16

SVD Applications

★ General pseudo inverse

⇒ If $A \neq 0$ has SVD $A = U\Sigma V^T$

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

is the pseudo-Inverse or Moore-Penrose Inverse of A.

⇒ If A is skinny & full rank

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

- gives the least-squares solution $x_{ls} = A^{+}y$

⇒ If A is fat & full rank

$$A^{+} = A^{T}(AA^{T})^{-1}$$

gives the least norm solution $x_{ln} = A^{+}y$

⇒ In general case

$$X_{ls} = \{ z \mid \|Az - y\| = \min_{w} \|Aw - y\| \}$$

is set of least-square solutions

⇒ $x_{pinv} = A^{+}y \in X_{ls}$ has minimum norm on X_{ls}.

{ x_{pinv} is the minimum norm least-squares solution }

★ Pseudo-inverse via regularization

⇒ for $\mu > 0$, let x_μ be (unique) minimizer of

$$\|Ax - y\|^2 + \mu\|x\|^2$$

$$x_\mu = (A^TA + \mu I)^{-1} A^T y$$

⇒ then we have $\lim_{\mu \to 0} x_\mu = A^+ y$

$$A^+ = \lim_{\mu \to \infty} (A^TA + \mu I)^{-1} A^T$$

★ Full SVD

⇒ Let SVD of $A \in \mathbb{R}^{m \times n}$ with $Rank(A) = r$

$$A = U_1 \Sigma_1 V_1^T \quad \text{—— ①}$$ {Called Compact SVD}

⇒ find $U_2 \in \mathbb{R}^{m \times (m-r)}$, $V_2 \in \mathbb{R}^{n \times (n-r)}$ st $U = [U_1 \; U_2] \in \mathbb{R}^{m \times m}$ & $V = [V_1 \; V_2] \in \mathbb{R}^{n \times n}$ are orthogonal.

⇒ add zero row/cols to Σ_1 to form $\Sigma \in \mathbb{R}^{m \times n}$

⇒ then we have

$$A = U \Sigma V^T \quad \text{—— ②}$$

{Called full SVD}

⋆ Sensitivity of linear equations to data error

⇒ Consider $y = Ax$, $A \in \mathbb{R}^{n \times n}$ invertible ; of course $x = A^{-1}y$

⇒ Suppose we have an error or noise in y.

then, $x \to x + \delta x$

$y \to y + \delta y$

$$\boxed{\delta x = A^{-1}\delta y}$$

⇒ hence we have $\|\delta x\| = \|A^{-1}\delta y\| \leq \|A^{-1}\|\|\delta y\|$

⇒ if $\|A^{-1}\|$ is large

- → Small error in y can lead to large errors in x
- → Can't solve for x given y (with small error)
- → hence, A can be considered singular in practice

⇒ A more refined analysis uses relative instead of absolute errors in x and y

⇒ Since $y = Ax$

$$\|y\| \leq \|A\|\,\|x\|$$

$$\frac{\|\delta x\|}{\|x\|} \leq \underbrace{\|A\|\,\|A^{-1}\|}_{K(A)} \frac{\|\delta y\|}{\|y\|}$$

$$K(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$ {Called conditional number of A}

$$\boxed{\begin{pmatrix}\text{relative error in}\\ \text{solution } x\end{pmatrix} \leq \begin{pmatrix}\text{Condition}\\ \text{number}\end{pmatrix} \cdot \begin{pmatrix}\text{relative error}\\ \text{in data } y\end{pmatrix}}$$

- A is well conditioned if κ is small.
- A is poorly conditioned if κ is large.

{definition of small & large depends on application}

⇒ Same analysis holds for least-square solutions with A nonsquare.

★ <u>Low rank approximations</u>

⇒ Suppose $A \in \mathbb{R}^{m \times n}$, $\text{Rank}(A) = r$,

with SVD $A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

⇒ We seek matrix $\hat{A}$, $\text{Rank}(\hat{A}) \leq p < r$

s.t $\hat{A} \approx A$ in the sense $\|A - \hat{A}\|$ is minimized.

<u>Solution</u>: Optimal rank p approximation is

$$\hat{A} = \sum_{i=1}^{p} \sigma_i u_i v_i^T$$

- hence $\|A - \hat{A}\| = \left\| \sum_{i=p+1}^{r} \sigma_i u_i v_i^T \right\| = \sigma_{p+1}$

<u>Interpretation</u>: SVD dyads $u_i v_i^T$ are ranked in order of importance.

{take p to get p rank approximation}

★ Distance to Singularity

⇒ Another interpretation of σ_i:

$$\sigma_i = \min \{ \|A-B\| \mid \mathrm{Rank}(B) \leq i-1 \}$$

{The distance to the nearest rank $i-1$ matrix}