For a symmetric matrix S

$$S = Q\Lambda Q^T$$

Q orthogonal

Λ diagonal

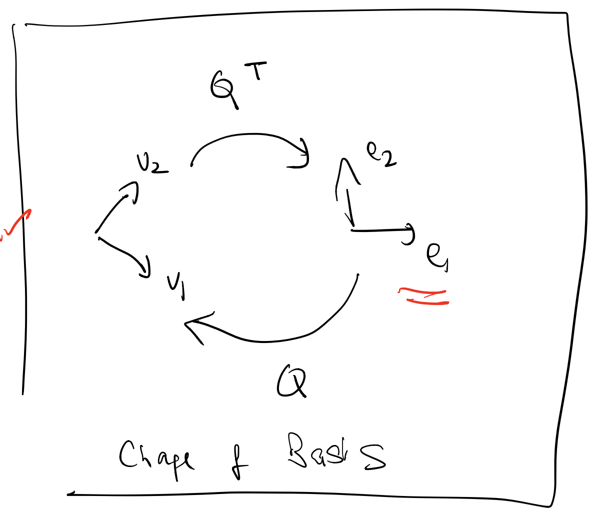
$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{bmatrix}$$ λ eigenvalues

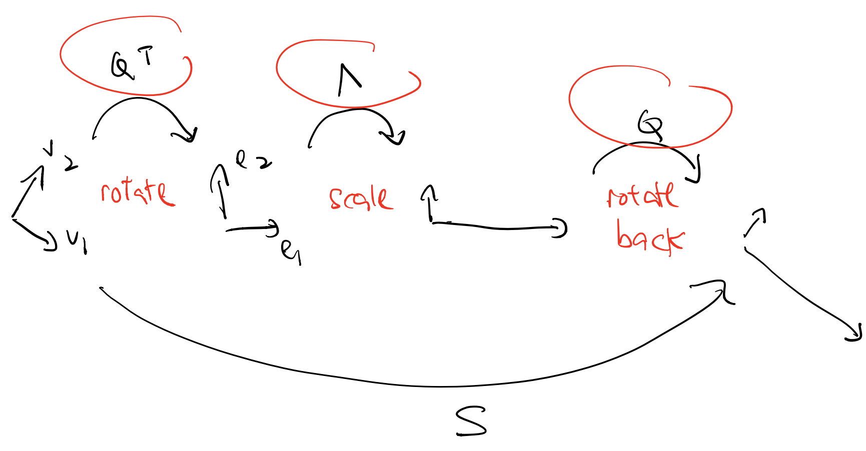
$$Q = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{bmatrix}$$ v eigenvectors

$$\begin{tikzcd} \cdot \arrow[r, "Q^T"] \arrow[d, "S"'] & \cdot \arrow[d, "\Lambda"] \\ \cdot & \cdot \arrow[l, "Q"] \end{tikzcd}$$

similar

$$S = Q\Lambda Q^T$$





$S \rightarrow$ symmetric signal ||

SVD works for ANY matrix A

rectangular

$$\boxed{A = U \Sigma V^T}$$

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

LU, Cholesky

QR Orthonormalization

$Q \Lambda Q^T$ Spectral

$U \Sigma V^T$ SVD

U, V orthogonal

Σ is "diagonal"

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & \end{array}\right]$$

$\sigma_i > 0$

$\sigma_1 \geq \sigma_2 \geq \dots$

$\sigma_1, \sigma_2, \dots, \sigma_r$ are ordered $\sigma_1 \geq \sigma_2 \dots$

$$\frac{\sigma_1}{\sigma_r} = \boxed{\frac{\sigma_{max}}{\sigma_{min}} = \underline{\text{Condition number}}}$$

CN

σ_i's are <u>singular</u> values

Can also write

$$\underset{m \times n}{A} = \underset{\underline{\underline{m}} \times r}{U} \; \underset{r \times r}{\Sigma} \; \underset{r \times \underline{\underline{n}}}{V^T}$$

