number of observations and thus the stability.

In ten dimensions we need to cover 63% (80%) of the range of each coordinate to capture 1% (10%) of the data.

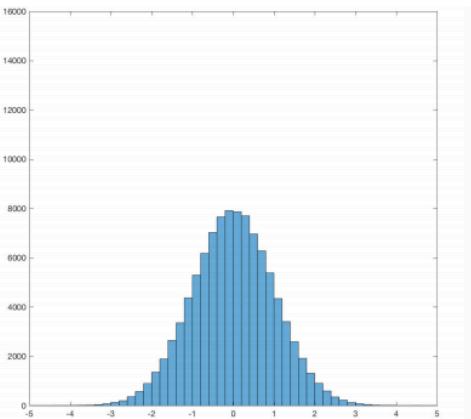
高维情况下, 距离, cosine similarity等度量方式失效.

e.g. 1维, 2维 Gaussian distribution PDF集中在均值附近, 但是在高维空间中, 大部分数据点都位于边界附近

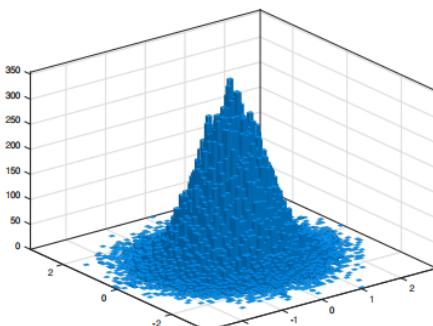
$$x \sim N_p(\mu, \Sigma)$$

Curse of Dimensionality (维度诅咒)

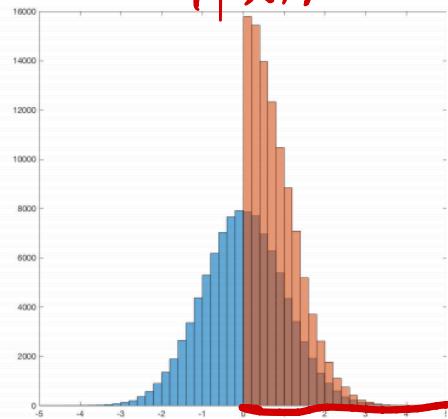
distribution of a one-dimensional Gaussian vector $x \in \mathbb{R}^1$ and of its length $\|x\| \in \mathbb{R}$,
with 100.000 samples



distribution of a two-dimensional Gaussian vector $x \in \mathbb{R}^2$ and of its length $\|x\| \in \mathbb{R}$,
with 100.000 samples

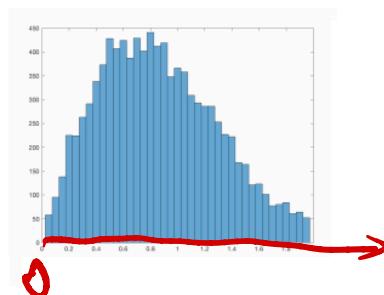
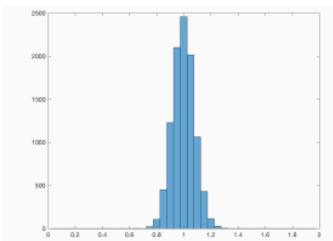
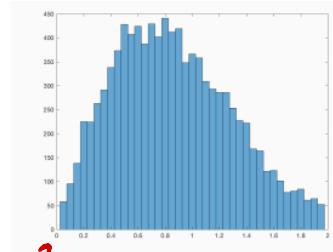


$$\|x\|$$



the length $\|x\| \in \mathbb{R}$ of a Gaussian vector $x \in \mathbb{R}^p$, for $p = 1, 2, 5, 20, 80, 320$;
with 10.000 samples in each case

