SINGULAR VALUE DECOMPOSITION AND ITS APPLICATIONS

DEEPESH K P

NITC

Spectral theorem for matrices

Let A be a symmetric matrix. Then we can orthogonally diagonalize A. That is, we can find orthogonal matrix P such that

$$A = PDP^T$$

where P is an orthogonal matrix and D is a diagonal matrix.

- D contains eigenvalues of A (real) as its diagonal entries
- P contains (orthonormal) eigenvectors of A as its columns.
- $P^T P = I = PP^T$, Orthogonal matrix
- P rotates any vector by an angle (no change in magnitude)..
- D stretches or compresses vectors

Further questions

- Why only square matrices?
- Can we do this when A is a rectangular real matrix?
- What are the uses?

FURTHER QUESTIONS

- Why only square matrices?
- Can we do this when A is a rectangular real matrix?
- What are the uses?
- Our aim is to have
$$A = U\Sigma V^T,$$
U, V orthogonal and Σ is 'diagonal' in a way!
- A is an $m \times n$ matrix, $\Sigma_{m\times n}$, $U_{m\times m}$, $V_{n\times n}$
- Plenty of applications!

If A is a 3×5 matrix,

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{bmatrix}$$

and U is 3×3, V will be 5×5.

- Let $A_{m \times n}$ be an $m \times n$ matrix.
- $A^T A_{n \times n}$ and $AA^T{}_{m \times m}$ are square, symmetric, positive matrices..
- Spectral theorem $\Rightarrow A^T A = PDP^T$ and $AA^T = Q\tilde{D}Q^T$.
- Eigenvalues of $A^T A$ and AA^T are same, except for zeros..

  $$A^T A v = \lambda v \Rightarrow AA^T(Av) = \lambda(Av) \Rightarrow Av \text{ an eigenvalue if } \neq 0.$$

  If $Av = 0$, then v corresponds to $\lambda = 0$
- If $A_{n \times n}$ has rank r, then $n - r$ eigenvalues are 0 (see $N(A)$)..
- Let $\lambda_1, \lambda_2, \ldots, \lambda_r$ be the nonzero eigenvalues ($r = rank(A)$)..
- λ_i are all positive..
- P contains (orthonormal) eigenvectors of $A^T A$ as its columns.
- Q contains (orthonormal) eigenvectors of AA^T as its columns.

The reverse calculation!

For a moment assume $A = U\Sigma V^T$.

- $A^T A = (V\Sigma^T U^T)(U\Sigma V^T) = V\Sigma^2 V^T$.
- $\Sigma^2 = \Sigma^T\Sigma$ contains eigenvalues (positive) of $A^T A$
- So Σ contains square roots of eigenvalues of $A^T A$
- Also V can be taken as orthogonal eigenvectors of $A^T A$. (i.e., P)
- $AA^T = (U\Sigma V^T)(V\Sigma^T U^T) = U\Sigma^2 U^T$.
- So U can be taken as orthogonal eigenvectors of AA^T. (i.e., Q)
- But what guarantees A can be written so?

If A is a 3×5 matrix,

$$\Sigma\Sigma^T = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{bmatrix} \sim \Sigma^2$$

and U is 3×3, V will be 5×5.

We have $A^T A = PDP^T$

- Since $A^T A$ (also AA^T) is a positive definite matrix, D contains only positive entries on the diagonal.
- $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_m)$, assume increasing order.
- Let $\sigma_i = \sqrt{\lambda_i}$ and $D^{\frac{1}{2}} = diag(\sigma_1, \sigma_2, \ldots, \sigma_m)$
- Define $|A| = VD^{\frac{1}{2}} V^T$ (modulus operator)
- Then $|A|^2 = A^T A$ ..
- What is connection between A and $|A|$?
- Polar decomposition!

