Singular Value Decomposition

Note Title 2024-10-02

$A \in \mathbb{R}^{n \times n}$ $\qquad A^T A$ is SPSD

$A^T A = V \Lambda V^T$ $\qquad V$ orthogonal $\qquad \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$

$$(AV)^T (AV) = V^T (A^T A) V = V^T V \Lambda V^T V = \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$\|(AV)_i\| = \sqrt{\lambda_i}$. We can define u_i such that

$(AV)_i = u_i \sigma_i \qquad \sigma_i = \sqrt{\lambda_i}$

u_i are the columns of an orthogonal matrix $U = [u_1 | u_2 | \dots | u_n]$

$$AV = U \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

Theorem: For every $A \in \mathbb{R}^{n \times n}$ there exist two orthogonal matrices $U, V \in \mathbb{R}^{n \times n}$ and a diagonal matrix

$$S = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \qquad \text{with } \sigma_i \geq 0$$

such that

$A = U S V^T$ $\qquad$ If we set $U = [u_1 | u_2 | \dots | u_n]$ $\quad V = [v_1 | v_2 | \dots v_n]$

$$A = [u_1 | u_2 | \dots u_n] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \begin{bmatrix} v_1^T \\ \hline v_2^T \\ \hline \vdots \\ v_n^T \end{bmatrix}$$

$$= u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \dots + u_n \sigma_n v_n^T = \sum_i u_i \sigma_i v_i^T$$

sum of n rank-1 matrices.

This dec. always exist, but we cannot use it to express powers of A:

$$A^2 = USV^TUSV^T \neq US^2V^T \quad \text{because} \quad V^TU \neq I$$

Generalization: for every $A \in \mathbb{R}^{m\times n}$, there exist $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ and a diagonal matrix $S \in \mathbb{R}^{m\times n}$ s.t.

$$A = USV^T$$

If $m > n$

$$A = U \, S \, V^T \qquad S = \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix} \in \mathbb{R}^{m\times n}$$

If $m < n$

$$A = U \, S \, V^T \qquad S = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_m & \end{bmatrix}$$

One can set $U = [u_1 | u_2 | \dots \, u_m]$ $\quad V = [v_1 | v_2 | \dots \, v_n]$

$$A = \sum_{i=1}^{\min(m,n)} u_i \sigma_i v_i^T$$

Singular value decomposition, or SVD.

One usually takes

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_{\min(m,n)} \geq 0$$

Thin SVD: Assume $m > n$. Then,

$$U = \begin{array}{c} \left[\begin{array}{c|c} U_0 & U_c \end{array}\right] m \\ \;\; n \quad\; m-n \end{array} \qquad S = \begin{array}{c} n \\ \left[\begin{array}{c} S_0 \\ \hline 0 \end{array}\right]\begin{array}{l} n \\ m-n \end{array} \end{array} \qquad S_0 = \left[\diagdown\right]$$

$$A = USV^T = \underbrace{\left[\begin{array}{c|c} U_0 & U_c \end{array}\right] \cdot \left[\begin{array}{c} S_0 \\ \hline 0 \end{array}\right]}_{= U_0 S_0 + U_c \cdot 0} V^T = U_0 S_0 V^T$$

$m \times n \quad n \times n \quad n$

If $m < n$ $\quad A = \square$

$$S = \boxed{S_0 \mid 0} \qquad V = \left[\begin{array}{c|c} V_0 & V_c \end{array}\right]$$

$$A = U \left[S_0 \;\; 0\right] \left[\begin{array}{c|c} V_0 & V_c \end{array}\right]^T = U \cdot S_0 \cdot V_0^T$$

$m \times m \quad m \times m \quad m \times n$

Thin (economy-sized) SVD.

Computational cost of thin SVD: $\quad O(mn \cdot \min(m,n))$

$[U0, S0, V] = \mathrm{svd}(A, 0)$ costs $\quad O(mn^2)$ if $m > n$

$O(m^2 n)$ if $m < n$

⚠ The full SVD has a higher cost $O(mn \cdot \max(m,n))$ because it needs to compute/return a potentially huge U (or V).

$A =$ $\square \; \square \; \square$

$m \times m \quad m \times n \quad n \times n$

⚠ σ_i are called singular values. They are ≥ 0

They are <u>not</u> the same as the eigenvalues

If $A = USV^T$ is an SVD, then

$$A^TA = (USV^T)^T(USV^T) = VS^TU^TUSV^T = V\,S^TS\,V^T$$

orthogonal, one the inverse of the other

S^TS is diagonal. If $m=n$

$$\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}^T \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}$$

eigenvalues of A^TA = squares of singular values of A

$m>n$

$$S^TS = \underbrace{\left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_n & \end{array}\right]}_{n\times m} \underbrace{\left[\begin{array}{cccc} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \\ \hline & & 0 & \end{array}\right]}_{m\times n} = \underbrace{\begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix}}_{n\times n}$$

$m<n$

$$S^TS = \underbrace{\left[\begin{array}{cccc} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \\ \hline & & 0 & \end{array}\right]}_{n\times m} \underbrace{\left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_m & \end{array}\right]}_{m\times n} = \underbrace{\left[\begin{array}{cccc|c} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \ddots & & 0 \\ & & & \sigma_m^2 & \\ \hline & & 0 & & 0 \end{array}\right]}_{n\times n}$$

eigenvalues of A^TA = squares of the sing. values of A padded with zeros if needed.

