SVD – PCA

ML – Lecture 17

③

Recall: vectors, vector space, span, basis, linear independence, eigen values, eigen vectors, orthogonality, symmetric positive definite, linear transformation, dimensionality; inner products, orthogonal decomposition

Some important Results:

Similarity
$\langle x, y\rangle = x^T y$
$= \|x\| \|y\| \cos$ (angle b/w x, y)

1. Orthogonal matrices preserve angles & lengths
2. Eigen values of a SPD are real & positive
3. Every real matrix (transformation) can be decomposed as

$$A = Q \Sigma Q^T ; \quad Q \equiv \text{orthogonal}, \quad \Sigma \equiv \text{upper triangular}$$

4. A set of orthogonal vectors are linearly independent
5. Eigen vectors corresponding to real & distinct eigen values are orthogonal
6. Transformations by orthogonal matrices (transformations) are reflections & rotations.
7. Any matrix-vector product is a linear transf!

(Angle preserving)

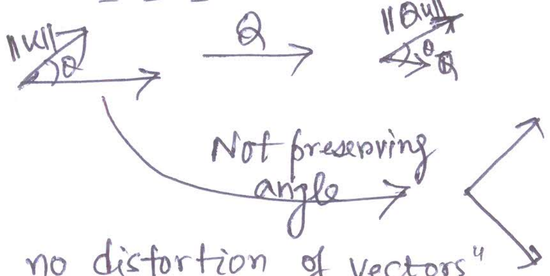
$\langle Qv, Qu\rangle = (Qv)^T Qu$
$= \langle Q^T Q v, u\rangle$
$= \langle v, u\rangle$; since $Q^T Q = I$

Q real; ALSO
$\langle Qu, Qu\rangle = \langle u, u\rangle$

$$\Rightarrow \underbrace{\|Qu\| = \|u\|}$$

(length preserving)

Ref: (i) Optimization model, Laurent el ghaoui.
(ii) CSPA, Raj Jain



" no distortion of vectors "

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\| \|v\|} ; \quad \cos\theta_T = \frac{\langle Qu, Qv\rangle}{\|Qu\| \|Qv\|} = \frac{\langle u, v\rangle}{\|u\| \|v\|}$$

" angle-preserving "

$$\boxed{A_{m\times n} = U_{m\times r} \Sigma_{r\times r} V^T_{n\times r}} \leftarrow \text{SVD}$$

An interesting way to break up matrices ⇒ matrix decomposition ⇒ Singular value decomposition (SVD).

Flashback: from Linear Algebra ⟩ $\underbrace{A = PDP^{-1}}_{\text{similarity transformation}}$, A is $(n \times n)$ ← (note)

P is invertible; D is diagonal

Useful for matrix exponentiation: $A^K = PD^K P^{-1}$ → (trivial)

columns of P are L.I eigenvectors of A

D contains eigenvalues of A

SVD: → FLASHBACK ←

Every matrix A factors into 3 pieces; $A = U\Sigma V^T$;

Σ: diagonal (stretching) with non-negative entries

U, V: orthogonal ⇓ (rotation)

Stretching ⇒ intuitive understanding of eigenvalues & eigenvectors

$$\underbrace{Ax}_{\text{matrix-vector product}} = \underbrace{\lambda x}_{\text{scalar x Vector}}$$

scaling the vector such that $\vec{x}$ doesn't fall off the span$(\vec{x})$

Details: My LA notes & supplementary sheets

$$A = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$$

← none square matrices

So, what's the difference b/w

$A = PDP^{-1}$ & $A = U\Sigma V^T$?

if P is orthogonal, $P^{-1} = P^T \Rightarrow A = PDP^T$ ⇒ still, not the same!

(i) A in Case I is square; A_{SVD} can be rectangular $(m \times n)$

(ii) P doesn't have to be orthogonal, U & V are orthogonal

D has eigenvalues; Σ has singular values

A & A_{SVD} have different bases!

