LECTURE 33 : APPLICATION & INTERPRETATION OF SVD

Certainly this lecture is not comprehensive, if you search "applications of singular value decomposition" you'll see many things not covered here. Let's review,

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

$$\underset{m\times n}{A} = \underset{(m\times m)}{U}\,\underset{(m\times n)}{\Sigma}\,\underset{(n\times n)}{V^T}$$

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & & \end{array}\right]$$

$\underbrace{\sigma_1, \sigma_2, \dots, \sigma_r}_{\text{singular values of } A} > 0$

$$U = \underbrace{[u_1 | u_2 | \cdots | u_m]}_{\text{left-singular vectors}}$$

AA^T's e-vectors orthonormalized

$$V = \underbrace{[v_1 | v_2 | \cdots | v_n]}_{\text{right-singular vectors}}$$

A^TA's e-vectors orthonormalized

$$u_j = \frac{1}{\sigma_j} A v_j \quad (j = 1, \dots, r) \qquad v_j = \frac{1}{\sigma_j} A^T u_j \quad (j = 1, \dots, r)$$

<u>Remark</u>: the formulas above show us how to generate <u>part</u> of U from V and vice-versa, however in both cases we may have to select $u_{r+1}, \dots, u_m$ or $v_{r+1}, \dots, v_n$ to orthonormally complete the basis for $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. We saw this in last lecture's examples.

And now for something new,

$$\begin{aligned}
A &= U\,[\sigma_1 e_1 | \sigma_2 e_2 | \cdots | \sigma_r e_r | 0 | \cdots | 0]\,V^T \\
&= [\sigma_1 U e_1 | \sigma_2 U e_2 | \cdots | \sigma_r U e_r | 0 | \cdots | 0]\,V^T \\
&= [\sigma_1 u_1 | \sigma_2 u_2 | \cdots | \sigma_r u_r | 0 | \cdots | 0]\,[v_1 | v_2 | \cdots | v_n]^T \\
&= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T
\end{aligned}$$

$\boxed{E1}$ $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$ we found $\sigma_1 = \sqrt{15}$

and $u_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $v_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

then $\sigma_1 u_1 v_1^T = \sqrt{15}\left(\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right)\left(\frac{1}{\sqrt{3}}[1,1,1]\right) = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix} \overset{\checkmark}{=} A.$

$\boxed{E2}$ $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ we found $\sigma_1 = \sqrt{6}$, $\sigma_2 = \sqrt{2}$

and $u_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$ and $u_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $v_1 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $v_2 = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$

then observe,

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = \frac{\sqrt{6}}{2\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}[1,1,1,1] + \frac{\sqrt{2}}{2\sqrt{2}}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}[-1,-1,1,1]$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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

$\boxed{E3}$ $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix}$ had $\sigma_1 = 2$, $\sigma_2 = \sqrt{2}$ with

left-singular vectors $u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ & $u_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

and right-singular vectors $v_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ then,

$$\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T = 2\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}[0,1] + \sqrt{2}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}[1,0]$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix} \overset{\checkmark}{=} A$$

THE MOORE - PENROSE PSUEDO INVERSE

We know A^{-1} only exists when A is $n \times n$ and $\det(A) \neq 0$. When A is rectangular the generalization of A^{-1} is known as the psuedoinverse. It seems <u>two</u> popular notations are in use; $A^{\dagger}$ or A^{+}

read as "A dagger"   read as "A plus"

Def Given A with r-nonzero singular values and SVD of $A = U\Sigma V^T$ we define for $A \in \mathbb{R}^{m \times n}$

$$\Sigma^{+} = \begin{bmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_r \\ & & \end{bmatrix} \in \mathbb{R}^{n \times m} \quad \text{and} \quad A^{+} = V\Sigma^{+}U^T$$

Let's study how AA^{+} and $A^{+}A$ simplify,

$$\begin{aligned} AA^{+} &= (U\Sigma V^T)(V\Sigma^{+}U^T) \\ &= U\Sigma\Sigma^{+}U^T \\ &= U\left[\sigma_1 e_1 | \cdots | \sigma_r e_r | 0 | \cdots | 0\right]\begin{bmatrix} (1/\sigma_1)f_1 \\ \vdots \\ (1/\sigma_r)f_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}U^T \\ &= U\left[\begin{array}{c|c} I_r & \\ \hline & \end{array}\right]U^T \\ &= \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] \end{aligned}$$

similarly, $A^{+}A = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]$

Sometimes the psuedoinverse has slick formula,

$$A^+A = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] \quad \text{and} \quad AA^+ = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]$$

where $\text{rank}(AA^T) = \text{rank}(A^TA) = r = \#$ of positive singular values. When A is full rank,

$\text{rank}(A) = \min\{m,n\}$
$A \in \mathbb{R}^{m\times n}$
defines "full rank"

$$\boxed{\begin{array}{ll} (1.)\ m < n, & A^+ = A^T(AA^T)^{-1} \\ (2.)\ m > n, & A^+ = (A^TA)^{-1}A^T \end{array}}$$

$\boxed{\text{E4}}$ $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix}$ has $\text{rank}(A) = 2 = \min\{2,3\}$

$m = 3$, $n = 2$ in case (2.)