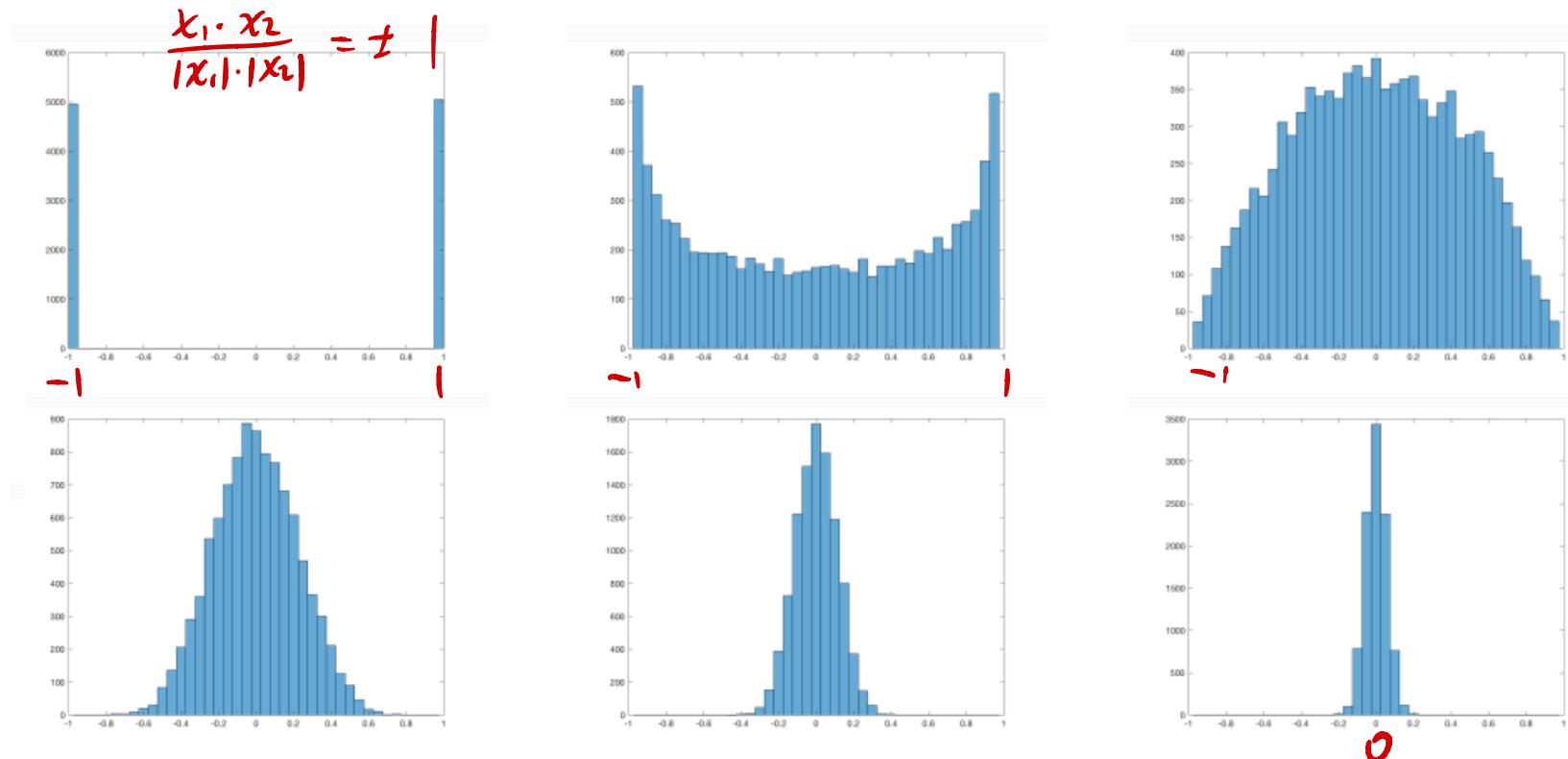
$$\|x\|$$



Curse of Dimensionality (维度诅咒)

Two Gaussian random vectors are nearly orthogonal in high dimensions

For 10.000 samples of pairs (x_1, x_2) of independent Gaussian vectors $x_1, x_2 \in \mathbb{R}^p$, the histogram of the normalized inner product $\frac{\langle x_1, x_2 \rangle}{\|x_1\| \cdot \|x_2\|}$ is shown for $p = 1, 2, 5, 20, 80, 320$. This shows that two such vectors are close to orthogonal in high dimensions. Note that 0 is, also in small dimensions, the expectation of the normalized inner product.



外差(Extrapolation), 内插(Interpolation).

外差是指在已知数据范围之外做预测的过程。在高维情况下，由于数据点之间的距离普遍较大，模型往往需要在大量空白或未探索的空间中进行预测，这通常会导致预测不准确。

内插是指在已知数据点之间进行预测的过程。在低维空间，数据点比较密集，内插通常可以较好地进行。然而在高维空间，即使是最近的邻居也可能相距甚远，这使得内插的效果大打折扣。

Review(Preview): Linear Algebra

- 为什么用到线性代数:
 - 线性代数是描述空间和变换的工具，让描述问题变得简单
 - 大量学习算法通过建模输入空间到输出空间的变换来解决问题
 - 线性代数的矩阵分解理论提供了寻找主成分的理论基础
- 用哪些线性代数:
 - 矩阵的基本运算和性质(回忆一下特殊矩阵：对称矩阵、对角矩阵、单位矩阵、正交矩阵、上三角矩阵)
 - 常用的两种矩阵分解: 特征值分解、SVD分解
 - 最小二乘法
 - 矩阵求导*(由于将向量记作行向量还是列向量有分歧，因此有两套矩阵求导公式，请注意如果没有特殊说明，我们均默认列向量)

linear algebra in CS

你们觉得你们现在学的东西没有用，并不是因为它真的没有用，只是你们还没有遇到要用到那门课的时候

linear algebra in CS

- 一个CS的学生, 在大学期间其实会多次重新学习线性代数
- 学习过的内容不需要你牢牢记住, 但是需要你知道它的存在, 以及它的用途, 以及清楚的知道你需要的时候去哪里找

linear algebra in CS



Tweet



貪心不足
@Tanxinbuzu

...

「大学招生简直诡计多端...某个uc数学系的博士生说，
现在愿意修线性代数的美国本科生越来越少了，后来他
们改了个名字大概叫Math Foundations of
MachineLearning，爆满，学生给教授写邮件要去上这
门课，其实还是线性代数。」

[Translate Tweet](#)

5:56 PM · 6/28/23 · 143K Views

147 Retweets 34 Quotes 947 Likes 77 Bookmarks



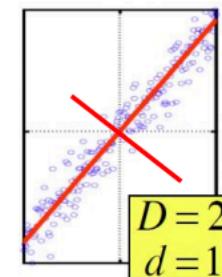
linear algebra in IML

Principal Component Analysis (PCA)

$(X X^T)v = \lambda v$, so v (the first PC) is the eigenvector of sample correlation/covariance matrix $X X^T$

Sample variance of projection $v^T X X^T v = \lambda v^T v = \lambda$

Thus, the eigenvalue λ denotes the amount of variability captured along that dimension (aka amount of energy along that dimension).



Eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots$

- The 1st PC v_1 is the eigenvector of the sample covariance matrix $X X^T$ associated with the largest eigenvalue
- The 2nd PC v_2 is the eigenvector of the sample covariance matrix $X X^T$ associated with the second largest eigenvalue
- And so on ...

linear algebra in IML

- for optimizations

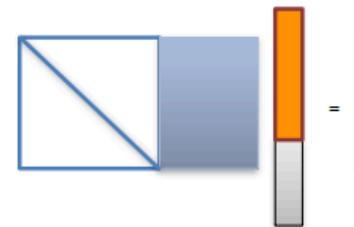
将 $Ax = b$ 的任一解 x 用非基变量表示为

$$x_1 = \bar{b}_1 - \sum_{j=m+1}^n y_{1j} x_j$$

$$x_2 = \bar{b}_2 - \sum_{j=m+1}^n y_{2j} x_j$$

\vdots

$$x_m = \bar{b}_m - \sum_{j=m+1}^n y_{mj} x_j$$



$$\begin{matrix} \textcolor{orange}{x_B} \\ + \end{matrix} \quad \begin{matrix} \textcolor{blue}{B^{-1}N} \\ (\textcolor{blue}{Y}) \end{matrix} \quad \begin{matrix} \textcolor{gray}{x_N} \\ = \end{matrix} \quad \begin{matrix} \textcolor{yellow}{B^{-1}b} \end{matrix}$$

$$\begin{matrix} \textcolor{orange}{x_B} \\ = \end{matrix} \quad \begin{matrix} \textcolor{yellow}{B^{-1}b} \\ - \end{matrix} \quad \begin{matrix} \textcolor{blue}{B^{-1}N} \\ (\textcolor{blue}{Y}) \end{matrix} \quad \begin{matrix} \textcolor{gray}{x_N} \end{matrix}$$

