Lecture 19

§1 Singular value decomposition (奇异值分解)

1. **Definition:** Singular value decomposition (SVD)

一个 real m×n 矩阵 A 的 singular value decomposition (SVD) 的形式为：

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

其中，

① $U \in R^{m\times m}$ 的 columns u_i 被称为 left singular vectors (左奇异向量)

② $V \in R^{n\times n}$ 的 columns v_i 被称为 right singular vectors (右奇异向量)

③ $\Sigma \in R^{m\times n}$ 为一 diagonal matrix with

$$\sigma_{ij} = \begin{cases} 0 & 若\ i \neq j \\ \sigma_i \geq 0 & 若\ i = j \end{cases}$$

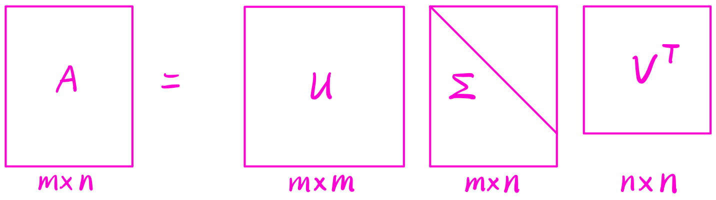
对角线元素 σ_i 为 A 的 singular values (奇异值)，且通常为 ordered：

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0 \quad , \quad p = \min\{m, n\}$$

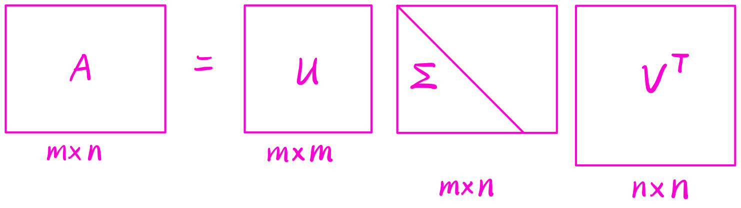
注：① 若 $U = V$，则 SVD 等价于 eigendecomposition (仅仅 $m=n$ 还不能得出该结论)

② SVD 的大致形式为：

Case 1: $m > n$



Case 1: $m < n$



2. **Theorem:** SVD 的 existence 与 construction

所有 real matrix $A \in R^{m\times n}$ 均有 SVD：$A = U\Sigma V^T$ (*)，且

① U 的 columns 为 $AA^T \in R^{m\times m}$ 的 eigenvectors

② V 的 columns 为 $A^TA \in R^{n\times n}$ 的 eigenvectors

③ singular values $\sigma_i = \sqrt{\lambda_i}$，$i = 1, \cdots, p$，其中 $\lambda_1, \cdots, \lambda_p$ 为 AA^T (若 $m<n$) 或 A^TA (若 $m>n$) (AA^T 与 A^TA 中维度较小的) 的 eigenvalues

证明： (此处不用讨论 m 与 n 谁大，因为 A^TA 与 AA^T 相差的特征值全为 0)

该证明非常 constructive，即我们先根据 AA^T 的 eigendecomposition 求出 U 和 Σ，再根据目标构造出 V.

- 由于 AA^T 为 symmetric 且 positive semi-definite ($x^TA^TAx = \|Ax\|^2 \geq 0$, $\forall x$), 有

  $AA^T = U\Lambda U^T$, U 为 orthorgonal matrix (实对称矩阵一定酉可对角化)

且 $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_m)$ with $\lambda_1 \geq \cdots \geq \underset{\sim}{\lambda_r} > 0 = \lambda_{r+1} = \cdots = \lambda_m$ (半正定矩阵特征值非负)

定义 $\Sigma := \sqrt{\Lambda} = \text{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_m})$

① 第一种情况: $r = m$ (注: 此时一定有 $m \leq n$, 否则一定有 $\lambda_i = 0$)

由于此时 U 与 Σ 可逆, 为了使 $A = U\Sigma V^T$, 直接构造 V 为:

$$V := A^TU\Sigma^{-1}$$

接下来仅需 check:

① V 为 orthorgonal: $V^TV = \Sigma^{-1}U^TAA^TU\Sigma^{-1} = \Sigma^{-1}\Lambda\Sigma^{-1} = I$ ($\Sigma = \sqrt{\Lambda} \Rightarrow \Lambda = \Sigma^2$)

② $A = U\Sigma V^T$: $U\Sigma V^T = U\Sigma\Sigma^{-1}U^TA = UU^TA = A$ (U 为 orthorgonal)

② 第二种情况: $r < m$ (注: 此时不一定有 $m \leq n$, 但有 $r \leq \min\{m,n\}$, 仅需保证 $\Sigma \in R^{m\times n}$, $V \in R^{m\times m}$)

此时 U 与 Σ 不可逆, 无法直接构造 V.

- 将 U partition 为 $U = [U_1, U_2]$, 其中 $U_1 \in R^{m\times \underset{\sim}{r}\,<m}$, $U_2 \in R^{m\times(m-r)}$.

