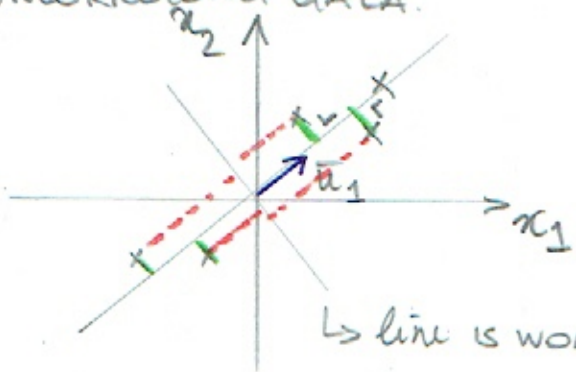
PCA: Principal Component Analysis

↳ describe lots of data with a smaller set of <u>relevant</u> features

→ transforms sets of correlated data to a smaller number of uncorrelated data



x_1, x_2 are <u>highly correlated</u>

↓

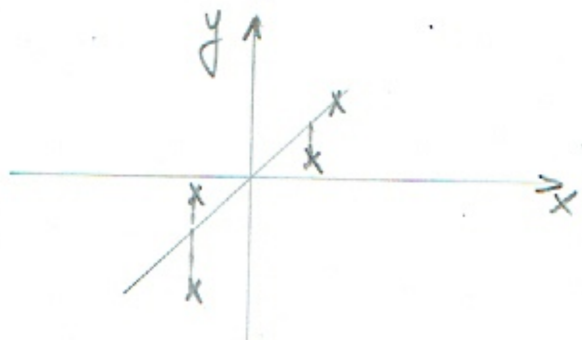
find $\bar{u}_1$ on which to project the data so that projection error is minimal

↓

goal is to keep variance as high as possible

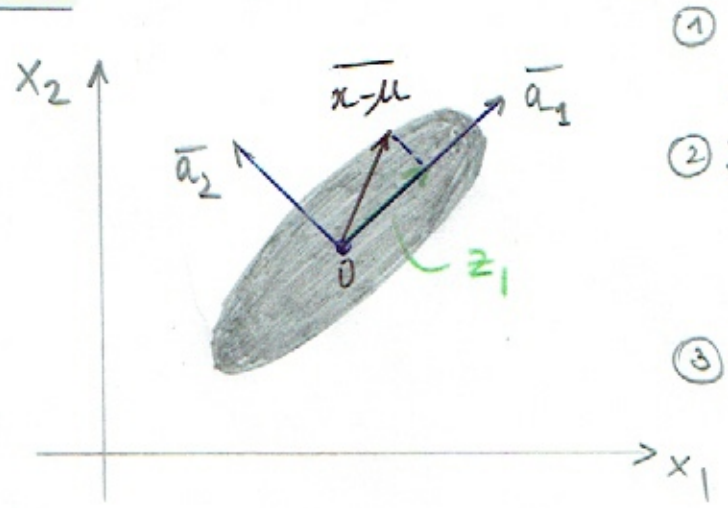
↳ line is worse → loss of variance

PCA ≠ linear Regression MSE



Advice ① only use when needed
② do feature normalization ! → $\frac{X-\mu}{\sigma}$
↳ zero mean

<u>Algorithm</u>



① select $\bar{a}_1$ and $\bar{a}_2$ in the directions of max variance

② $\bar{a}_1$ and $\bar{a}_2$ should be independent

⇒ $\text{cov}(a_1, a_2) = 0$

③ $\bar{a}_1^T \cdot \bar{a}_1 = 1$ per def

→ $\boxed{\bar{a}_1^T \cdot \bar{a}_1 - 1 = 0}$ (1)

Step 1: Mean normalization

$$z_1 = \bar{a}_1^T \cdot (\overline{x-\mu}) \quad : \text{proj of } (\overline{x-\mu}) \text{ onto } \bar{a}_1$$

$z_1 \downarrow$ SCALAR.

Find $\bar{a}_1$ so that $VAR(z_1)$ is MAXIMIZED

$$VAR(z_1) = var\left(\bar{a}_1^T \cdot (\overline{x-\mu})\right) \overset{cov(Ax) = A\, cov\, A^T}{=} \bar{a}_1^T cov(x)\, \bar{a}_1$$

$\bar{a}_1$: fixed unit vector

$(\overline{x-\mu})$: column vector of Random Variables.