Review(Preview) Outline

- Trace, Transpose, Inverse, Symmetric, Determinant, Rank
- Quadratic Form
- Positive (Semi) Definite Matrix
- Orthogonal Matrix
- Eigenvalues Decomposition
- Singular Value Decomposition

Trace 迹

only for square matrix

$$\text{Tr}(A) = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i \leftarrow$$

$$A = P' \Lambda P$$

$$\text{Tr}(A) = \text{Tr}(P^{-1} \underline{\Lambda} P) = \text{Tr}(\underline{\Lambda} P P^{-1})$$

- $\text{Tr}(AB) = \text{Tr}(BA)$
- matrix inner product

$$\langle A, B \rangle = \text{Tr}(A^\top B) = \text{Tr}(B^\top A)$$

$$\text{Tr}(A^\top) = \overline{\text{Tr}}(A)$$

$$(A^\top B)^\top = B^\top A$$

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \underline{\mathbf{y}} = \underline{\mathbf{y}}^\top \mathbf{x}$$

- $\text{Tr}(aA + bB) = a \text{Tr}(A) + b \text{Tr}(B)$

Trace is a linear operator

Transpose 转置

$$(A^\top)_{ij} = (A)_{ji}$$

- $\underline{(AB)^\top} = \underline{B^\top A^\top}$
- $\overrightarrow{AA^\top}$ must be a symmetric matrix

$$(AB)^{-1} = \underline{\overleftarrow{B^{-1}A^{-1}}}$$

$$(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A$$

Inverse 逆

- nonsingular matrix \leftrightarrow invertible matrix

$$\underbrace{AB = BA}_{AB = I_n}$$

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

otherwise, A is singular and has no inverse

$$(|A| = 0)$$

$$[a]_{1 \times 1}$$

singular: 奇异的. 所以一个所有元素都随机的矩阵概率是可逆的

i.e. 奇异矩阵是不可逆矩阵

e.g. $A \in \mathbb{R}^{1 \times 1}$, A is singular if and only if $\underline{a_{11}} = 0$

Properties of Inverse

- inverse matrix of A is unique

B, C are A 's inverse matrices

$$B = BI_n = B(AC) = (BA)C = I_n C = C$$

- $(A^{-1})^{-1} = A$

- $(AB)^{-1}$ = $B^{-1}A^{-1}$

| proof: $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I_n$

- $(A^\top)^{-1}$ = $(A^{-1})^\top$ $\rightarrow A^{-T}$

$A^{-\top}$

| proof: $A^\top(A^{-1})^\top = (A^{-1}A)^\top = I_n$

Symmetric matrix 对称矩阵

- $A^\top = A$

all properties of symmetric matrix are based on this definition(this time)
(more properties later such as similarity and diagonalizable)

- $\forall A_{m \times n}, \underline{AA^\top} \text{ or } \underline{A^\top A}$ are symmetric matrix
- 全体对称矩阵的集合: \mathbb{S}^n

\mathbb{S}^n

S^n

S_{++}^n

Determinant 行列式

$$\begin{bmatrix} A & I \\ I & A^{-1} \end{bmatrix}$$

- a function mapping a matrix A into a scalar $\det(A)$ or $|A| = \prod_{i=1}^n \lambda_i$
- $A^{-1} = \frac{1}{|A|} A^*$
- $A^* = [C_{ij}]^\top$, C_{ij} is the cofactor of a_{ij}
- $C_{ij} = (-1)^{i+j} M_{ij}$, M_{ij} is the minor of a_{ij}
- the most simple usage: invertibility

Determinant Properties

compare with the elementary row(column) operations

1. B is obtained from A by interchanging two rows(columns)

$$|B| = -|A|$$

2. B is obtained from A by multiplying one row(column) by a nonzero scalar k

$$|B| = \cancel{k} |A|$$

3. B is obtained from A by adding a multiple of one row(column) to another row(column)

$$|B| = |A|$$

Determinant Properties

$$|A^\top| = |A|$$

$$\textcircled{|\lambda A| = \lambda^n |A|}$$

$$|AB| = |A||B|$$

$$|A^{-1}| = \frac{1}{|A|}$$

Triangular Matrix 上/下三角矩阵

- upper triangular matrix

the elements below the diagonal are all zero

(the elements on the diagonal can be zero or not)

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix},$$

then elements on the diagonal of A^k are $a_{11}^k, a_{22}^k, \dots, a_{nn}^k$

- similar to the lower triangular matrix

Triangular Matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

- $\det(A) = \prod_{i=1}^n a_{ii}$
- and the lower triangular matrix is the same

Diagonal Matrix 对角矩阵

- for diagonal matrix Λ , $\Lambda_{ij} = 0$ for $i \neq j$
so it can be written as $\Lambda = \text{diag}(d_1, d_2, \dots, d_n)$
- the power of diagonal matrix is easy to compute
 $\Lambda^k = \text{diag}(d_1^k, d_2^k, \dots, d_n^k)$
 $\overbrace{\quad\quad\quad}^{\rightarrow \text{similarity and diagonalizable}}$
- the diagonal matrix Λ is invertible if and only if $\forall i, d_i \neq 0$

$$|\Lambda| = \prod_{i=1}^n d_i$$

$$\begin{aligned}\underline{\Lambda^k} &= (\underline{P^{-1}} \underline{\Lambda} \underline{P})^k \\ &= P^{-1} \underline{\Lambda^k} P\end{aligned}$$

Row space, Column space and Null space

A is a $m \times n$ matrix

- row space 行空间

$$\text{row}(A) = \text{span}\{\mathbf{r}_1, \dots, \mathbf{r}_m\}$$

- column space 列空间

$$\text{col}(A) = \text{span}\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$$

- null space 零空间

$$\text{null}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$$

- left null space 左零空间

$$\text{null}(A^\top) = \{\mathbf{x} \in \mathbb{R}^m : A^\top \mathbf{x} = \mathbf{0}\}$$

Fundamental Matrix Spaces

Definition 4.31. 对 $m \times n$ -矩阵 A , 以下六个向量空间被称为 A 的基本空间(*the fundamental spaces of A*):