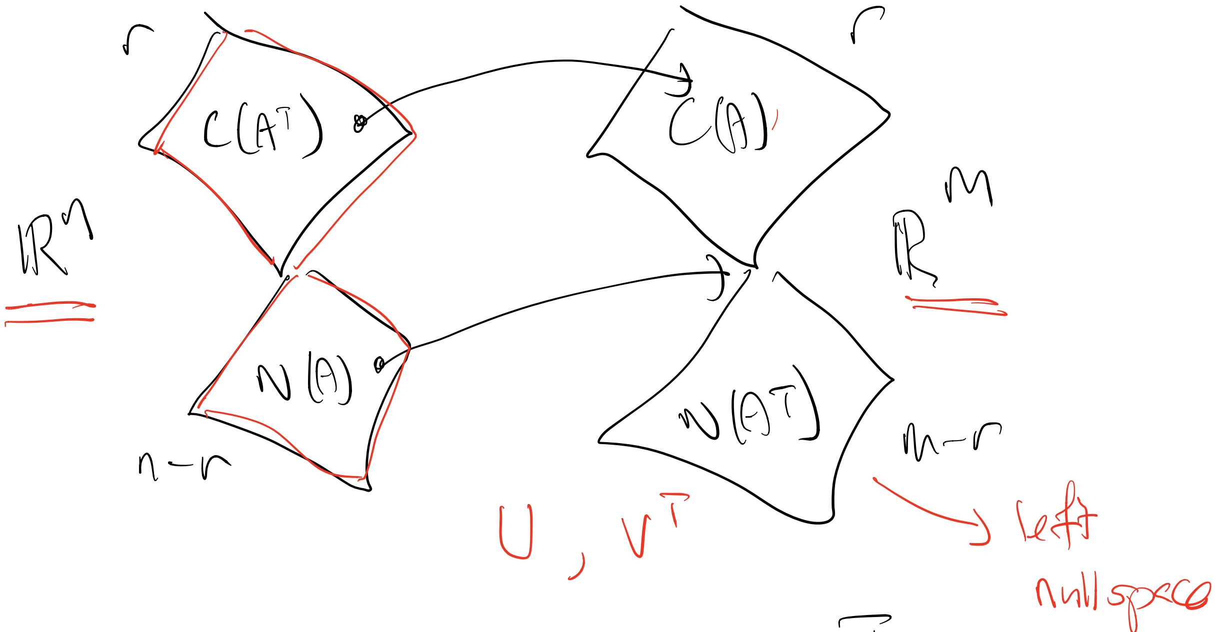
V eigenvectors of A^TA
U eigenvectors of AA^T $\Big\}$ symmetric

A^TA & AA^T have same set of r eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r$

where

$$\underline{\lambda_i = \sigma_i^2}$$

↓
eigenvalue of
A^TA or AA^T



$$\underset{m\times n}{A} = \underset{m\times m}{U}\ \underset{m\times n}{\Sigma}\ \underset{n\times n}{V^T}$$

$\overset{\mathbb{R}^m}{\{u_1, \dots, u_r\}}$ Basis for $\overset{\mathbb{R}^m}{C(A)}$

$\overset{\mathbb{R}^m}{\{u_{r+1}, \dots, u_m\}}$ Basis for $\overset{\mathbb{R}^m}{N(A^T)}$

$\overset{\mathbb{R}^n}{\{v_1, \dots, v_r\}}$ Basis for $\overset{\mathbb{R}^n}{C(A^T)}$

$\overset{\mathbb{R}^n}{\{v_{r+1}, \dots, v_n\}}$ Basis for $\overset{\mathbb{R}^n}{N(A)}$

