19.08.2022

CLASS - V

Cor. $A \in M_{m,n}(\mathbb{R})$, $z \in \mathbb{R}^m$.

$$\sigma_{max}([A|z]) \geq \sigma_{max}(A) \geq \sigma_2([A|z]) \geq \sigma_2(A) \geq \dots$$

(Ch. 3 : Matrix Analysis by R. Bhatia)

Standard Technique for obtaining singular values inequalities for A.

(1) Apply eigenvalue inequalities to A^*A or AA^*.
(eigenvalues are squares of singular values of A).

(2) Apply eigenvalue inequalities to $\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} = H_A$
(eigenvalues of $\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$ are $\pm$ of the singular value of A, and 0, with multiplicities at least $|m-n|$

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} = \begin{bmatrix} AA^* & 0 \\ 0 & A^*A \end{bmatrix}$$

Exc. $A_{m\times n}$ $B_{n\times m}$ $Sp(AB)\setminus\{0\} = Sp(BA)\setminus\{0\}$

$n \geq m$ $\chi_{AB}(t) = t^{n-m}\chi_{BA}(t)$

$$Sp(H_A^2) \subseteq \{\text{squares of singular values of } A\}$$

$$Sp(H_A^2) = \{\lambda^2 : \lambda \in Sp(H_A)\}$$

$$\Rightarrow Sp(H_A) \subseteq \{\pm \text{ singular values of } A\}$$

To Show. $Sp(H_A)\setminus\{0\} = \{\pm \text{ singular values of } A\}\setminus\{0\}$

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} = \lambda\begin{bmatrix} u \\ v \end{bmatrix} \quad \lambda \neq 0. \qquad u \in \mathbb{R}^m, v \in \mathbb{R}^n$$

$Av = \lambda u$ $\qquad A^*Av = \lambda^2 v$

$A^*u = \lambda v$

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sigma_i}Av_i \\ v_i \end{bmatrix} = \begin{bmatrix} Av_i \\ \sigma_i v_i \end{bmatrix} = \sigma_i\begin{bmatrix} \frac{1}{\sigma_i}Av_i \\ v_i \end{bmatrix}$$

$$\begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}\begin{bmatrix} -\frac{1}{\sigma_i}Av_i \\ v_i \end{bmatrix} = \begin{bmatrix} Av_i \\ -\sigma_i v_i \end{bmatrix} = -\sigma_i\begin{bmatrix} -\frac{1}{\sigma_i}Av_i \\ v_i \end{bmatrix}$$

APPLICATIONS OF SVD

(1) Solving Linear Equations

$A \in M_{m,n}(\mathbb{R})$ of rank $r \le \min\{m,n\}$ & $\vec{b} \in \mathbb{R}^m$.

Want to solve $A\vec{x} = \vec{b}$

Modified Problem :- $\arg \min_{\vec{x} \in \mathbb{R}^n} \|A\vec{x} - \vec{b}\|_2$ (arg = argument)

$J(\vec{x}) = \|A\vec{x} - \vec{b}\|_2^2 = (\vec{b} - A\vec{x})^T (\vec{b} - A\vec{x})$

$\frac{\partial J}{\partial \vec{x}} := \vec{\nabla} J = -2A^T(\vec{b} - A\vec{x}) = 0.$

$\Rightarrow \boxed{A^T A \vec{x} = A^T \vec{b}}$ (Normal equation)

If $A^T A$ is invertible ($n \le m$), then $\vec{x} = (A^T A)^{-1} A^T \vec{b}$

Fundamental Theorem of Linear Algebra :-

$A \in M_{m,n}(\mathbb{R})$ of rank $r \le \min\{m,n\}$ $\quad A : \mathbb{R}^n \to \mathbb{R}^m$.

$R(A) \subseteq \mathbb{R}^m$, dim $= r$ $\quad N(A) \subseteq \mathbb{R}^n$, dim $= n-r$.

$R(A^T) \subseteq \mathbb{R}^n$, dim $= r$ $\quad N(A^T) \subseteq \mathbb{R}^m$, dim $= m-r$.

$R(A^T) = [N(A)]^\perp$ $\quad N(A^T) = [R(A)]^\perp$

Proof:- $\vec{x} \in R(A^T)$, $\vec{x} = A\vec{y}$. If $z \in N(A) \Rightarrow$

$\langle A^T y, z \rangle = \langle y, Az \rangle = \langle y, 0 \rangle = 0 \quad \Rightarrow \langle x, z \rangle = 0.$

$\Rightarrow x \in [N(A)]^\perp$.

Theorem. Singular Value Decomposition of A provides orthonormal bases for these four subspaces and A, A^T are put in diagonal form w.r.t. these base.

$R(A) = \text{span}\{u_1, \dots, u_r\}$ $\quad P_{R(A)} = [u_1 | \cdots | u_r] \begin{bmatrix} u_1^T \\ \vdots \\ u_r^T \end{bmatrix}$ $m \times m$ matrix.