- A 的行空间 $\text{Row}(A)$,
- A 的列空间 $\text{Col}(A)$,
- A^\top 的行空间 $\text{Row}(A^\top)$,
- A^\top 的列空间 $\text{Col}(A^\top)$,
- A 的零空间 $\text{Null}(A)$,
- A^\top 的零空间 $\text{Null}(A^\top)$ 。

- 行空间和零空间互为正交补
- 列空间和左零空间互为正交补

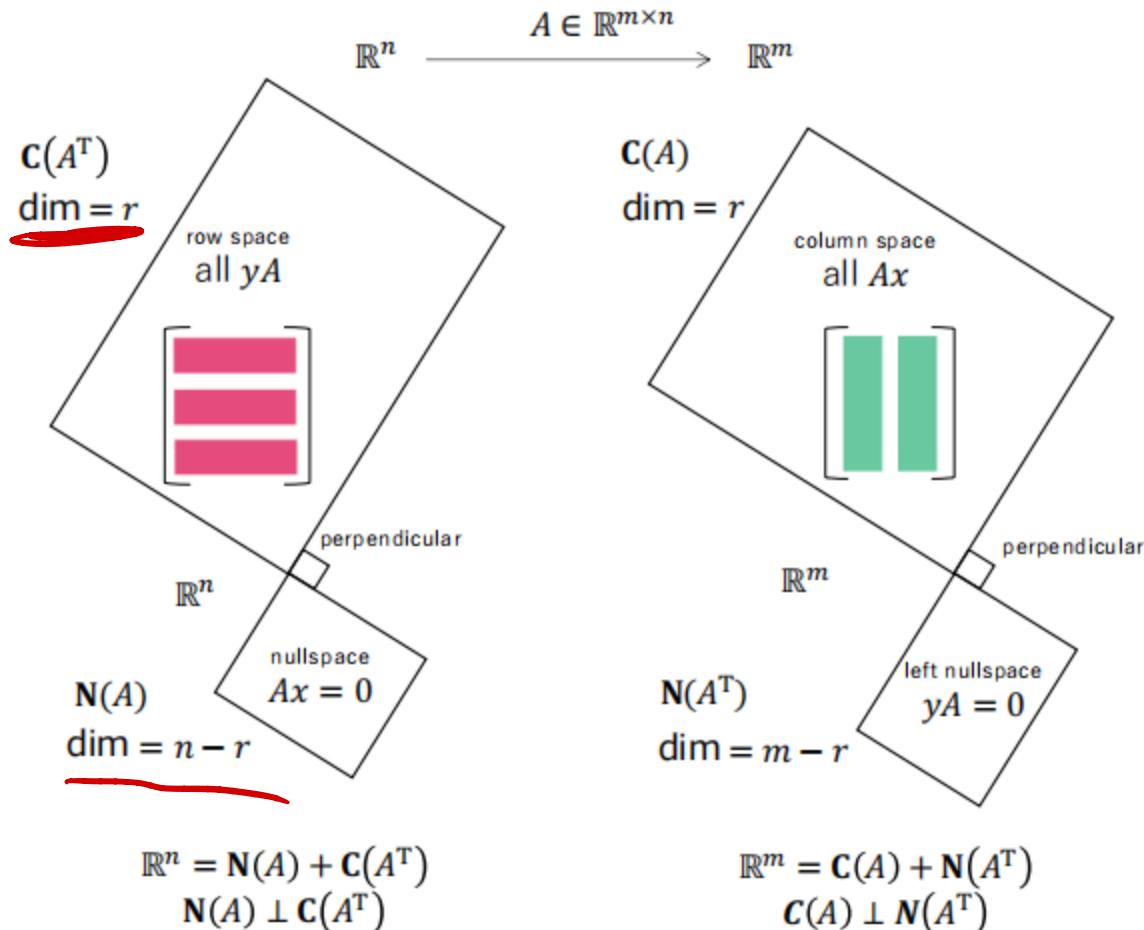
正交补(Orthogonal Complements)

正交: $\text{col}(A) \perp \text{null}(A)$, 补: $\text{col}(A) + \text{null}(A) = \mathbb{R}^n$

$$Ax = 0$$

$m \times n$

Row space, Column space and Null space



$$A_{m \times n}$$

$$\text{rank}(A) + \text{nullity}(A) = n$$

$$\Delta \quad \delta$$

Figure 5: 四个子空间

Rank, Nullity 秩, 零化度

- $\text{rank}(A) = \dim(\text{row}(A)) = \dim(\text{col}(A))$

A 的秩 = A 的行阶梯矩阵的首一个数

最本质: 非0奇异值的个数

$$\Rightarrow \text{rank}(A) \leq \min(n, m)$$

- $\text{rank}(A)$ 可看作行阶梯矩阵的首一(非零行/主元) 个数

$\text{nullity}(A) = \dim(\text{Null}(A))$ 可看作自由元的个数

$$\Rightarrow \text{rank}(A) + \text{nullity}(A) = n$$

$$A = \underbrace{P^{-1}}_0 \underbrace{\Lambda}_{\Delta} \underbrace{P}_0$$

$$\sigma_i \geq 0$$

$$\sigma_i = \sqrt{\lambda_i}$$

rank property

- $A \in \mathbb{R}^{m \times n}$
 $rank(A) \leq \min(m, n)$
- $rank(AB) \leq \min(rank(A), rank(B))$
- $rank(A^\top A) = rank(A)$

Equivalent expression

$A \in M_{n \times n}$

- A 可逆;
- $A\mathbf{x} = \mathbf{0}$ 只有平凡解;
- A 的简化阶梯型为单位矩阵;
- A 是一组初等矩阵的乘积;
- $A\mathbf{x} = \mathbf{b}$ 对任何 $n \times 1$ 的列向量 \mathbf{b} 都有解;
- $A\mathbf{x} = \mathbf{b}$ 对任何 $n \times 1$ 的列向量 \mathbf{b} 都有且只有一个解;
- $\det(A) \neq 0$;
- A 的所有 n 个行向量线性无关;
- A 的所有 n 个列向量线性无关;
- $\text{span}(\text{Row}(A)) = \mathbb{R}^n$;
- $\text{span}(\text{Col}(A)) = \mathbb{R}^n$;
- A 的所有 n 个行向量构成 \mathbb{R}^n 的一组基底;
- A 的所有 n 个列向量构成 \mathbb{R}^n 的一组基底;
- $\text{rank}(A) = n$;
- $\text{Null}(A) = \mathbf{0}$ 。

