MAT 653: Statistical Simulation

Instructor: Dr. Wei Li

2023-08-29

Back **Updated:** 2023-08-29 Statistical Simulation, Wei Li

Eigenvalue decomposition

Given a square matrix M so that $Mx = \lambda x$ for some nonzero vector x and some λ(scalar). We call λ an eigenvalue of M whose associated eigenvector is x.

Note: A is a square matrix, $\det(A) = \prod_{i=1}^{n} \lambda_i$. A square matrix is invertible if and only if its eigenvalues are non-zero.

Eigen-decomposition: A is a square and symmetric matrix, then we can write $A = P\Lambda P^T$ where

$$\begin{aligned} P &= [u_1, u_2, \cdots, u_n], \quad u\text{'s are the orthonormal eigenvector of } A \\ \Lambda &= diag\{\text{eigenvalues of } A\}. \end{aligned}$$

Note: If this symmetric matrix A is invertible, then we have $A^{-1} = P\Lambda^{-1}P^T$ where

$$\Lambda^{-1} = \begin{pmatrix} \ddots & 0 & 0 \\ 0 & \frac{1}{\lambda_i} & 0 \\ 0 & 0 & \ddots \end{pmatrix}$$

Example: $A = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$, $Av = \lambda v \implies (A - \lambda I)v = 0$. Here we need a non-zero solution, so

$$\det(A - \lambda I) = \begin{vmatrix} 9-\lambda & 0 \\ 0 & 4-\lambda \end{vmatrix} = 0 \implies \lambda_1 = 9, \lambda_2 = 4$$

For $\lambda_1 = 9$, since we know $Av_1 = 9v_1 \implies (A - 9I)v_1 = 0$, then we have

$$\begin{pmatrix} 0 & 0 \\ 0 & -5 \end{pmatrix} v_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

For $\lambda_2 = 4$, since we know $Av_2 = 4v_2 \implies (A - 4I)v_2 = 0$, then we have

$$\begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix} v_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Now $A = P\Lambda P^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ where $P = [v_1, v_2]$.

Singular value decomposition

For a $m \times n$ matrix A, $A^T A$ is a symmetric matrix. Let $\{v_1, \dots, v_n\}$ be the collection of all the orthonormal eigenvectors of $A^T A$ (eigen-decomposition); let $\lambda_1, ..., \lambda_n$ be the associated eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. We have $\|Av_i\|^2 = \lambda_i$, since

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

The singular values of A are squared roots of the eigenvalues of $A^T A$, denoted by $\sigma_i = \sqrt{\lambda_i}$; $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$

Fact: The rank of a matrix A is equal to the number of positive singular value of A.

SVD: Let A $(m \times n)$ be a matrix of rank r. There exists a matrix $\Sigma_{m\times n}$ of the following form, with D being a diagonal matrix whose entries are the first r (non-zero) 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 Singular Value Decomposition of A is:

$$A = U_{m\times m} \Sigma_{m\times n} V^T_{n\times n}$$

Where U and V are both orthogonal matrices: $V = [v_1, \cdots, v_n]$ consists of all orthonormal eigenvectors of $A^T A$; $U = [u_1, \cdots, u_m]$ consists the column u_i as follows

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

and these columns $\{u_1, \dots, u_r\}$ can be extended to $\{u_1, \dots, u_m\}$ as the orthonormal basis.

Matrix U and V are not uniquely determined in general, but have the property: $Col(U)$ spans $Col(A)$ and $Col(V)$ spans $Row(A)$.

Reduced (thin) 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 thin SVD to solve the Least Squares Problems as following:

$$\begin{aligned} Ax &= b, \\ A^T A x &= A^T b, \\ (U_r D V_r^T)^T (U_r D V_r^T) x &= (U_r D V_r^T)^T b, \\ V_r D I D V_r^T x &= V_r D U_r^T b, \\ D V_r^T x &= U_r^T b. \end{aligned}$$

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

Algorithm

1. Find the SVD of $A = U_r D V_r^T$;

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

3. Solve the diagonal system $Dw = y$, giving $w^* = D^{-1} y$;

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

The solution of the equation is $x^* = V_r D^{-1} U_r^T b = V \Sigma^{-1} U^T b$. We notice that this SVD method allows A, b to be arbitrary. Notice that $V_r D^{-1} U_r^T$ is like the "inverse" of A. Here we have the concept of generalized inverse.

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

2.LS-Problem

In LS Problem

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

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

$$A^T A x = A^T b.$$

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 $\lambda > 0$, 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

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

So

$$x^* = \lim_{\lambda \to 0} A(A^T A + \lambda I)^{-1} A^T = A^+ b.$$

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

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

Back **Updated:** 2023-08-29 Statistical Simulation, Wei Li