MAT 653: Statistical Simulation

Instructor: Dr. Wei Li   Scribe: Jiangyu Yu

Sep 14th, 2021

Singular value decomposition

A, a $m \times n$ matrix, $A^T A$ is a symmetric matrix. Let $\{v_1, \ldots, v_n\}$ consists of all orthonormal eigenvectors of $A^T A$.
Let $\lambda_1, ..., \lambda_n$ the associated eigenvalues, $\lambda_1 \geq \lambda_2 \geq \lambda_n \geq 0$.
We have $\|Av_i\|^2 = \lambda_i$. Since

$$A^T A v_i = \lambda_i v_i \Rightarrow v_i^T A^T A v_i = \lambda_i v_i^T v_i \lambda_i$$

Denote $\sigma_i = \sqrt{\lambda_i}$, λ_i is the ith eigenvalue of $A^T A$.

Fact: rank of A is equal to the number of positve singular value of A.

SVD: Let A be a rank r matrix. There exists a matrix $\Sigma_{m\times n}$ with diagonal entries in D are the first r singular value. That is

$$\Sigma_{m\times n} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

$$D = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$$

The decomposition of A is:

$$A = U_{m\times n} \Sigma_{m\times n} V_{n\times n}^T$$

Where U and V are bothy orthogonal matrices. V consists of all the eigenvectors of $A^T A$. U consists the column u_i as

$$u_i = \frac{Av_i}{\|Av_i\|} = \frac{Av_i}{\sigma_i}, 1 \leq i \leq r$$

Those $\{u_i\}_1^r$ can extend to $\{u_i\}_1^m$ as the orthonormal basis.
Matrix U and V have the property:

$$Col(U) = Col(A), Col(V) = Row(A)$$

Reduced SVD

For those $A_{m\times n}$, $U_{m\times m}$, $V_{n\times n}$, $\Sigma_{m\times n}$ above, we have the partition:

$$U = [U_r, U_{m-r}], V = [V_r, V_{n-r}], \Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

Consequently, the SVD of A can be represented as:

$$A = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T$$

Application

1. Linear Least Squares

We can use SVD to solve the Least Squares Problems as following:

$$\begin{aligned}
Ax &= b, \\
A^T Ax &= A^T b, \\
(UDV^T)^T (UDV^T)x &= (UDV^T)^T b, \\
\Rightarrow VD^2 Dx &= VDU^T b, \\
\Rightarrow D^2 V^T x &= DU^T b, \\
\Rightarrow DV^T x &= U^T b.
\end{aligned}$$

Denote $w = V^T x$,$y = U^T b$, we have the following algorithm:

Algorithmn

1. Find the SVD of $A = UDV^T$;

2. Compute $y = U^T b$;

3. Solve $w^* = D^{-1} y$;

4. $V^T x = w^* \Rightarrow x = V w^*$;

The solution of the equation is $x = V_r D^{-1} U_r^T b$, and we can notice that this SVD method allows A,b to be arbitrary. Notice that $V_r D^{-1} U_r^T$ is the inverse of A. Here we have the concept of generalized inverse.

Moore-Penrose inverse: $A^+ = V_r D^{-1} U_r^T$

2.LS-Problem

In LS Problem

$$\min_x \|Ax - b\|$$

,

where A and b are not restricted at all, we have:

$$A^T Ax = A^T b.$$

Let $\mathfrak{L}$ is the set of all the minimizers to the LS problem. We have following facts.

Fact 1 $x^* = A^+ b \in \mathfrak{L}$.

Fact 2 $A\tilde{x_1} = A\tilde{x_2}$, for any $\tilde{x_1}, \tilde{x_2} \in \mathfrak{L}$.

Fact 3 For the optimization problem $\min_{x \in \mathfrak{L}} \|x\|$, there is a unique solution $x^* = A^+ b$.

Fact 4 We already have the result that if we choose some λ, LS problem will has a unique solution to the "modified" normal equation:

$$(A^T A + \lambda I)x = A^T b$$

that is,

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

In fact, let $\lambda \to 0$, we have

$$(A^T A + \lambda I)^{-1} A^T \to A^+$$

Fact 5 The projection of b on Col(A) is given by $A(A^T A)^{-1} A^T b$ (assuming $A^T A$ is invertible), here is a more general result

$$\hat{b} = AA^+ b = \lim_{\lambda \to 0} [A(A^T A + \lambda I)^{-1} A^T] b$$