$A = U\Sigma V^T$

$\Rightarrow \boxed{AV = U\Sigma}$

$\Rightarrow Av_i = \sigma_i u_i$

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

symmetric

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

$= \boxed{U\Sigma\Sigma^T U^T}$

U is eigenvector matrix for AA^T

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

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

$= \boxed{V\Sigma^T\Sigma V^T}$

V is eigenvector matrix for A^TA

Also AA^T & A^TA have same r eigenvalues

$\lambda_i = \sigma_i^2$

A^TA, AA^T

Low rank version

$$\boxed{A = U \Sigma V^T}$$

$$= \sum \sigma_i u_i v_i^T$$

$m \times 1 \quad 1 \times n$

$m \times n$

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$$

$u_i v_i^T$

u_i is $m \times 1$
v_i^T is $1 \times n$ } $u_i v_i^T$ is $m \times n$

So A is a sum of rank 1 matrices $u_i v_i^T$

whose contribution to A is weighted by σ_i

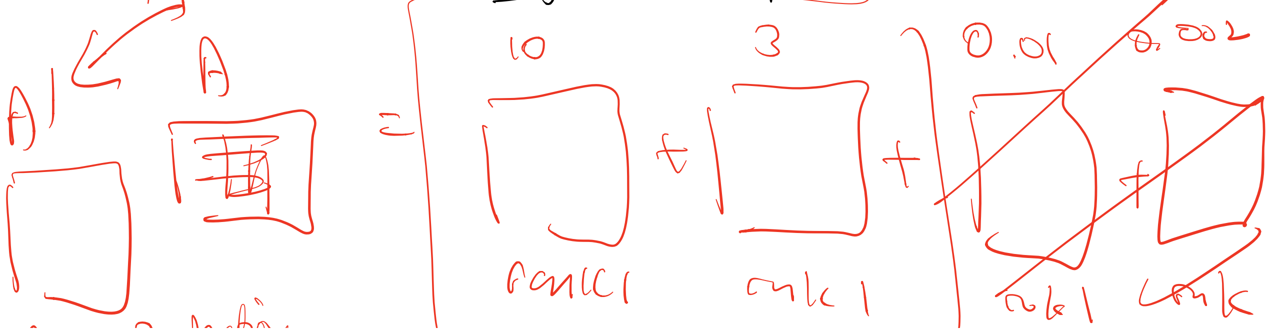
$\Rightarrow$ we can drop rank 1 matrices whose $\sigma_i \ll 1$

Suppose A is $1000 \times 1000 = 10^6$ values

Suppose only 100 σ_i's are non-negligible

Store 100 σ_i — 100 values
100 u_i — 100,000 values
100 v_i — 100,000 values } 200,100 values

Used for lossy data compression



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

$$\Sigma^+ = \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & 0 \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_r} \\ & & 0 & \end{bmatrix}$$

Σ^+ = Take reciprocals of non-zero values of Σ

Note $(\Sigma^+)^+ = \Sigma$ even though Σ is NOT invertible.

$\boxed{\text{Pseudoinverse}}$

$\boxed{A^{-1} = V\Sigma^{-1}U^T}$ if A square

$A = U\Sigma V^T$

$\boxed{A^+ = V\Sigma^+U^T}$

General solution $Ax = b$

$\hat{x} = (A^TA)^{-1}A^Tb$

CN ↑.

$Ax^+ = b$

$x^+ = A^+b = V\Sigma^+U^Tb$

Works even if A has dependent columns ⇒ Normal equations fail because A^TA is not invertible.

→ Shortest least squares solution

$$A = \underset{m\times r}{U} \quad \underset{r\times r}{\Sigma} \quad \underset{r\times n}{V^T}$$

Bonus factorisation

$$V^T V = I$$

$$A = \left(\underset{m\times r}{U} \quad \underbrace{\underset{r\times n}{V^T} \Big) \Big(\underset{n\times r}{V}}_{} \quad \underset{r\times r}{\Sigma} \quad \underset{r\times n}{V^T} \right)$$

