16/10/42

S Deepak Narayanan

classmate

Date

Page

(2) $A = U\Sigma V^T$

Suppose $A \in \mathbb{R}^{n\times d}$ $U \in \mathbb{R}^{n\times r}$ $\Sigma \in \mathbb{R}^{r\times r}$ $V^T \in \mathbb{R}^{r\times d}$

(i) AA^+. Suppose rank $(A) = r$

$\Rightarrow U\Sigma V^T V \Sigma^{-1} U^T$

$\Rightarrow U\Sigma\Sigma^{-1}U^T$

$\Rightarrow UU^T$

V^TV and U^TU are always identity matrices.

UU^T = Identity only if

~~U^TU is~~ U is square, since then only

$UU^T = U^TU = I$.

U must be square, orthonormal.

ii) $U \in \mathbb{R}^{n\times r}$, $U \notin \mathbb{R}^{n\times n}$ for $r \neq n$

$\Rightarrow$ Inverse when U is square orthogonal.

(ii) A^+A

$V\Sigma^{-1}U^TU\Sigma V^T$

$\Rightarrow V\Sigma^{-1}\underbrace{U^TU}_{I}\Sigma V^T \Rightarrow V\Sigma^{-1}\Sigma V^T$

$\Rightarrow VV^T$

$VV^T = I$ only when V has full rank $(\because V^TV = I)$

(iii) $AA^+A \Rightarrow \underbrace{UU^T}_{I}.U\Sigma V^T$

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

UU^T and VV^T are projection matrices onto column & row spaces of A respectively. Makes sense as that's all we need to project onto the subspace.

Scanned by CamScanner

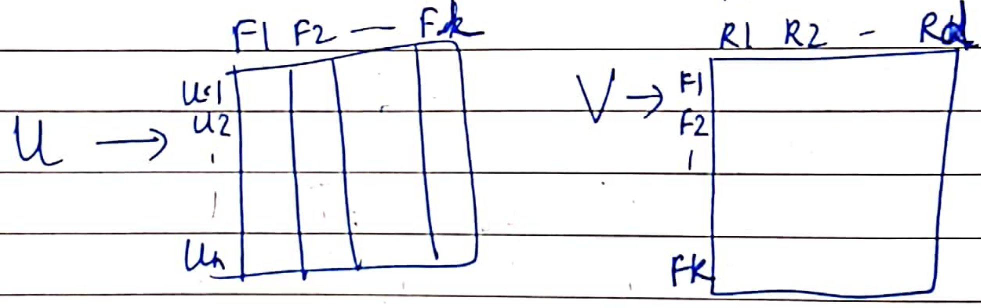
5) $A = U\Sigma V^T$

$$U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} — v_1^T — \\ \vdots \\ — v_r^T — \end{bmatrix}$$

I am taking v_i to be the ~~first~~ i^{th} Column vector of V which is the ~~first~~ i^{th} row vector of V^T.
u_i — Column vector of U.

Essentially, we need to project restaurants and users into a common space.
This could be viewed as a "low-rank" approximation of A (User - Restaurant) matrix.

~~Each of these~~
We can view as a matrix A as follows



Each user and their preference ~~for~~ feature i is u_i^T

u_i represents how feature i varies across all the users, or how different users rate feature i

v_i represents how feature i varies across all the restaurants.
These features could be Ambience, food, service, crowd, waiting time etc.

$\sigma_1 - \sigma_2$ is essentially the difference in the most important features. Suppose $\sigma_1 >> \sigma_2$, then only one feature is dominant, if $\sigma_1 \approx \sigma_2$, then both features (they're already the top ones) have almost equal importance.

(1) Consider any vector $\vec{v} \in \mathbb{R}^d$

Consider the subspace spanned by columns ~~space~~ of an matrix $\equiv$ Subspace spanned by the U in the SVD of A.

Suppose A had rank r;
consider the following SVD

$$\underset{n \times d}{A} = \underset{n \times r}{U} \underset{r \times r}{\Sigma} \underset{r \times d}{V},$$ where ~~each vector~~ $\{u_i\}_{i=1}^{r}$ is a linearly independent

The projection of $\vec{v}$ is, by definition, onto $\vec{u_i}$, is,

$$u_i^T \vec{v} \in \mathbb{R}.$$

We need the projection vector which is,

$$(u_i^T v)\vec{u_i} \text{ or,}$$

$$\vec{u_i}(\vec{u_i}^T \vec{v}).$$

For all u_i, we take to compute projection onto the subspace spanned by $\vec{u_i}$.

$$\Rightarrow \quad U U^T \vec{v} \text{ is the projection matrix.}$$

Now, consider $A(A^TA)^{-1}A^T$.

Clearly, if this has to be the projection matrix, A^TA has to be invertible.