Matrix Multiplication

Suppose v_1, v_2, v_3 are column vectors and $B = [v_1\ v_2\ v_3]$. Then

$$AB = A\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} Av_1 & Av_2 & Av_3 \end{bmatrix}$$

Suppose u_1, u_2, u_3 are row vectors and $A = [u_1\ u_2\ u_3]^T$. Then

$$AB = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} B = \begin{bmatrix} u_1 B \\ u_2 B \\ u_3 B \end{bmatrix}$$

$$AD = A\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} = \begin{bmatrix} av_1 & bv_2 & cv_3 \end{bmatrix}$$

How to establish SVD

Consider $A_{m\times n}$ (matrix with $n \leq m$) (other case is similar)

- $A : \mathbb{R}^n \to \mathbb{R}^m$
- $A^T A$ is real symmetric, positive matrix $n \times n$ matrix.
- $A^T A$ has positive eigenvalues $\lambda_i = {\sigma_i}^2$ and orthonormal eigenvectors $v_1, v_2, \ldots, v_n$ in $\mathbb{R}^n$.
- $rank(A) = r \leq n \leq m$. (Think $r = 3$, $n = 4$, $m = 5$)
- $\sigma_{r+1}, \ldots, \sigma_n = 0$..
- Define $u_1 = \frac{1}{\sigma_1} Av_1,\ u_2 = \frac{1}{\sigma_2} Av_2, \ldots\ u_r = \frac{1}{\sigma_r} Av_r,$ in $\mathbb{R}^m$
- Or $Av_1 = \sigma_1 u_1,\ Av_2 = \sigma_2 u_2, \ldots, Av_r = \sigma_r u_r$

- $\{u_1, u_2, \ldots, u_r\}$ are orthonormal vectors in $\mathbb{R}^m$

$$\langle u_i, u_j \rangle = \delta_{ij}$$

- Hence independent and can be extended to a basis for $\mathbb{R}^m$
- Gramm-Schmidt gives ONB for $\mathbb{R}^m$ (extension).
- Let it be $\{u_1, u_2, \ldots, u_r, \ldots, u_m\}$.
- Call $U_{m \times m} = [u_1\ u_2\ \ldots\ u_r\ \ldots\ u_m]$, $V_{n \times n} = [v_1\ v_2\ \ldots\ v_n]$
- $\Sigma_{m \times n} = diag(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, 0, \ldots, 0)$ (rectangular)

If $m = 5, n = 4, r = 3$

$$\Sigma_{5\times 4} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

CLAIM: $A = U\Sigma V^T$

- Enough to show $AV = U\Sigma$
- $A : \mathbb{R}^n \to \mathbb{R}^m$
- $v_i \in \mathbb{R}^n$, $Av_i \in \mathbb{R}^m, u_j \in \mathbb{R}^m$
- $Av_1 = \sigma_1 u_1,\ Av_2 = \sigma_2 u_2, \ldots, Av_r = \sigma_r u_r, Av_{r+1} = 0, \ldots, Av_n = 0$

$$\begin{aligned} AV &= A[v_1\ v_2\ \ldots\ v_r\ \ldots\ v_n] \\ &= [Av_1\ Av_2\ \ldots\ Av_r\ \ldots\ Av_n]_{m\times n} \\ &= [\sigma_1 u_1\ \ \sigma_2 u_2\ \ \ldots \sigma_r u_r\ 0u_{r+1}\ \ \ldots\ 0u_n] \end{aligned}$$

$$AV = \begin{bmatrix} u_1 & u_2 & \dots & u_n & \dots u_m \end{bmatrix}_{m \times m} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & 0 = \sigma_n \\ 0 & 0 & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & 0 \end{bmatrix}_{m \times n}$$

This gives $AV = U\Sigma$ or

$$A = U\Sigma V^T,$$

which is the SVD.

Example

Consider

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Then

$$A^T A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \text{ and } A^T A = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- Eigenvalues are $\lambda = 3, 1$ (and 0)
- Singular values are $\sqrt{3}, 1$ (and 0)

Orthogonal Eigenvectors of $A^T A$ gives

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Take $u_1 = \frac{1}{\sqrt{3}} A v_1$ and $u_2 = \frac{1}{1} A v_2$ and u_3 as any orthonormal vector in $\mathbb{R}^3$:

$$U = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

we get $A = U \Sigma V^T$.

EFFECTIVE RANK OF A MATRIX

- Even minor calculation errors can change the rank of a matrix
- The matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4.01 \end{bmatrix}$$

  has rank 2, whereas one row is 'almost' dependent on the other.
- Practically rank is 1. But any program will give the rank as 2.
- $rank(A) = rank(AA^T)$(why?)
- For a diagonalizable matrix, rank is the number of nonzero eigenvalues..
- $rank(A)$= number of positive singular values for any matrix..
- For A, singular values are 5.0080... and 0.00199 (negligible)
- Hence, the size of singular values decides the size of the 'effective data' or the rank.