$$= \quad Q \qquad\qquad S$$

rotation $\qquad$ scaling

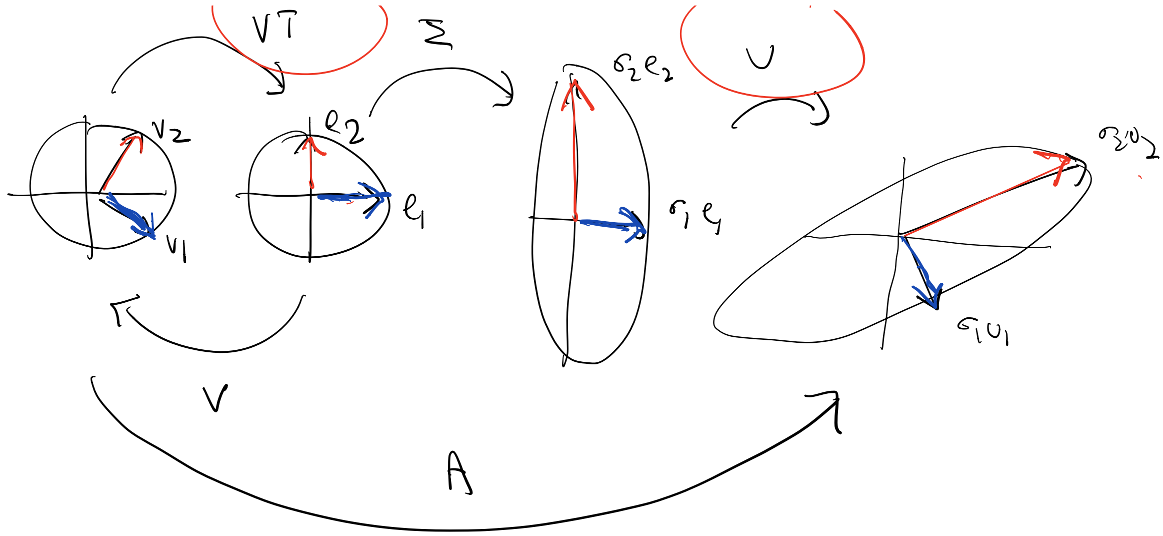
$A = LU$

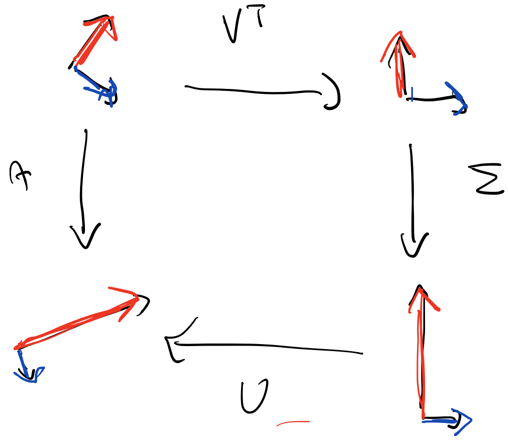
$A = QR$

$$Q^T Q = (UV^T)^T (UV^T) \qquad QQ^T = (UV^T)(UV^T)^T$$

$$= V U^T U V^T \qquad\qquad = UV^T V U^T$$

$$= I \qquad\qquad = I$$





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

$$A V = U \Sigma$$

PCA   let X be a matrix of features with mean = 0 (centered)

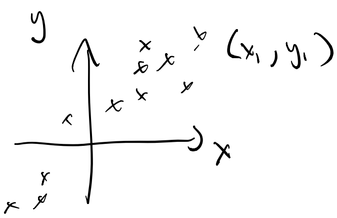
Then $\text{Cov}(X) = \underbrace{\frac{1}{n-1} XX^T}_{\text{symmetric positive definite}} = C$

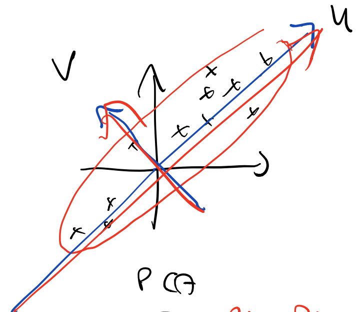
Note how SVD & PCA are related
SVD starts with X

$\cancel{X}$ 0

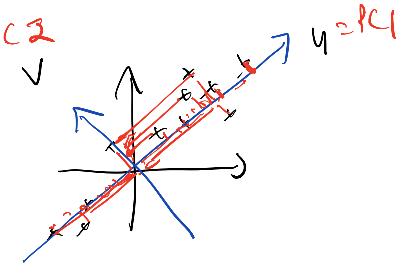
$$\boxed{C = Q\Lambda Q^T}$$

High "dependence" among feature vectors





PC2
v



x	y
x_1	y_1
x_2	y_2
x_3	y_3
⋮	
x_n	y_n

PCA →

PC1	PC2
u	v
u_1	v_1
u_2	v_2
u_3	v_3
⋮	
u_n	v_n

Drop v →

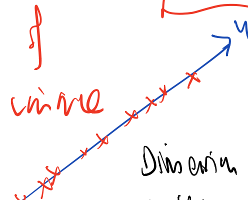
u
u_1
u_2
u_3
⋮
u_n

$$\frac{\lambda_1 \text{ the PCs}}{\sum \lambda_i} > 95\%$$

$= \lambda_i$ proportion of total variance

C □ X ...

$$\begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_n \end{bmatrix}$$



Dimension reduction

SVD & PCA

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

$$\frac{1}{n-1} XX^T = \boxed{U\Sigma V^T V \Sigma^T U^T = \boxed{U} \frac{\Sigma\Sigma^T}{n-1} U^T}$$

Trace

Explained variance = 95%

Principal component eigenvalues are given by $\frac{\Sigma^2}{n-1}$ $\boxed{\lambda_i = \frac{\sigma_i^2}{n-1}}$

Principal component vectors are given by U

Summary (Highlights of undergraduate linear algebra in 3 lectures)

① Vectors & vector spaces, norms

② matrix * vector → linear combination

$$\begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

③ Linear combinations

④ Linear independence

⑤ Rank, span, basis

⑥ Gaussian elimination → Fundamental subspaces

→ LU

LDU

LDL^T

CC^T (Cholesky)

$LD^{1/2} \; D^{1/2} L^T$

⑦ Orthogonalization

inner product

projection

Gram schmidt → QR

$\{u_1, u_2, \dots, u_n\}$

$\{v_1, v_2, \dots, v_n\}$

⑧ Change of basis → $S = Q \Lambda Q^T$

eigenvector basis

eigenvalues

(9) SVD $\longrightarrow$ $A = U \Sigma V^T$

$\sum \sigma_i u_i v_i^T$ sum of rank 1 matrices

PCA

Pseudoinverse

e.g. solve least squares problem even if A^TA singular

QS decomposition

(10) Now for some Python examples.