L10-1.a

Lecture 10

$A \in \mathbb{R}^{m \times n}$ with rank$(A) = r$ and $A = U \Sigma V^T$ be the SVD.

Lets revise

Rank of a mt. $(r(A))$

↳ is the positive integer r such that there exists at least one r-rowed square mt with non-vanishing determinant while $(r+1)$ or more rowed matrices have vanishing determinants.

Thus rank of a mt is the largest order of a non-zero minor of mt.

→ Rank of A and A^T is same

→ Rank of null mt is zero.

→ for $A^{m \times n}$. rank$(A) \leq \min(m, n)$

⇒ Only the first r $(= \text{rank}(A))$ singular values of A are non-zero.

⇒ range(A) is given by first 'r' columns of mt U

⇒ null(A) is given by last '$n-r$' columns of V.

⇒ range(A^T) is given by first 'r' columns of mt V

⇒ null(A^T) is given by last '$m-r$' columns of mt U.

Scanned with OKEN Scanner

Pseudo Inverse

The (Moore-Penrose) pseudoinverse of a mt generalizes the notion of an inverse, somewhat like SVD generalizes diagonalization.

As we know, not every mt has an inverse.

But every mt has a pseudoinverse, even non-square mt

We know $A = U \Sigma V^T$

Let rank(A) = r

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0.$$

$$A^+ = (U \Sigma V^T)^+ = (V^T)^{-1} \Sigma^+ U^{-1} = V \Sigma^+ U^T$$

Let A be a 3x3 mt + singular values are $\sigma_1, \sigma_2, 0$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \Sigma^+ = \begin{bmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & \boxed{0} \end{bmatrix}$$

$\frac{1}{0} \rightarrow 0$ replace

Scanned with OKEN Scanner

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{bmatrix} \qquad \Sigma^+ = \begin{bmatrix} 1/\sigma_1 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 \\ 0 & 0 & 1/\sigma_3 & 0 \end{bmatrix}$$

4×3 mx $\longrightarrow$ 3×4 mx

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \end{bmatrix} \longrightarrow \Sigma^+ = \begin{bmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 1/\sigma_3 \\ 0 & 0 & 0 \end{bmatrix}$$

3×4 mx $\qquad$ 4×3 mx

Remember if σ is closer to zero or zero, replace $\frac{1}{\sigma}$ by zero.

Ex:

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$$

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

$$= \begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix} \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} +1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{pmatrix}$$

Then

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

Scanned with OKEN Scanner

$$A^+ = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} \frac{1}{6\sqrt{10}} & 0 \\ 0 & \frac{1}{3\sqrt{10}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}^T$$

Find range(A) and null(A)

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

SVD

$$= \underbrace{\begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix}}_{U} \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix} \underbrace{\begin{pmatrix} 1/3 & -2/3 & +2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{pmatrix}^T}_{V}$$

$$\text{range}(A) = \left(\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -3 \end{pmatrix} \right)$$

$$\text{Null}(A) = (2, -2, 1)$$

Scanned with OKEN Scanner

Matrix norms

p-norm of a mt is simply an extension of the vector p-norm to matrices.

→ Although the definition of a mt p-norm vary based on the context.

They are generally based on two forms:-

1) Induced p-Norm (Operator p-Norm)
2) Entrywise p-Norm (Element-wise p-Norm)

1) Induced p-Norm

→ is derived from vector p-norms.

For m×n mt A, it is defined as

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

sup → supremum
concept from mathematical analysis that refers to the least upper bound in a set

In norm, it supremum captures largest possible value of certain expression over all valid inputs.

Scanned with OKEN Scanner

x is vector of appropriate size
(usually with dim n, as A is m×n)

$\|Ax\|_p$ is p-norm of mtx product Ax

$\|x\|_p$ is p-norm of vector x

sup → looking for max. possible value of the ratio $\frac{\|Ax\|_p}{\|x\|_p}$, over all vectors x except x=0.

→ Induced p-norm measures the maximum scaling factor by which mtx can increase p-norm of a vector.
→ how much a mtx can "stretch" a vector.

// p=1 (Induced 1-Norm)
Maximum absolute column sum

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}|$$

// p=∞ (Induced Infinity Norm)
Induced ∞ norm is max. absolute row sum

Scanned with OKEN Scanner

// $p = 2$ (Spectral Norm)

The induced 2-norm is the largest singular value of the mtx A.

This gives insight into the "stretching" of mtx done on vectors in term of the Euclidean distance

$$\|A\|_2 = \sigma_1(A) = \max_{\lambda_i}\{|\lambda_i|\}$$

where σ_1 denotes largest singular value

→ Entry wise p-Norm (Element-wise p-Norm)

Another way to apply norm is to elements of the mtx directly, As it is based on elements, it is called Entry wise p-norm or element wise p-norm.

For mtx $A_{m\times n}$

// $p=1$ (Entry wise 1-Norm)

Sum of absolute values of all the elements in the mtx.

$$\|A\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$$

Scanned with OKEN Scanner

$p = 2$ (Frobenius Norm).

Square root of the sum of squares of all elements.

$$\|A\|_2 = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2} = \|A\|_F$$

Also known as Euclidean Norm $\|A\|_E$

$p = \infty$ (Entry wise ∞ - Norm)

Maximum absolute value among all elements of mat

$$\|A\|_\infty = \max_{i,j} |a_{ij}|$$

For a $m \times n$ mat A; the Frobenius norm can be written as $\|A\|_F = \sqrt{\text{trace}(A^T A)}$

trace $(A^T A)$ is sum of the diagonal elements of the resulting mat $A^T A$.