$N(A^T) = [u_{r+1} | \dots | u_m]$ $\quad P_{N(A^T)} = [u_{r+1} | \cdots | u_m] \begin{bmatrix} u_{r+1}^T \\ \vdots \\ u_m^T \end{bmatrix}$

$N(A) = [v_{r+1} | \dots | v_n]$ $\quad P_{N(A)} = [v_{r+1} | \dots | v_n] \begin{bmatrix} v_{r+1}^T \\ \vdots \\ v_n^T \end{bmatrix}$

$R(A^T) = [v_1 | \dots | v_r]$ $\quad P_{R(A^T)} = [v_1 | \dots | v_r] \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$

(1) If $m > n = r$, $\vec{x} = A^+\vec{b}$ for $A^+ = (A^TA)^{-1}A^T$.

(2) If $n > m = r$, $\vec{x} = A^+\vec{b} + \vec{x}_1$ for $A^+ = A^T(AA^T)^{-1}$

(3) If $m = n = r$, $\vec{x} = A^+\vec{b}$ for $A^+ = A^{-1}$

(i) $N(A) = \{0\}$ $R(A^T) = \mathbb{R}^n$ (over determined)

(least square solution is unique)

(ii) (under determined)

$\vec{x} = \vec{x}_0 + \vec{x}_1$, $\vec{x}_0 \in R(A^T) = N(A)^\perp$

$\vec{x}_1 \in N(A)$

$\vec{x}_0 = A^T\vec{c}$ for some $\vec{c} \in \mathbb{R}^m$.

$\vec{b} = A\vec{x} = AA^T\vec{c}$

$\Rightarrow \vec{c} = (AA^T)^{-1}\vec{b}$

$\Rightarrow \vec{x}_0 = A^+(AA^T)^{-1}\vec{b}$

$\vec{x}_0 \perp N(A)$

Moore - Penrose Pseudoinverse

$A = U\Sigma V^T$ (of full rank $n\times n$ matrix)

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

For a scalar x, we define its pseudoinverse: $x^+ = \begin{cases} \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$

Given $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)$ $\Sigma^+ = \text{diag}(\sigma_1^+, \ldots, \sigma_p^+)^T$

$\Sigma = \left[\begin{array}{c|c} \Sigma_r & 0 \\ \hline 0 & 0 \end{array}\right]$ so $\Sigma^+ = \left[\begin{array}{c|c} \Sigma_r^{-1} & 0 \\ \hline 0 & 0 \end{array}\right]$

$A^+ = V\Sigma^+U^T$ (if $A = U\Sigma V^T$)

The matrix A^+ satisfies (and is uniquely determined by) the four Penrose conditions

(i) $AA^+A = A$ (iii) $A^+AA^+ = A^+$

(ii) $(AA^+)^T = AA^+$ (iv) $(A^+A)^T = AA^+$

$(A^+A)(A^+A) = (A^+A)$

$A^+A \rightarrow$ projection with range $R(A^T)$ $A^+A = P_{R(A^T)}$

$AA^+ \rightarrow$ projection matrix with range $R(A)$.

The most general least squares solution of $A\vec{x} = \vec{b}$ is given by $\vec{x} = \vec{x}_0 + \vec{x}_1$, $\vec{x}_1 = N(A)$. $\boxed{\vec{x}_0 = A^+\vec{b}}$

Reduced Rank Approximation

$A = \sum \sigma_i u_i v_i^T$ $\quad A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ is the best rank k approximation to A.

Application:- Compression of Image.

"Netflix" Recommendation System :-

$$\begin{array}{cc} \text{customers} & \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \\ & \text{Movies} \end{array}$$

[Find A of min rank $A_{ij} = v_{ij}$ for i, j filled out according to preference]

Regularization of Ill Conditioned Problems

$A = U\Sigma V^T$ is replaced by 'filtered' or 'regularized' version.

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

$$A_f^+ = f(A)A^+ = \sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i} v_i u_i^T$$

$$f(\sigma) = \frac{\sigma^2}{\sigma^2 + \lambda} \quad \text{(Tychonoff)}$$

$\frac{f(\sigma)}{\sigma}$ is nearly 0 for small σ (high pass filter)

Modified Least Square Solution

$$\arg\min_{x \in \mathbb{R}^n} \|A\vec{x} - \vec{b}\|_2^2 + \lambda \|\vec{x}\|_2^2$$

$$\vec{\nabla} J = 2A^T(A\vec{x} - \vec{b}) + 2\lambda\vec{x} = 0$$

$$\Rightarrow (A^TA + \lambda I)\vec{x} = A^T b$$

$$\Rightarrow \vec{x} = (A^TA + \lambda I)^{-1} A^T b$$

(λ Ridge Regression Parameter)

Bias Variance Trade off

PCA (Principal Component Analysis)

$X_1, \ldots, X_n \to$ Normal Random Variables

$$\Sigma = \begin{bmatrix} Var(X_1) & & Cov(X_i, X_j) \\ & \ddots & \\ & & Var(X_n) \end{bmatrix}$$

Σ = positive definite symmetric Matrix.

$U \Sigma U^T = D \to$ diagonal matrix with strictly positive entries.

$X_1, \ldots, X_n$

$y_1 = u_{11} X_1 + \ldots + u_{1n} X_n$

$\vdots$

$y_n = u_{n1} X_1 + \ldots + u_{nn} X_n$

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

$Cov(y_i, y_j) = 0 \quad i \neq j$

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$

Variable set : $y_1, \ldots, y_k$ (if $\lambda_1, \ldots, \lambda_k$ belong to the image group).