Norm 范数

满足:

- triangle inequality

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

when $p = 2$: triangle inequality

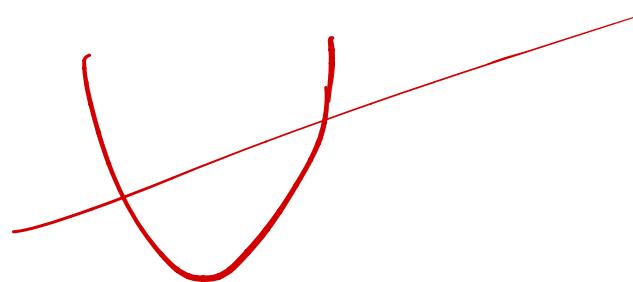
- $\|\mathbf{x}\| \geq 0, \|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
- $\forall a, \|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$

则 $\|\cdot\|$ 是一个范数

| all norm $f : \mathbb{R}^n \rightarrow \mathbb{R}$ are convex.

$$\forall x, y, \theta \in [0, 1]$$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$$



Norm

- p-norm of a vector $\mathbf{v} = (v_1, \dots, v_n)$
$$\|\mathbf{v}\|_p = \sqrt[p]{|v_1|^p + |v_2|^p + \dots + |v_n|^p}$$
- 0-norm: the number of nonzero entries in \mathbf{v}
0 – norm 不是范数 (non – convex)!

- 1-norm / L1-norm:

$$\|\mathbf{v}\|_1 = \underbrace{|v_1| + |v_2| + \dots + |v_n|}_{}$$

- 2-norm(Euclidean norm, L2-norm):

$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

- ∞ -norm:

$$\|\mathbf{v}\|_\infty = \underbrace{\max\{|v_1|, |v_2|, \dots, |v_n|\}}_{}$$

$$L = \sum \| \cdot \|_2^2 + \lambda \|\theta\|_2$$

$\rightarrow \lambda \|\theta\|_2$

Norm

$$p\text{-norm} : \|\mathbf{v}\|_p = \sqrt[p]{|v_1|^p + |v_2|^p + \cdots + |v_n|^p}, p \geq 1$$

- 0 norm is actually not a norm, but metric(度量)
- Hölder's inequality (赫尔德不等式)

$$p, q \geq 1, \frac{1}{p} + \frac{1}{q} = 1$$

$$\sum_{i=1}^n |a_i b_i| \leq \underline{\|x\|_p \|y\|_q}$$

$$|a^\top b| = |a| \cdot |b| \cdot |\cos \theta| \leq |a| \cdot |b|$$

$$\sum_{i=1}^n |a_i b_i| \leq \|x\|_1 \|y\|_\infty$$

- Cauchy inequality ($p = q = 2$)

$$\sum_{i=1}^n |a_i b_i| \leq \sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}$$

Norms for matrices

$$\left\{ \begin{array}{l} \|y - \hat{y}\|_2^2 \\ \|Y - \hat{Y}\|_F^2 \end{array} \right.$$

- p -norm: $\|A\|_p = \max_{\|\mathbf{x}\|_p \neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$
- 1-norm(列范数): $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$
- 2-norm(spectral norm, 谱范数): $\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)} = \sup_{\|\mathbf{x}\|_2 \neq 0} \frac{\|A\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}$
- ∞ -norm(行范数): $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$
- Frobenius norm: $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\text{Tr}(A^\top A)}$
- Nuclear norm(核范数): $\|A\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i(A)$

Dual Norm 对偶范数

范数 $\|\cdot\|$ 的对偶范数 $\|\cdot\|_*$ 定义为

$$\|\mathbf{z}\|$$



$$\|\mathbf{x}\|_* = \sup_{\|\mathbf{z}\| \leq 1} \langle \mathbf{x}, \mathbf{z} \rangle$$

e.g.

- $\|\mathbf{x}\|_1$ & $\|\mathbf{x}\|_\infty$
- $\|\mathbf{x}\|_2$ & $\|\mathbf{x}\|_2$
- $\|\mathbf{X}\|_2$ & $\|\mathbf{X}\|_*$

$$\|\mathbf{x}\|_p \|\mathbf{x}\|_q \quad \frac{1}{p} + \frac{1}{q} = 1$$

性质:

$$\mathbf{x}^\top \mathbf{z} \leq \underbrace{\|\mathbf{x}\|_* \|\mathbf{z}\|}_{\mathbf{z}}$$

$$\text{proof: } \|\mathbf{x}\|_* = \sup_{\|\mathbf{z}\| \leq 1} \mathbf{x}^\top \mathbf{z} \geq \mathbf{x}^\top \frac{\mathbf{z}}{\|\mathbf{z}\|} \Rightarrow \mathbf{x}^\top \mathbf{z} \leq \|\mathbf{x}\|_* \|\mathbf{z}\|$$

点到点的距离

the Euclidean distance between **u** and **v**

$$d(u, v) = d(u, v) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}$$

点到平面的距离

$$\text{Proj}_V u = \frac{u \cdot v}{\|v\|^2} \cdot v$$

$$Ax + By + C = 0$$

$$w^T x + b = 0$$

点 $x_0 \in \mathbb{R}^n$ 到高维超平面 (hyperplane) $H = \{\mathbf{x} \in \mathbb{R}^n : \underline{w^T x + b = 0}\}$ 的距离为