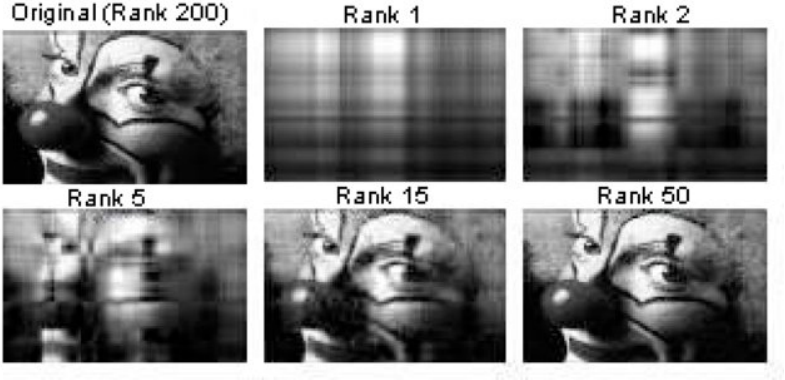
WHY $rank(A) = rank(A^T A)$?

- Note that $R(A)$ and $R(A^T A)$ may be in two different spaces!
- But $N(A)$ and $N(A^T A)$ are subspaces of same space.
- We can show $N(A) = N(A^T A)$..
- Hence nullity($A^T A$)=nullity(A).
- So $rank(A) = rank(A^T A)$

Image processing

- Suppose we want to send a picture having 1000×1000 pixels from a satellite
- Then the matrix has 1000000 entries in it
- You need to send a huge sized data to transfer the file
- The rank of the matrix may be 200 (remaining singular values will be zeros)
- many of these rows also may be negligible (singular values may be minimal (so effective rank will be smaller than 200)
- By sending the biggest 100 singular values and the corresponding vectors, it will give a very good picture.

IMAGE PROCESSING



Polar decomposition

- In complex numbers, we can split a complex number into its magnitude and angle form - the polar form $z = re^{i\theta}$.
- Can we split a square matrix $A = QS$ where Q is orthogonal, S is symmetric, positive matrix?
- Polar decomposition of a matrix ($A = SQ$ also)
- $A = U\Sigma V^T = U(V^T V)\Sigma V^T$
- Call $Q = UV^T$ and $S = V\Sigma V^T$ (RECALL $|A|$)
- Q does not change the size of any vector (rotates/reflects), S stretches or compresses
- Helpful in continuum mechanics
- when rotation happens and when the size gets changed during a continuous process!

MOORE-PENROSE INVERSE

B is said to be a Moore-Penrose inverse of A if

- $ABA = A$
- $BAB = B$
- $(AB)* = AB$
- $(BA)* = BA$

We denote $B = A^{\dagger}$

- $A^{\dagger}$ exists
- $A^{\dagger}$ is unique
- If A^{-1} exists, $A^{\dagger} = A^{-1}$.

Why it exists?

$\frac{1}{x}$ exists whenever $x \neq 0$. So for the $m \times n$ block diagonal matrix

$$D = \begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix}$$

where D_1 is a square diagonal matrix, and is invertible. So define

$$D^\dagger = \begin{bmatrix} {D_1}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

where $D^\dagger$ is an $n \times m$ matrix and $DD^\dagger D = D$ ... Why unique?

M-P Inverse for diagonal matrices

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \Sigma^\dagger = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_3} & 0 \end{bmatrix}$$

- Calculate $\Sigma\Sigma^\dagger$ and $\Sigma\Sigma^\dagger\Sigma$?
- For an $m \times n$ matrix A, we can take $A^\dagger = V\Sigma^\dagger U^T$
- $A^\dagger$ is the Moore Penrose inverse of A.

Why it is unique?

Suppose there is another matrix B like $A^\dagger$.

- Observe that $AA^\dagger = AB$

$$\begin{aligned} AA^\dagger &= ABAA^\dagger \\ &= (AB)^*(AA^\dagger)^* = B^*A^*A^{\dagger *}A^* = B^*(AA^\dagger A)^* \\ &= (AB)^* = AB \end{aligned}$$

- Similarly $A^\dagger A = BA$
- Hence $A^\dagger = B$

$$A^\dagger = A^\dagger AA^\dagger = A^\dagger AB = BAB = B$$

Solution of system of equations

We are interested in solving system $Ax = b$.