$\operatorname{Im} A$ = set of vectors of the form Ax for some x

Ker A = set of vectors z s.t. $Az = 0$

$$A = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + u_3 \sigma_3 v_3^T + \dots + \underbrace{u_{min} \sigma_{min} v_{min}^T}$$

Assume that

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_{min} = 0$$

(r non-zero sig. values)

Then, $A = u_1 \sigma_1 v_1^T + \dots + u_r \sigma_r v_r^T$.

$$Ax = u_1 (\sigma_1 v_1^T x) + u_2 (\sigma_2 v_2^T x) + \dots + u_r (\sigma_r v_r^T x)$$

Any vector $Ax \in \mathrm{Im}\, A$ can be written as a lin. comb. of $u_1, u_2, \dots u_r$

$\Rightarrow$ A has rank r

$\boxed{\text{rank of } A = \#\text{ of nonzero sig. values}}$

$\boxed{\mathrm{Im}\, A = \mathrm{span}(u_1, u_2, \dots u_r)}$

Ker A = vectors orthogonal to $v_1, v_2 \dots v_r$

$= \mathrm{span}(v_{r+1}, v_{r+2}, \dots v_{min})$

Matrix norms:

$\|v\|_2 = \sqrt{v^T v} = \sqrt{v_1^2 + v_2^2 + \dots v_n^2}$ for a vector v

$\|Uv\| = \|v\|$ for orthogonal U and for all v.

Def: the "induced norm" of a matrix is

$$\|A\| = \max_{v \in \mathbb{R}^n} \frac{\|Av\|}{\|v\|}$$ for a matrix $A \in \mathbb{R}^{m \times n}$

Because of this, $\boxed{\|Av\| \leq \|A\| \cdot \|v\|}$ for all $v \in \mathbb{R}^n$

Properties: for all matrices A, B, vectors v, scalars α:

$\|A\| \geq 0$ $\qquad \|A\| = 0$ iff $A = 0$

$\|\alpha A\| = |\alpha| \cdot \|A\|$

$\|A + B\| \leq \|A\| + \|B\|$

$\|AB\| \leq \|A\| \cdot \|B\|$ $\qquad$ ⚠ unlike $|\alpha\beta| = |\alpha| \cdot |\beta|$

$\|Av\| \leq \|A\| \cdot \|v\|$ for all vectors v

If you use this construction with $\|v\|_2$ as the base norm, you get the Euclidean matrix norm.

Def: the Frobenius norm of a matrix A is

$$\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$$

This is also a matrix norm, but it is different from the Euclidean norm.

Ex: Let $S = \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ & & \ddots \\ 0 & & \sigma_n \end{bmatrix}$ be a diagonal matrix.

Then, $\|S\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}$

$$\|S\|_2 = \max_{v \in \mathbb{R}^n} \frac{\|Sv\|}{\|v\|} = \max \frac{\left\| \begin{bmatrix} \sigma_1 v_1 \\ \vdots \\ \sigma_n v_n \end{bmatrix} \right\|}{\left\| \begin{bmatrix} v_1 \\ v_2 \\ v_n \end{bmatrix} \right\|}$$

$$= \max_v \frac{\sqrt{\sigma_1^2 v_1^2 + \dots + \sigma_n^2 v_n^2}}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \leq \frac{\sqrt{\sigma_{max}^2}\, \cancel{\sqrt{v_1^2 + \dots + v_n^2}}}{\cancel{\sqrt{v_1^2 + \dots v_n^2}}} = \sigma_{max}$$

For all v, $\sigma_1^2 v_1^2 + \sigma_2^2 v_2^2 + \dots + \sigma_n^2 v_n^2 \leq \sigma_{max}^2 (v_1^2 + \dots + v_n^2)$

$$\|S\|_2 = \max(\sigma_i) \neq \|S\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_n^2}$$

From the definition of these norms:

$\|A\| = \|AU\|$ for all orthogonal U

and similarly $\|A\| = \|UA\|$ for all orthogonal U

For both the Euclidean and Frobenius norm.

If you have an SVD of A available

$A = USV^T$, then

$$\|A\| = \|USV^T\| = \|S\| \begin{cases} \text{Euclidean: } \|S\| = \sigma_1 \\ \text{Frobenius: } \|A\|_F = \sqrt{\sigma_1^2 + \dots + \sigma_n^2} \end{cases}$$

Eckhart-Young theorem:

The solution of the problem

$$X_k = \arg \min_{\text{rank } X \leq k} \|A - X\|$$

is equal to

$$X_K = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T + \dots + u_K \sigma_K v_K^T$$

where $A = USV^T$, $U = [u_1 \cdots u_m]$, $V = [v_1 \cdots v_n]$ $S = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{\min} \end{bmatrix}$

is an SVD of A.

This is true for <u>both</u> the Euclidean and the Frobenius norm.

	Test 1	Test 2	Test 3
Student A	8	5	6
B	10	9	10
C	3	2	2
D	7	7	7

$= A$

Q_1: what happens if A has rank 1?

rank 2?

Can we give an interpretation of the fact in terms of the performance of our students?