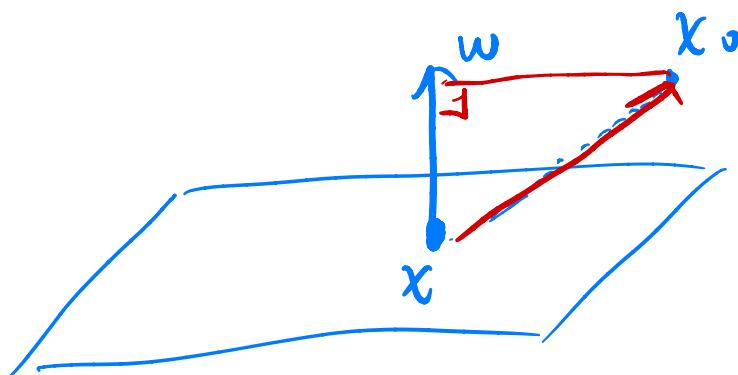
$$d = \frac{|w^T x_0 + b|}{\|w\|}$$

$$\|\text{proj}_w(x_0 - x)\|$$

$$\left\| \frac{(x_0 - x)^T w}{\|w\|^2} \cdot w \right\|$$

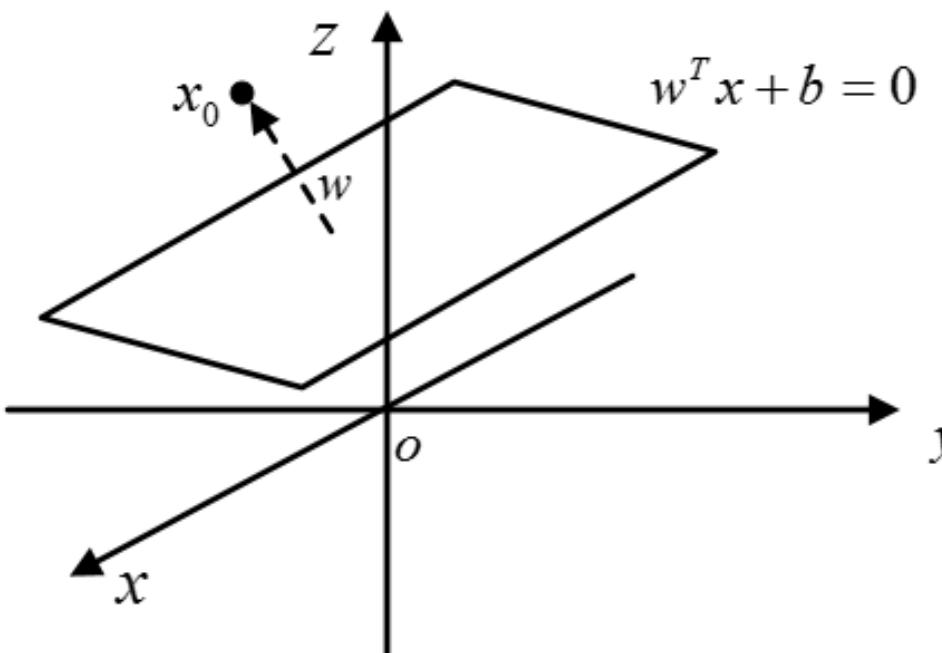
$$\left| \frac{(x_0 - x)^T w}{\|w\|^2} \right| \cdot \|w\|$$

$$\frac{|x_0^T w + b|}{\|w\|}$$



$$\begin{aligned} & (x_0 - x)^T w \\ &= x_0^T w - \underline{x^T w} \end{aligned}$$

$$= x_0^T w + b$$



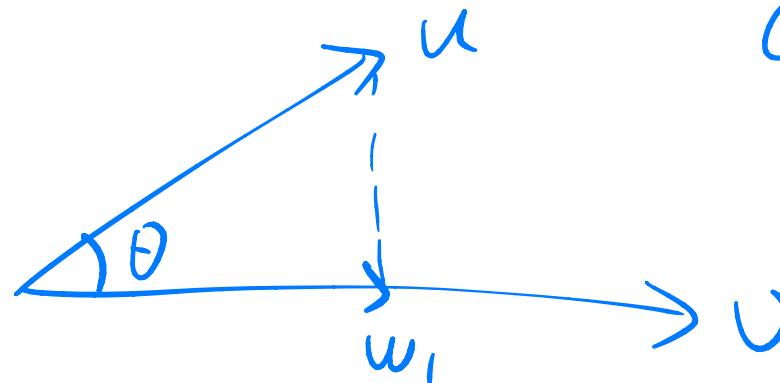
Projection Theorem

- orthogonal projection of \mathbf{u} on \mathbf{v}

$$\mathbf{w}_1 = \text{proj}_{\mathbf{v}}(\mathbf{u}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v}$$

- the vector component of \mathbf{u} orthogonal to \mathbf{v}

$$\mathbf{w}_2 = \mathbf{u} - \mathbf{w}_1 = \mathbf{u} - \text{proj}_{\mathbf{v}}(\mathbf{u}) = \mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v}$$



$$\cos \theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}$$

$$\mathbf{w}_1 = \frac{\mathbf{v}}{\|\mathbf{v}\|} (\mathbf{u} \cdot \mathbf{v}) \cos \theta$$

$$= \frac{\mathbf{v}}{\|\mathbf{v}\|} \|\mathbf{u}\| \frac{\mathbf{u}^T \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}$$

$$= \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|^2} \cdot \mathbf{v}$$

Quadratic Form 二次型

- $Q(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x}$ $A^\top = A$
- $Q(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$

e.g.

$$Q(\mathbf{x}) = x_1^2 + 2x_1 x_2 + 3x_2^2$$

$$Q(\mathbf{x}) = [x_1 \quad x_2] \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Positive (Semi) Definite Matrix (半)正定矩阵

- $A \in \mathbb{S}^n$: symmetric matrix

$$A^\top = A$$

- $A \in \mathbb{S}_+^n$: symmetric positive semi-definite matrix

$$A^\top = A, A \succeq 0 : \forall \mathbf{x} \in \mathbb{R}^n, \underline{\mathbf{x}^\top A \mathbf{x} \geq 0}$$

- $A \in \mathbb{S}_{++}^n$: symmetric positive definite matrix

$$A^\top = A, A \succ 0 : \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \underline{\mathbf{x}^\top A \mathbf{x} > 0}$$

| $\forall A \in \mathbb{S}_+^n$, A can be decomposed as $\underline{A = BB^\top}$

$$\begin{aligned} A &= P \underline{\Delta} P^\top \\ &= P \underline{\Lambda^{\frac{1}{2}}} \cdot \underline{\Lambda^{\frac{1}{2}} P^\top} = (\underline{P \Lambda^{\frac{1}{2}}}) (\underline{P \Lambda^{\frac{1}{2}}})^\top \end{aligned}$$