→ trace of a square mat is equal to the sum of its eigenvalues, assuming mat has a full set of eigenvalues. $tr(B) = \sum_{i=1}^{n} \lambda_i$

Scanned with OKEN Scanner

Frobenius norm of mtx A is related to its singular value.

If $\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ are singular values of A (r is the rank of the mtx)

Then Frobenius norm expressed as

$$\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2}.$$

singular values are related to eigen values of $A^T A$.

Ex

$$A = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix}$$

$$\|A\|_1 = \max(5+1+2,\ 4+2+1,\ 2+3+0)$$

$$= 8$$

$$\|A\|_\infty = \max(5+4+2,\ 1+2+3,\ 2+1+0)$$

$$= 11$$

$$\|A\|_F = \sqrt{25+16+4+1+4+9+4+1+0}$$

$$= \sqrt{64} = 8$$

Scanned with OKEN Scanner

Low rank approximation

Let $A \in \mathbb{R}^{m \times n}$ having rank$(A) \leq \min\{m, n\}$

The low rank approximation of A is to find another mt $A_k \in \mathbb{R}^{m \times n}$ which is having rank $k \leq r$ and approximate A.

SVD provide an easy way to get low-rank approximation solution

Suppose $A \in \mathbb{R}^{m \times n}$
where $m \geq n$

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

$$= \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

Then

Scanned with OKEN Scanner

4a

Then the best-k approximation to A is

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T \qquad \text{where } k \leq \text{rank}(A)$$

in the sense that

$$\|A - A_k\| \leq \|A - \tilde{A}\|$$

for any $\tilde{A} \in \mathbb{R}^{m \times n}$ with rank$(\tilde{A}) \leq k$

Let A be a 6×6 mt with

$\sigma_1 = 4, \sigma_2 = 2; \sigma_3 = 1; \sigma_4 = 0.3, \sigma_5 = 0.1, \sigma_6 = 0.02$

rank is 6 as all σ are non-zero

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

$$A = U \begin{bmatrix} \sigma_1 & & & & & 0 \\ & \sigma_2 & & & & \\ & & \sigma_3 & & & \\ & & & \sigma_4 & & \\ & & & & \sigma_5 & 0 \\ 0 & & & & & \sigma_6 \end{bmatrix} V^T$$

I am interested in first 4 σ then

$$A_4 = \underset{6 \times 6}{U} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \\ 0 & \sigma_2 & 0 & 0 & \\ 0 & 0 & \sigma_3 & 0 & \\ 0 & 0 & 0 & \sigma_4 & \\ & & & & 0 \end{bmatrix} \underset{6 \times 6}{V^T}$$

$$\underset{6 \times 4}{\begin{bmatrix} & \end{bmatrix}} \underset{4 \times 4}{\begin{bmatrix} & \end{bmatrix}} \underset{4 \times 6}{\begin{bmatrix} & \end{bmatrix}}^T = A_{4_{6 \times 6}}$$

First 4 column of U

First 4 row of V^T

rank 4

Scanned with OKEN Scanner

Measure of quality of approximation is given by

$$\frac{\|A_k\|_F^2}{\|A\|_F^2} = \frac{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots + \sigma_k^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots + \sigma_r^2}$$

$$k \le r$$

Example

$$A = \begin{bmatrix} 3 & 2 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A = U S V^T$$

$$\begin{bmatrix} 0.91 & 0.42 & 0.02 \\ 0.41 & -0.87 & -0.26 \\ 0.09 & -0.24 & 0.97 \end{bmatrix} \begin{bmatrix} 4.04 & 0 & 0 \\ 0 & 1.70 & 0 \\ 0 & 0 & 0.87 \end{bmatrix} \begin{bmatrix} 0.67 & 0.73 & 0.08 \\ 0.65 & -0.54 & -0.53 \\ 0.35 & -0.41 & 0.84 \end{bmatrix}^T$$

$$= \begin{bmatrix} 0.91 & 0.42 \\ 0.41 & -0.87 \\ 0.09 & -0.24 \end{bmatrix} \begin{bmatrix} 4.04 & 0 \\ 0 & 1.70 \end{bmatrix} \begin{bmatrix} 0.67 & 0.73 \\ 0.65 & -0.54 \\ 0.35 & -0.41 \end{bmatrix}^T$$

$$A_2 = \begin{bmatrix} 2.99 & 2.01 & 0.98 \\ 0.02 & 1.88 & 1.1 \\ -0.07 & 0.45 & 0.29 \end{bmatrix}$$

This is important when one deals with large mp such as image.

One can utilize it in the image processing.

Scanned with OKEN Scanner

Let say we have collection of photographs.
each of which is 640x 480 pixel.
Then it can be represen by $\mathbb{R}^{640\times 480}$, $\mathbb{R}^{307200}$ ~~vecto~~.
vector in $\longrightarrow$

Given a collection of such images
↳ one can translate into collection of vectors in
high dimensional space.

Assuming that faces occupy a very small region of the
high dimensional space, one ought to be able to
find a relatively small dimensional subspace W that
captures most of the data.
Such a space would a low dimensional approx. to
the column space of photo.

Scanned with OKEN Scanner