$cov(x-\mu) = cov(x)$

$$cov(Ax) = A \Sigma A^T$$

Σ ↳ cov. matrix.

$$cov(Ax) = E\left[(Ax - E(Ax))\,(Ax - E[Ax])^T\right]$$

$$= E\left[(Ax - AE(x))\,(Ax - AE(x))^T\right]$$

$$= E\left[A(x - E(x))\,(A(x - E(x)))^T\right]$$

$$= A \underbrace{E(x - E(x))\,(x - E(x))^T}_{VAR(x) = cov(x) = \Sigma} A^T$$

CONSTRAINT: $\bar{a}_1^T \bar{a}_1 - 1 = 0$

$\xrightarrow{\text{LAGRANGE}}$ $L(\bar{a}_1, \lambda) = \overbrace{\bar{a}_1^T cov(x)\, \bar{a}_1}^{\|a_1\|^2} - \lambda\,(\overbrace{\bar{a}_1^T \cdot \bar{a}_1}^{\|a_1\|^2} - 1)$

$\longrightarrow$ $\nabla L(\bar{a}_1, \lambda) = 0$ ↑ SET

$$\nabla L = \cancel{2}\, cov(x)\, \bar{a}_1 - \cancel{2} \lambda a_1 = 0$$

$$\rightarrow \boxed{cov(x)\, \bar{a}_1 = \lambda\, \bar{a}_1}$$

$\bar{a}_1$ is an eigenvector of the COVARIANCE MATRIX!

NOW: $VAR(z_1) = \bar{a}_1^T cov(x)\bar{a}_1 = \bar{a}_1^T \lambda \bar{a}_1 = \lambda$

$\bar{a}_1$ is unit vector!

$\Rightarrow$ VARIANCE Along $\bar{a}_1$ is the eigenvalue itself.

$\rightarrow$ Eigenvector corresponding to the highest eigenvalue is the 1st Principal component!

SECOND Principal Component

CONSTRAINT 2: $\bar{a}_2^T \bar{a}_1 = \bar{a}_1^T \bar{a}_2 = 0 \qquad \bar{a}_1 \perp \bar{a}_2$

$$\nabla L(a_2, \alpha, \beta) = \bar{a}_2^T cov(x)\bar{a}_2 - \alpha(\bar{a}_2^T \cdot \bar{a}_2 - 1) - \beta(\bar{a}_2^T \bar{a}_1)$$

$\rightarrow \underset{\text{SET}}{\nabla L = 0} \rightarrow 2\,cov(x)\bar{a}_2 - 2\alpha\bar{a}_2 - \beta\bar{a}_1 = 0 \quad (**)$

$\rightarrow \nabla L\, \bar{a}_1^T = 2\bar{a}_1^T cov(x)\bar{a}_2 - \underbrace{2\alpha\bar{a}_2\bar{a}_1^T}_{=0} - \beta\bar{a}_1^T\bar{a}_1 = 0$

↓ swap $\bar{a}_1$ & $\bar{a}_2$

$\rightarrow \nabla L \cdot \bar{a}_1^T = 2\bar{a}_2^T \underbrace{cov(x)\bar{a}_1}_{\lambda\bar{a}_1} - \beta\underbrace{\bar{a}_1^T\bar{a}_1}_{\|1\|} = 0$

$\rightarrow 2\underbrace{\bar{a}_2^T\bar{a}_1}_{=0}\lambda - \beta = 0 \Rightarrow \boxed{\underset{\text{MUST}}{\beta = 0}}$

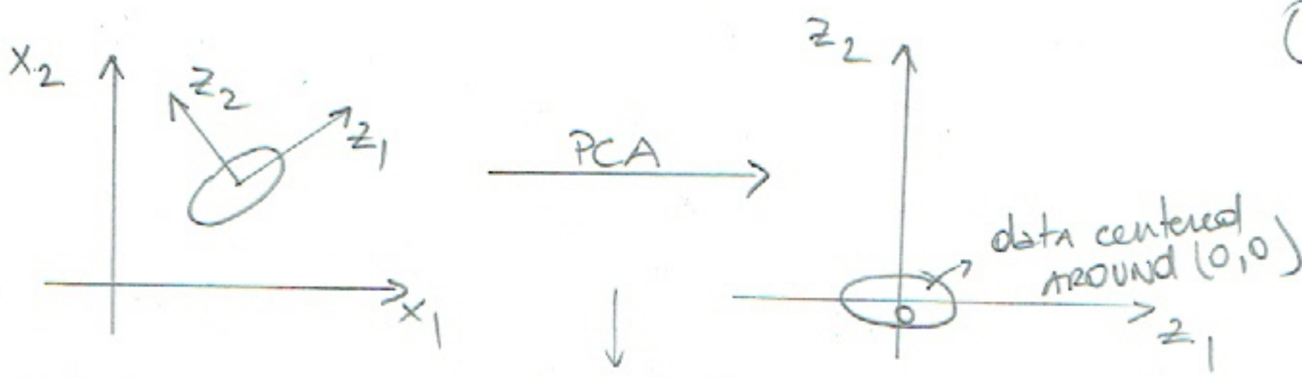
$(**) \rightarrow \nabla L = \cancel{2}cov(x)\bar{a}_2 - \cancel{2}\alpha\bar{a}_2 = 0$