- For the equation $Ax = b$, there may or may not be solutions.
- Plenty of solutions/unique solution/no solution are possible.
- If there are plenty of solutions, find the solution of minimal norm
- If no solutions, find approximate solution of minimal norm

Solution of system of equations

- The equation $Ax = b$ has a solution if and only if $b \in R(A)$.
  - Since $R(A)$ is spanned by column vectors, b must be in the column space for a solution.
  - This means b should be a depending on column vectors
  - This happens iff $rank(A) = rank(Ab)$
- There is unique solution iff $N(A) = 0$
  - Suppose x_0 and x_1 are two solutions of $Ax = b$
  - $x_1 - x_0$ should be solution of $Ax = 0$. That is $x_1 - x_0 \in N(A)$
  - $x_1 \neq x_0$ iff $N(A) \neq \{0\}$.
  - So unique solution iff $N(A) = \{0\}$.
  - Rank-nullity theorem implies $rank(A) = rank(Ab) = n$

Solution of system of equations

- There are infinite number of solutions iff $N(A) \neq \{0\}$..
  - We can show that, if x_0 is a solution, each $z \in N(A)$ gives $x_0 + z$ a solution of same equation.
  - Conversely, whenever x_1 is any other solution, $x_1 = x_0 + z$ for some $z \in N(A)$, since $x_1 - x_0 \in N(A)$.
  - Due to rank-nullity theorem, $rank(A) < n$, the number of variables iff there are infinite number of solutions.

Summary

$Ax = b$ has

- No solution iff $rank(A) \neq rank(Ab)$
- Unique solution iff $rank(A) = rank(Ab) = n$
- Infinitely many solutions iff $rank(A) = rank(Ab) < n$

LEAST SQUARE METHOD

If there are infinite number of solutions, how to find a 'suitable' solution?

- Solution of smallest size!
- $Ax = b$ solvable iff $b \in R(A) = R(AA^\dagger)$
  - $y \in R(A) \implies y = Ax = AA^\dagger Ax \in R(AA^\dagger)$
  - $y \in R(AA^\dagger) \implies y = AA^\dagger x \in R(A)$
- But $P = AA^\dagger$ is a projection. So $b = AA^\dagger b$.
- So $Ax = b = AA^\dagger b$ gives $A^\dagger b$ is a solution!
- Among all solutions of $Ax = b$, $A^\dagger b$ has the minimal norm..

Why it is Least square solution?

Note that $A^\dagger b$ satisfies the equation $Ax = b$.

- Take any other solution x_0 of $Ax = b$
- $x_0 \in H = R(A^\dagger A) \oplus N(A^\dagger A)$, orthogonal sum
- $x_0 = A^\dagger b + (x_0 - A^\dagger b)$
- $\|x_0\|^2 = \|A^\dagger b\|^2 + \|(x_0 - A^\dagger b)\|^2$
- $\|x_0\| \geq \|A^\dagger b\|$
- Among all solutions of $Ax = b$, $A^\dagger b$ has minimal norm.
- So it gives least square solution.

Approximate solution for inconsistent systems

The equation $Ax = b$ has a solution if and only if $b \in R(A)$.

- Suppose $b \notin R(A)$. Then no solution.
- Consider $A^\dagger Ax = A^\dagger b$ has a solution since $b \in R(A^\dagger A) = R(A^\dagger)$..
- $x = A^\dagger b$ is a solution to the above also.
- But it need not satisfy $Ax = b$. So it is an approximate solution.
- If A invertible, it coincide with the actual solution.
- $\|Ax - b\|^2 = \|Ax - AA^\dagger b\|^2 + \|AA^\dagger b - b\|^2$
- See that $\langle Ax - AA^\dagger b, AA^\dagger b - b\rangle = 0$
- $\|Ax - b\|^2 \geq \|AA^\dagger b - b\|^2$
- Among all approximate solutions of $Ax = b$, $A^\dagger b$ gives least error solution.

QNS

- Why low rank is decided by size of singular values?
- What is the relation between eigenvalues and singular values?
  λ_i lies between largest and smallest singular values
- What is Carl inequality?
  sum of modulii of eigenvalues is less than or equal to sum of singular values.
- How approximation numbers emerges out of singular values?
  Using a maximum criterion for singular values.

REFERENCES

Gilbert Strang: Linear Algebra and its applications, Fourth Edn, Cengage, 2006.

THANK YOU