? $A = U\Sigma V^T$; what are U & V^T?

$$A^T A = (U\Sigma V^T)^T A = V\Sigma^T \underbrace{U^T U}_{I} \Sigma V^T = V \boxed{\Sigma^T \Sigma} V^T$$

(diagonal) → WHY?

$$\Rightarrow A^TA = \underbrace{V(\text{diag})V^T}_{\boxed{\text{Same matrix}}} \quad ; \quad \text{Compare w/} \quad A = PDP^T$$

⇒ V has eigenvectors of A^TA	P has eigenvectors of A
Diag has eigen values of A^TA i.e A^2	D has eigen values of A

$$AA^T = U\Sigma V^T(U\Sigma V^T)^T$$

$$= U\Sigma \underline{V^T} V \Sigma^T U^T$$

$$AA^T = \underbrace{U\Sigma\Sigma^T U^T}$$

{ U has eigenvectors of AA^T
V ———— of AA^T }

eigenvalues (A^TA) = eigenvalues (AA^T)
= singular values of (A)

$$\text{diag} = \underbrace{\Sigma^T\Sigma}_{\text{square matrix \& symmetric}} \Rightarrow \text{eigenvalues } (A^TA) = \sigma^2 \text{ for } A$$

$$\boxed{\text{Singular values of } A \equiv \text{eigenvalues } (A^TA)}$$

$$\boxed{\Sigma^T\Sigma \text{ is symmetric \& PD}}$$

Ex:

$$A = \underbrace{\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}}_{\text{singular}} = \underset{\uparrow \atop U}{\frac{\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}}{\sqrt{5}}} \underset{\uparrow \atop \Sigma}{\begin{bmatrix} \sqrt{10} & \\ & 0 \end{bmatrix}} \underset{\uparrow \atop V^T}{\frac{\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}{\sqrt{2}}}$$

Application: Goal: $u_1 \sigma_1 v_1^T$

u_1: first eigenvector of U

σ_1: greatest σ in the band of σ's $\underbrace{(\sigma_1 \text{ -- } \sigma_n)}_{\text{spectrum}}$ contain the most information ⇒ = greatest variance in the matrix of rank 1 (Low rank projection)

v_1^T: first eigenvector of V

PCA ⟸

u_1 → a combination of some input data

v_1 → a combination of some other input data

σ_1 → largest singular value ⟹ $\boxed{u_1\sigma_1 v_1^T}$ ← principal component

Ex: (i) Page Rank

(ii) Article ranking within a journal (captures qualitative information)

Ⓐ [quality features ⟹ $\underbrace{| \ | \ \text{- - -} \ |}_{\text{articles}}$ $\{u_1\sigma_1v_1^T,\ u_2\sigma_2v_2^T,\ \text{- - - -},\ u_n\sigma_nv_n^T\}$ from $U\Sigma V^T$

Worked out example: $A = U\Sigma V^T \Rightarrow$ • $A^TA = V\Sigma^T\Sigma V^T$

• $AV = U\Sigma$

1) $A^TA = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \quad \end{bmatrix}$

$\text{Det}(A^TA - \lambda I) = \text{Det}\begin{bmatrix} \quad \end{bmatrix} = \cdots \Rightarrow \text{eig}(A^TA) = \underset{\lambda_1,\ \lambda_2}{\cdots}$

2) Use λ_1, λ_2 to compute eigenvectors (A^TA)

$A^TA - \lambda_1 I$; $v_1' = \begin{pmatrix} \quad \end{pmatrix}$; normalize ; $v_1 = \frac{v_1'}{\|v_1'\|}$

$A^TA - \lambda_2 I$; $v_2' = \cdots \longrightarrow v_2 =$

$V = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$

3) Singular values:

$\equiv$ sqroot of (eigenvalues)

$\Rightarrow (\sigma_1, \sigma_2) = \text{sqrt}(\lambda_1, \lambda_2) \cdots$

4) U matrix: Use $AV = U\Sigma$; solve ....