$\rightarrow cov(x)\bar{a}_2 = \boxed{\alpha}\bar{a}_2$

↳ MUST be second highest eigenvalue after λ

Conclusion: WE will find d eigenvectors from the COV MATRIX Σ

$\rightarrow \Sigma$ is always symmetric AND eigenvectors ARE orthogonal to EACH other!



PCA centers the data and then projects the data onto axes along which the variance is maximum

→ Variance along z_2 could be ignored if negligible!

How much of variance is retained?

$$\mathrm{cov}(z) = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix} \qquad \lambda_1 > \lambda_2 > \lambda_3 \cdots > \lambda_d$$

keep k principal components → $\dfrac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_d}$

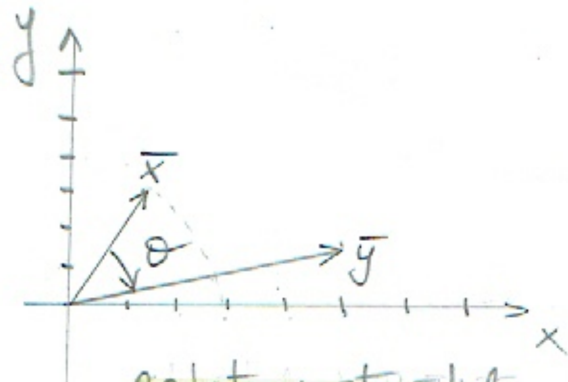
target is to retain 99% of the variance. Can also be 95%, 90%, 85%

Singular Value Decomposition

$\bar{x} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \quad A = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$

$\bar{y} = A\bar{x} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$

MATRIX WORKING ON VECTOR



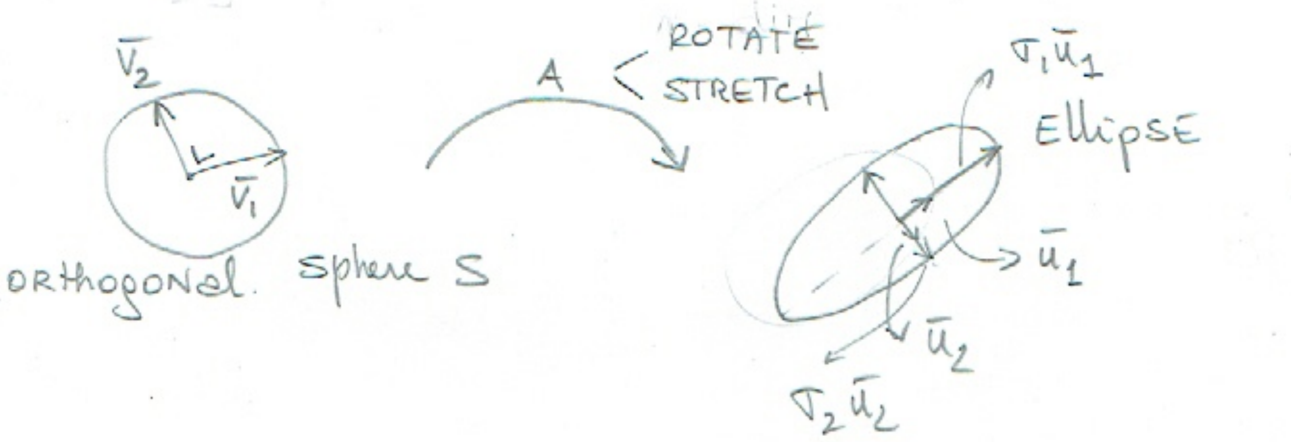
Rotate + stretch

MATRIX = linear transformation in space!