  定义 $\tilde{\Sigma} = \text{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_r}) \in R^{r\times r}$

- 构造 $V_1 := A^TU_1\tilde{\Sigma}^{-1} \in R^{n\times \underset{\sim}{r}\,<n}$, 接下来我们可以证得:

  ① $V_1^TV_1 = I$

$$\begin{aligned} V_1^TV_1 &= \tilde{\Sigma}^{-1}U_1^TAA^TU_1\tilde{\Sigma}^{-1} \\ &= \tilde{\Sigma}^{-1}U_1^TU_1\tilde{\Sigma}^2U_1^TU_1\tilde{\Sigma}^{-1} \end{aligned}$$

$$\left(AA^T = U\Lambda U^T = [U_1\ U_2]\begin{bmatrix} \tilde{\Sigma}^2 & 0 \\ 0 & 0 \end{bmatrix}[U_1\ U_2]^T = U_1\tilde{\Sigma}^2U_1^T \right)$$

$$\begin{aligned} &= \tilde{\Sigma}^{-1}\tilde{\Sigma}^2\tilde{\Sigma}^{-1} \\ &= I \end{aligned}$$

  ② $U_1\tilde{\Sigma}V_1^T = A$

$$\begin{aligned} U_1\tilde{\Sigma}V_1^T &= U_1\tilde{\Sigma}\tilde{\Sigma}^{-1}U_1^TA \\ &= U_1U_1^TA \\ &= A \end{aligned}$$

$$\left(U^TU = \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix}[U_1\ U_2] = \begin{bmatrix} U_1^TU_1 & U_1^TU_2 \\ U_2^TU_1 & U_2^TU_2 \end{bmatrix} = I \quad \Rightarrow \quad U_1^TU_2 = 0 \right.$$

$$\Rightarrow U_2^TAA^TU_2 = U_2^TU_1\tilde{\Sigma}^2U_1^TU_2 = 0$$

$$\Rightarrow \text{trace}(U_2^TAA^TU_2) = \|A^TU_2\|_F^2 = 0$$

$$\Rightarrow U_2^TA = 0$$

$$UU^T = U_1U_1^T + U_2U_2^T = I \quad \Rightarrow \quad U_1U_1^T = I - U_2U_2^T$$

$$\left. \Rightarrow U_1U_1^TA = (I - U_2U_2^T)A = A - U_2\underbrace{U_2^TA}_{=0} = A \right)$$

· 构造 $V := [V_1, V_2] \in R^{n\times n}$ ($V_2 \in R^{n\times(n-r)}$ 任选，确保 V 为 orthorgonal)

$$\Sigma := \begin{bmatrix} \tilde{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \in R^{m\times n}$$

因此有：

$$V^T V = I$$

$$U \Sigma V^T = U_1 \tilde{\Sigma} V_1^T = A$$

3. Definition: thin SVD (Economy size)

令 r 为 matrix A 的非零 singular values 数，即

$$\sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_p \, .$$

则我们有：

$$A = [\, \overset{r}{\underbrace{U_1}} \; \overset{m-r}{\underbrace{U_2}} \,] \begin{array}{c} r\{ \\ m-r\{ \end{array} \begin{bmatrix} \overset{r}{\tilde{\Sigma}} & \overset{n-r}{0} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \begin{array}{c} \}r \\ \}n-r \end{array}$$

其中：$U_1 \in R^{m\times r}$, $U_2 \in R^{m\times(m-r)}$, $V_1 \in R^{n\times r}$, $V_2 \in R^{n\times(n-r)}$, $\tilde{\Sigma} = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$

则 thin (economy size) SVD 为：

$$A = U_1 \tilde{\Sigma} V_1^T$$

4. 对 Left & Right singular vectors 的理解

$$A = U \Sigma V^T \iff \begin{cases} AV = U\Sigma \\ U^T A = \Sigma V^T \end{cases}$$

$$\iff \begin{cases} A\mathbf{v}_i = \sigma_i \mathbf{u}_i \\ \mathbf{u}_i^T A = \sigma_i \mathbf{v}_i \end{cases}$$

($\mathbf{v}_i$ 为右奇异向量，作用类似右乘特征向量)

($\mathbf{u}_i$ 为左奇异向量，作用类似左乘特征向量)

$$\forall i = 1, \cdots, p$$

例 1: 求下述矩阵的 SVD：

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

① Step 1: 计算 A^TA 与 AA^T 中维度更小 (更易于计算) 的一个

$$AA^T = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

② Step 2: 计算 A^TA 或 AA^T 的特征值，求出 Σ

$$\det(AA^T - \lambda I) = (2-\lambda)^2 - 1 = 4 - 4\lambda + \lambda^2 - 1 = (\lambda-1)(\lambda-3) = 0$$

$$\Rightarrow \lambda_1 = 3, \ \lambda_2 = 1$$

$$\Rightarrow \sigma_1 = \sqrt{3}, \ \sigma_2 = 1$$

$$\Rightarrow \Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}$$

③ Step 3: 计算 A^TA 或 AA^T 的特征向量，求出 V 或 U

- $AA^T - \lambda_1 I = 0 \iff \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 1 & -1 & 0 \end{array}\right] \iff \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$

  取自由未知量 $x_2 = 1$，有 eigenvector $[1, 1]^T$

  Normalize 后有：$u_1 = \frac{1}{\sqrt{2}}[1, 1]^T$

- $AA^T - \lambda_2 I = 0 \iff \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 1 & 0 \end{array}\right] \iff \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$

  取自由未知量 $x_2 = 1$，有 eigenvector $[-1, 1]^T$

  Normalize 后有：$u_2 = \frac{1}{\sqrt{2}}[-1, 1]^T$

$$\Rightarrow U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

④ Step 4：利用 $A = U\Sigma V^T$ 求解另一个 orthorgonal matrix

$$A = U\Sigma V^T \iff A^T = V\Sigma U^T \iff V = A^T U \Sigma^{-1}$$

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

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

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