Also consider $X = A(A^TA)^{-1}A^T$.

$$X^T = \left(A(A^TA)^{-1}A^T\right)^T$$

$$\Rightarrow (A^T)^T\left((A^TA)^{-1}\right)^T(A)^T$$

$$\Rightarrow A\left((A^TA)^T\right)^{-1}(A^T)$$

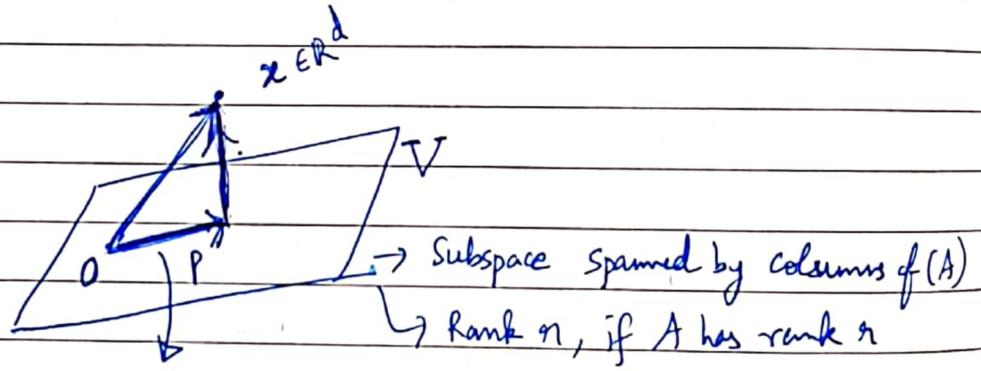
$$\Rightarrow A\left((A^TA)^{-1}\right)A^T$$

$$\Rightarrow X$$

$\Rightarrow X^T = X$

$\Rightarrow$ Projection matrix is symmetric if X is a projection matrix

Consider:



Projection of x onto V

We can see that $\vec{x} - \vec{p}$ is perpendicular to the subspace $\Rightarrow$ It is perpendicular to each vector ~~in V~~ that spans the colspace(A)

$$\Rightarrow A = \begin{bmatrix} | & | & \cdots & | \\ a_1 & a_2 & & a_d \\ | & | & & | \end{bmatrix}$$

Scanned by CamScanner

A^TA is invertible if we have full column rank for A.

$\Rightarrow \quad \vec{a_i}^T(\vec{x}-\vec{p}) = 0 \quad \forall\, i \in \{1,2,\cdots d\}$

$\Rightarrow \quad A^T(\vec{x}-\vec{p}) = 0$. But ~~project is really~~,

~~p lies on the V~~ $\Big\{$ ~~Consider $\vec{a_1},\vec{a_2},\cdots \vec{a_d}$~~

~~$\Rightarrow p = \sum_{i=1}^{d} \alpha_i a_i$~~ ~~$\vec{p} = Ax$~~

~~Projection is~~

~~$\vec{p}$ lies on V~~

~~$\Rightarrow \vec{p} = \sum_{i=1}^{d} \alpha_i \vec{a_i}$, a_i~~ $= \vec{p} = A\hat{x}$ for.

~~$\Rightarrow \vec{p} \in$ colspace(A)~~ $\hat{x} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{bmatrix}$

~~$\Rightarrow \vec{p} =$~~

$\vec{p}$ lies on V. If A had rank r, then,

$$\vec{p} = \sum_{i=1}^{r} \alpha_i \vec{a_i}.$$

$\Rightarrow \quad A^T(\vec{x} - A\vec{\alpha}) = 0$

$\Rightarrow \quad A^T\vec{x} = A^TA\vec{\alpha}$

$\Rightarrow$ If A^TA is invertible

$\Rightarrow \quad \vec{\alpha} = (A^TA)^{-1}A^T\vec{x}$

But we know, projection is $A\vec{\alpha}$

$\Rightarrow \quad UU^T$ ($U \to$ Span of Cols of A)

is same as

$$A(A^TA)^{-1}A^T$$

Hence proved.

~~Also $\Rightarrow UU^T = U^TU$~~ $\Rightarrow$

anan

3, 5, 4
2, 4, 6

2

p

$\frac{1}{3}+\frac{3}{3}\frac{1}{3}$
$1=2$

(4) Let P be the projection matrix.

We need $\vec{v}$, such that $\frac{\|P\vec{v}\|}{\|V\|}_{\|V\|_2=1}$ is maximised.

$\Rightarrow \|P\vec{v}\|_2$ must be maximized

By projection, we're taking "components of v.."

$\Rightarrow \|P\vec{v}\|_2 \leq \|v\|_2$

Really, because $\|v\|_2$ ~~be~~ must be 1,

$\|Pv\|_2 \leq 1$.

$\Rightarrow$ the v, that maximises the LHS and still satisfies the inequality is the one such that $\|PV\| = 1$

$\Rightarrow \sigma_1 = 1$.

Similarly, because we have a k rank subspace,

$\sigma_1, \sigma_2, \ldots \sigma_k = 1$ (k mutually orthonormal vectors).

$\Rightarrow \sigma_{k+1} = \sigma_{k+2} \ldots = \sigma_n = 0$

Physically, it means that the stretch that the projection matrix causes on the k dimensions is ~~almost~~ 1 $\Rightarrow$ It preserves the initial stretch or ~~only "shrinks"~~ the value of any k-d object along the k-dimensions.

U, V, only rotate. Its really σ that stretches the object.