$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \rightarrow$ ROTATION by θ

$A = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \rightarrow \alpha =$ stretching factor

$\bar{y} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$



$\bar{v}_1, \bar{v}_2, \dots, \bar{v}_m \xrightarrow{A} \bar{u}_1, \bar{u}_2, \dots, \bar{u}_m \rightarrow$ orthonormal unit vectors

orthonormal.

$\underbrace{\sigma_1, \sigma_2, \dots, \sigma_m}_{\text{SINGULAR VALUES}} \rightarrow$ "stretching"

↳ MEASURE of VARIENCE IN DATA

$$A\bar{v}_1 = \sigma_1 u_1 \quad \longleftrightarrow \quad A\bar{v} = \lambda \bar{v}$$

(Eigenvalue Def)

$$\underbrace{A\bar{v}_j = \sigma_j \bar{u}_j \quad j = 1, \dots, n}$$

$$\underbrace{A}_{m\times n}\underbrace{\left[\bar{v}_1, \bar{v}_2, \dots, \bar{v}_n\right]}_{m\times n} = \underbrace{\left[u_1, u_2, \dots, \bar{u}_n\right]}_{\text{pick up ROTATION}} \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}}_{\text{stretching factors}}$$

$\sigma_1 \geq \sigma_2 \geq \dots \geq 0$

→ $\boxed{A \underbrace{V}_{\text{ROTATION}} = U \underbrace{\Sigma}_{\text{stretching}}}$ (1)

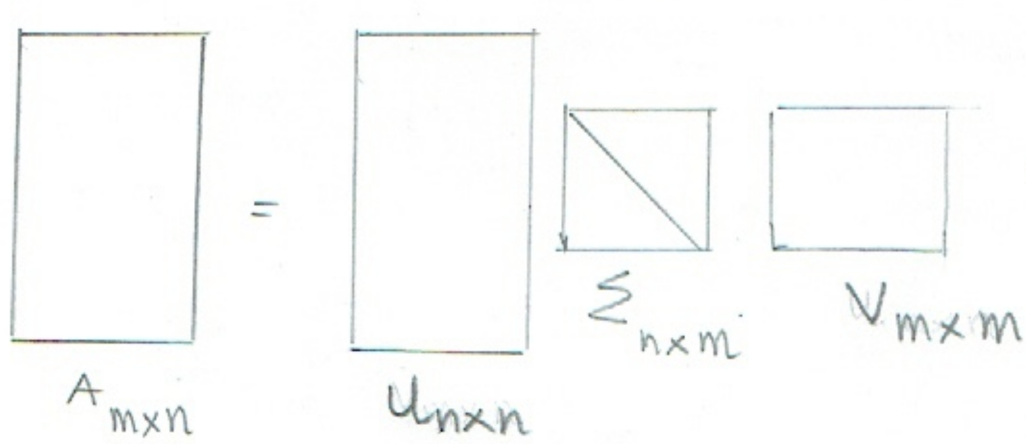
ROTATION

if orthogonal $\begin{cases} V^{-1} = V^T \\ U^{-1} = U^T \end{cases}$

(1) $A\underbrace{VV^{-1}}_{I} = U\Sigma V^{-1}$

→ $\boxed{A = \underbrace{U}_{\text{ROT}} \underbrace{\Sigma}_{\text{stretch}} \underbrace{V^T}_{\text{ROT}}}$ → SINGULAR VALUE DECOMPOSITION — always works!

= decomposition of complex transformation in 3 simpler transformation.



How to compute $U \Sigma V^T$

1. $A^T A = (U \Sigma V^T)^T (U \Sigma V^T)$

$$= V \Sigma^T \underbrace{U^T U}_{I} \Sigma V^T \rightarrow A^T A = V \Sigma^2 V^T$$

EIGENVECTORS

$$\underline{A^T A V} = V \Sigma^T \Sigma V^T V = \underline{V \Sigma^2} \rightarrow \text{FIND } V$$

Eigenvalue problem $\rightarrow \lambda_j = \sigma_j^2$

$\rightarrow$ Find Eigenvalues

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

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

Eigenvector

$$A A^T U = U \Sigma^2 \underbrace{U^T U}_{I} = A A^T U = U \Sigma^2 \rightarrow \text{FIND } U$$

$\hookrightarrow$ Eigenvalue problem

Dimensionality Reduction $\rightarrow$ keep σ_j^2s that ARE significant!

$$\frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2}{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_m^2} \rightarrow \text{\% of variance retained}$$