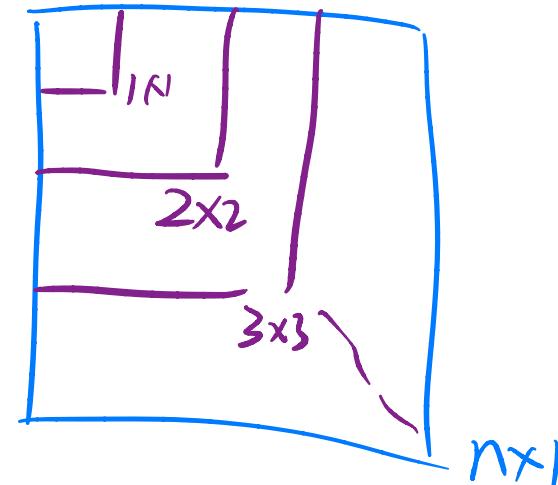
正定矩阵的判定方法

- 所有的特征值均 ≥ 0
- 二次型 $Q(\mathbf{x}) = \mathbf{x}^\top A \mathbf{x} \geq 0, \forall \mathbf{x} \neq 0$
- 顺序主子式均 ≥ 0

e.g. $A^\top A$ 是正定矩阵

$$(A\mathbf{x})^\top (A\mathbf{x})$$

$$|A| > 0 = \prod_{i=1}^n \lambda_i$$

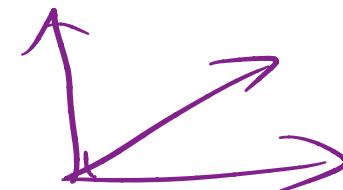


$$\forall \mathbf{x}, \underbrace{\mathbf{x}^\top A^\top A \mathbf{x}}_{\geq 0} = \|A\mathbf{x}\|^2 \geq 0 \Leftrightarrow A^\top A \succeq 0 \Leftrightarrow A^\top A \text{ 的特征值都是非负的}$$

同理, AA^\top 的特征值一定都是非负的

Orthogonal 正交

- \mathbf{u}, \mathbf{v} are orthogonal iff $\mathbf{u} \cdot \mathbf{v} = 0$
- orthogonal set
 $\mathbf{v}_1, \dots, \mathbf{v}_n$ are orthogonal
 $\|\mathbf{v}_1 + \dots + \mathbf{v}_n\| = \|\mathbf{v}_1\| + \dots + \|\mathbf{v}_n\|$
~~proof in homework~~



Orthogonal Matrix 正交矩阵

- $\underline{A^\top A = AA^\top = I_n}$
 $A^{-1} = A^\top$

e.g. rotation matrix $R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$

正交矩阵性质

- A 是正交矩阵 $\Leftrightarrow A$ 的行/列向量组成的集合是正交规范集合

- 正交矩阵的行/列向量是标准正交基底

$$A^\top A = I$$

- A 是正交矩阵 $\Leftrightarrow A^\top$ 是正交矩阵

- A 是正交矩阵 $\Leftrightarrow \|Ax\| = \|\mathbf{x}\|$

$$\|Ax\|_2^2 = \mathbf{x}^\top \underline{A^\top A} \mathbf{x} = \mathbf{x}^\top \mathbf{x}$$

- A 是正交矩阵 $\Leftrightarrow Ax \cdot Ay = \mathbf{x} \cdot \mathbf{y}$

- A 是正交矩阵 $\Leftrightarrow A^{-1}$ 也是正交矩阵

- A, B 是正交矩阵 $\Rightarrow AB$ 也是正交矩阵

- A 是正交矩阵 $\Rightarrow |A| = 1$ or $|A| = -1$

eigenvalue 特征值

$$\begin{aligned} & \underline{Ax = \lambda x} \\ & (\lambda I - A)x = \mathbf{0} \end{aligned}$$

$$\mathbf{x} \neq \mathbf{0}$$

$$\underline{|\lambda I - A| = 0}$$

- $p(\lambda) = |\lambda I - A|$: eigen polynomial 特征多项式
- $p(\lambda) = 0$: characteristic equation 特征方程

eigenvector 特征向量

$$A\mathbf{x} = \lambda\mathbf{x}$$

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

$$\mathbf{x} \neq \mathbf{0}$$

- the nontrivial solutions of $(A - \lambda I)\mathbf{x} = \mathbf{0}$
- $\mathbf{x} \in \text{null}(A - \lambda I)$
- \mathbf{x} : the eigenvectors(特征向量) of A corresponding to λ
- $\text{null}(A - \lambda I)$: the eigenspace(特征空间) of A corresponding to λ

The number of the eigenvectors of A corresponding to λ_i is same as the multiplicity of roots of λ_i of $p(\lambda)$

eigenvalue and eigenvector

$$A = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$$

find the eigenvalues and eigenvectors of A

相似对角化 Diagonalization

若一个矩阵 A 可写作 $\Lambda = P^{-1}AP$, 即 $A = P\Lambda P^{-1}$, 则称 A 可对角化(diagonalizable)

usage:

$$\underbrace{A^n}_{\textcircled{d}} = P\Lambda P^{-1}P\Lambda P^{-1}\dots P\Lambda P^{-1} = \underbrace{P\Lambda^n P^{-1}}_{\textcircled{d}}$$

特征值分解 / 谱分解 Eigenvalues Decomposition

将 A 对角化为 $\underbrace{A = P\Lambda P^{-1}}$:

- 求 A 的特征值和特征向量
- 将特征向量组成 P

原因: $A = P\Lambda P^{-1} \Leftrightarrow \boxed{AP = P\Lambda}$

$$\lambda \leftrightarrow \varphi$$

$$A\varphi = \lambda\varphi$$

Λ : 特征值

P : 特征空间的基拼成(对应特征值)

$$|\lambda I - A| = 0$$

若某个特征值的 几何重数(特征空间的维度) 小于代数重数(特征值的重数), 则 A 不可对角化

$$\lambda \quad \text{nullity}(A - \lambda I) \quad < \quad \lambda_1 = \lambda_2 = \dots = \underline{\quad}$$

$$A = P \Lambda P^{-1} \rightarrow A = P \Lambda P^T$$

Orthogonal Diagonalization 正交对角化

- 实对称矩阵不同特征值对应的特征向量彼此正交

proof:

设 $\lambda_1 \neq \lambda_2$, 其对应的特征向量为 $\mathbf{x}_1, \mathbf{x}_2$

$$A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1, A\mathbf{x}_2 = \lambda_2 \mathbf{x}_2$$

$$\mathbf{x}_1^\top A \mathbf{x}_2 = \mathbf{x}_1^\top \lambda_2 \mathbf{x}_2 = \lambda_2 \mathbf{x}_1^\top \mathbf{x}_2$$

$$(A\mathbf{x}_1)^\top \mathbf{x}_2 = (\lambda_1 \mathbf{x}_1)^\top \mathbf{x}_2$$

$$(\lambda_2 - \lambda_1) \mathbf{x}_1^\top \mathbf{x}_2 = \mathbf{x}_1^\top A \mathbf{x}_2 - \mathbf{x}_1^\top A^\top \mathbf{x}_2 = 0$$

$$\lambda_1 = \lambda_2 \neq \lambda_3$$

Orthogonal Diagonalization 正交对角化

- 若将**实对称矩阵**的每个特征值对应的特征向量的基施密特正交化得到 P , 则 P 是正交矩阵, i.e. $P^\top P = I$

- 实对称矩阵一定可以相似对角化 \Rightarrow 一定可以正交对角化

$$A = P\Lambda P^{-1} \Rightarrow A = P\Lambda P^\top$$

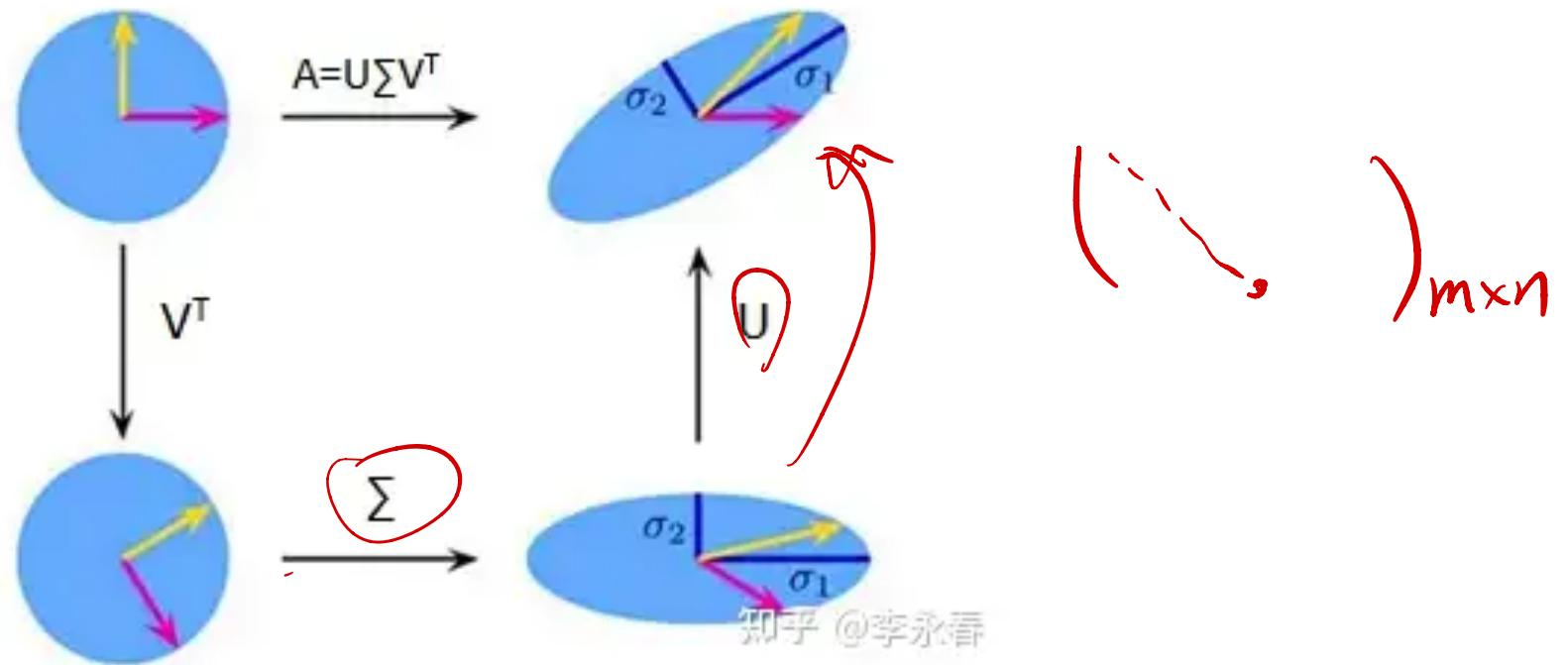
$$P^\top P = I$$



$$|| \cdot || = |$$

$$A = U \Sigma V^T \leftrightarrow A = P \Lambda P^{-1}$$

Singular Value Decomposition(SVD) 奇异值分解



整个 SVD 分解过程可以视为: 首先通过 V^T 把数据旋转到一个新的坐标系, 在这个坐标系中通过 Σ 对数据进行不同程度的拉伸(或压缩), 最后通过 U 可以将这些变化映射回原始或另一个适当的空间.

$$A^T A \rightarrow \lambda_1, \dots, \lambda_n \iff \sigma_i = \sqrt{\lambda_i}$$

$v_1 \dots v_n$

SVD

记录 A 的奇异值分解为 $A = U \Sigma V^\top$, 其中 U 和 V 是正交矩阵, Σ 是对角矩阵.

- $V \& \Sigma$

$$A^\top A = V \Sigma^\top U^\top U \Sigma V^\top = V \Sigma^\top \Sigma V^\top$$

$$A A^\top V = V \Sigma^2$$

对于 V 的每个列向量 v_i , $A^\top A v_i = \sigma_i^2 v_i$

所以 Σ 为 $A^\top A$ 的特征值的平方根, V 为正交规范化的特征向量拼成的矩阵

- U : 同理可得 $A A^\top u_i = \sigma_i^2 u_i$

设奇异值 σ_i 的左奇异向量为 u_i , 右奇异向量为 v_i , 则

$$A v_i = U \Sigma V^\top v_i = U \Sigma e_i = \sigma_i u_i$$

同理: $A^\top u_i = \sigma_i v_i$

由于 U 是正交矩阵, 所以求解方程组 $x \cdot u_i = 0$ 的解即可得到 U

$$u_i = \frac{1}{\sigma_i} A v_i$$

$$u_1, u_2 \in \mathbb{R}^{3 \times 3}$$

$$u_3 = u_1 \times u_2$$

SVD

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$