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

$$\begin{aligned} A^+ &= (A^TA)^{-1}A^T \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}^{-1} A^T \\ &= \begin{bmatrix} 1/2 & 0 \\ 0 & 1/4 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} \\ &= \underline{\begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix}} \end{aligned}$$

Let's see if we calculate from the def$^{=}$ we find same,

$$\begin{aligned} A^+ = V\Sigma^+U^T &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/\sqrt{2} & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1/\sqrt{2} & 0 \\ 1/2 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix}. \end{aligned}$$

THE FOUR SUBSPACES

row (A)

null (A)

row (A) $\perp$ null (A)

row (A) & null (A) span all $\mathbb{R}^n$

col (A)

null (A^T)

col (A) $\perp$ null (A^T)

col (A) & null (A^T) span all $\mathbb{R}^m$

natural basis for all of the above can be found with $A = U \Sigma V^T$

$$U = [u_1 \mid u_2 \mid \cdots \mid u_r \mid u_{r+1} \mid \cdots \mid u_m] \in O(m, \mathbb{R})$$

$$V = [v_1 \mid v_2 \mid \cdots \mid v_r \mid v_{r+1} \mid \cdots \mid v_n] \in O(n, \mathbb{R})$$

$A^T A v_j = 0$ for $j = r+1, .., n$, $v_j \in \text{Null}(A)$

$A A^T u_j = 0$ for $j = r+1, .., m$, $u_j \in \text{Null}(A^T)$

$\text{Col}(A) = \text{span}\{u_1, .., u_r\}$

$\text{Row}(A) = \text{span}\{v_1^T, .., v_r^T\}$

$\text{Col}(A^T) = \text{span}\{v_1, .., v_r\}$

$\dim(\text{Col}(A)) = \dim(\text{Row}(A)) = r = \text{rank}(A).$

<u>Remark</u>: $A^TAx = A^Tb$ has unique solution of $x = (A^TA)^{-1}A^Tb$ provided $\det(A^TA) \neq 0$

Notice $A^+ = (A^TA)^{-1}A^T$ in this case

hence $A^TAx = A^Tb$ has solution $\underline{x = A^+b}$

Therefore, when we know A^+ we can solve least squares (in case $\det(A^TA) \neq 0$). I mentioned in-class when $Ax = b$ has non-unique sol$^{\underline{n}}$ then A^+b gives sol$^{\underline{n}}$ of minimal norm.

$\boxed{E5}$ COURTESY OF CHATGPT

Let's walk through a concrete least-squares problem and see exactly how the **Moore–Penrose inverse** solves it.

Problem: Fit a line to data

Suppose you have three data points:

$$(0,1), \quad (1,2), \quad (2,2)$$

You want the "best fit" line:

$$y = ax + b$$

Step 1: Write as a linear system

We rewrite this as:

$$Ax \approx b$$

where

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}, \quad x = \begin{pmatrix} a \\ b \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

This system is **overdetermined** (3 equations, 2 unknowns), so no exact solution exists.

Step 2: Use the Moore–Penrose inverse

7

The least-squares solution is:

$$x = A^\dagger b$$

Since A has full column rank, we use:

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

Step 3: Compute $A^T A$

$$A^T A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 3 \\ 3 & 3 \end{pmatrix}$$

Step 4: Invert $A^T A$

$$(A^T A)^{-1} = \frac{1}{(5)(3) - (3)(3)} \begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix}$$

Step 5: Compute $A^\dagger$

$$A^\dagger = (A^T A)^{-1} A^T = \frac{1}{6} \begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix} \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix}$$

Multiply:

$$A^\dagger = \frac{1}{6} \begin{pmatrix} -3 & 0 & 3 \\ 5 & 2 & -1 \end{pmatrix}$$

Step 6: Solve $x = A^\dagger b$

$$x = \frac{1}{6} \begin{pmatrix} -3 & 0 & 3 \\ 5 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

Compute:

- First component:

$$\frac{1}{6}(-3 \cdot 1 + 0 \cdot 2 + 3 \cdot 2) = \frac{1}{6}(3) = \frac{1}{2}$$

- Second component:

$$\frac{1}{6}(5 \cdot 1 + 2 \cdot 2 - 1 \cdot 2) = \frac{1}{6}(7) = \frac{7}{6}$$

Final answer

$$a = \frac{1}{2}, \quad b = \frac{7}{6}$$

So the best-fit line is:

$$\boxed{y = \frac{1}{2}x + \frac{7}{6}}$$