$\Rightarrow$ ~~[illegible]~~

Summary:

$A_{m\times n} = U_{m\times r}\, \hat{\Sigma}_{r\times r}\, V^T_{n\times r}$ not symmetric

A: i/p data matrix ; m×n (m documents, n terms)

U: Left singular vectors, m×r (m documents, r concepts)

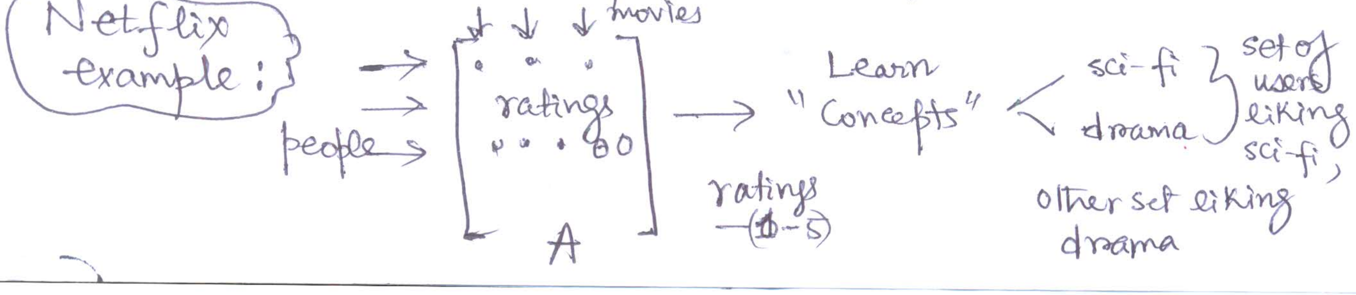
Σ: Singular values (~~r &~~ r ≡ rank(A); strength of each concept)

V: Right singular vectors, n×r (n terms, r concepts)

$m >> r,\ r = r,\ r >> n$

$\sigma_1 \geq \sigma_2 \geq \cdots 0$, U, V orthogonal

A-decomposition is unique.



Use MATLAB Command: ⑤

$$A = \begin{array}{c} \begin{array}{ccccc} \multicolumn{3}{c}{\text{sci-fi}} & \multicolumn{2}{c}{\text{drama}} \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}_{7\times5} \\ \begin{array}{ccccc} \uparrow & \uparrow & \uparrow & & \\ \text{movie 1} & \text{mov 2} & \text{mov 3} & \text{mov 4} & \text{mov 5} \end{array} \end{array} = \overset{U}{\begin{array}{r} \text{user1} \rightarrow \\ \\ \\ \end{array}\begin{bmatrix} \square & & \\ & \text{some \#'s} & \\ & & \end{bmatrix}_{7\times3}} \overset{\Sigma}{\begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & & \textcircled{\sigma_3} \end{bmatrix}_{3\times3}} \overset{V^T}{\begin{bmatrix} & \text{some \#'s} & \end{bmatrix}_{3\times5}}$$

7×3 ↑ = "user-concept similarity"

3×3 ↑ strength of concepts → sci-fi, drama

3×5 "movie to concept" similarity

$A \Longrightarrow$ first 3 columns are "sci-fi", last 2 columns → "drama"

$\Longrightarrow$

$$U \rightarrow \begin{array}{c} \begin{array}{ccc} \text{sci-fi} & \text{drama} & \\ \downarrow & \downarrow & \text{noise} \end{array} \\ \underbrace{\begin{bmatrix} 0.13 & 0.02 & \\ 0.41 & 0.07 & \\ 0.55 & 0.09 & \\ 0.68 & 0.11 & \\ 0.15 & -0.59 & \\ 0.07 & -0.73 & \\ 0.07 & -0.29 & \end{bmatrix}}_{\text{user-concept similarity}} \end{array}$$

people → rows

$$\Sigma = \begin{bmatrix} \textcircled{12.4} & & \\ & \textcircled{9.5} & \\ & & 1.3 \end{bmatrix}$$

12.4: strength of sci-fi > 9.5: strength of drama

$$V^T = \underbrace{\begin{array}{r|ccccc|} & \text{mov 1} & \text{mov 2} & \cdots & & \\ \text{sci-fi} \rightarrow & 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ \text{drama} \rightarrow & 0.12 & -0.02 & 0.12 & -0.69 & -0.69 \\ \text{noise} \rightarrow & & & & & \end{array}}_{\text{movie-concept similarity}}$$