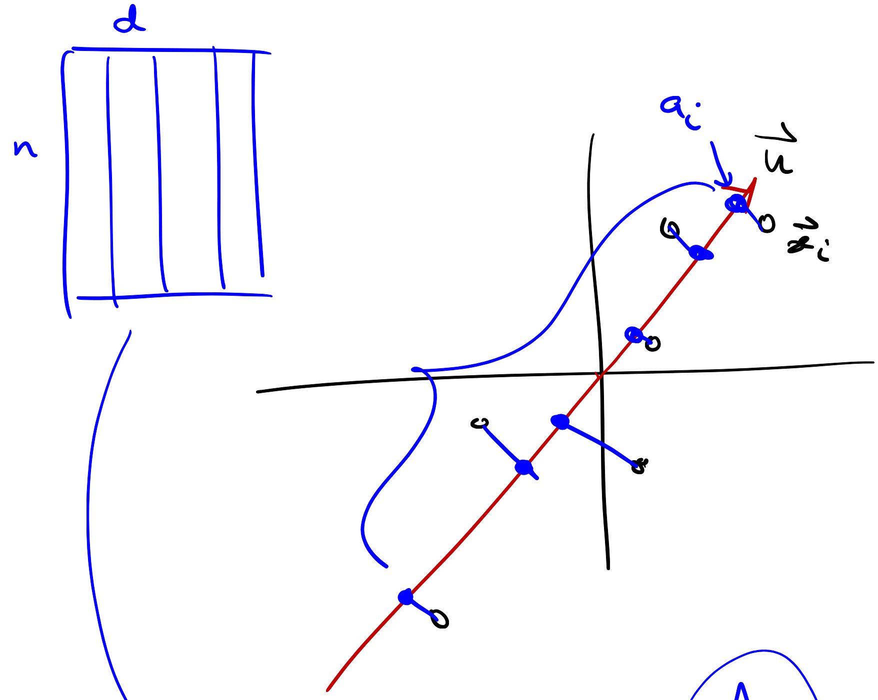


Center the data

$\bar{D}$

$$a_i = \underbrace{\text{proj}_{\vec{u}}(\vec{x}_i)}_{\text{scalar}} = \vec{x}_i^T \vec{u}$$

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \rightarrow \sigma_u^2 = \vec{u}^T \Sigma \vec{u}$$

cov matrix for D

$$J = \max_{\vec{u}} \vec{u}^T \Sigma \vec{u}, \quad \text{s.t.} \quad \underbrace{\vec{u}^T \vec{u} = 1}$$

$$J = \max_{\vec{u}} \vec{u}^T \Sigma \vec{u} - \alpha \underbrace{(\vec{u}^T \vec{u} - 1)}$$

$$\frac{\partial J}{\partial \vec{u}} = 2 \Sigma \vec{u} - 2 \alpha \vec{u} = 0$$

$$\Rightarrow \boxed{\Sigma \vec{u} = \alpha \vec{u}}$$

eigen-equation

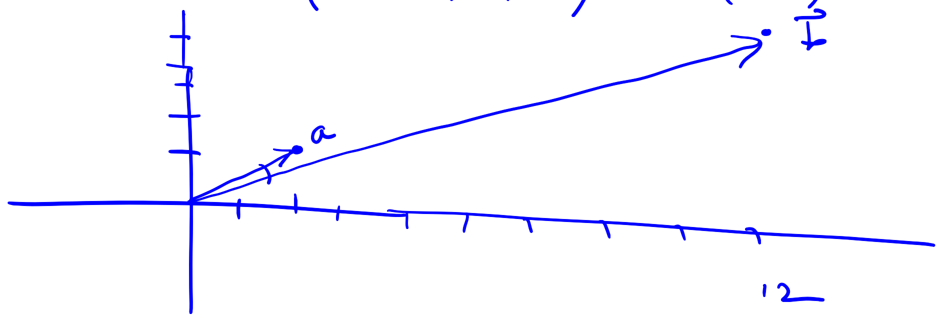
Σ: $d \times d$ $\quad$ $\vec{u}$: $d \times 1$ $\quad$ α: scalar 1×1

The direction that captures the most variance has to be an eigenvector $\vec{u}$

(α is called the eigenvalue)

$$\underset{d\times d}{\Sigma} = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \qquad \underset{d\times 1}{\vec{a}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$\Sigma \cdot \vec{a} = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 12 \\ 5 \end{pmatrix} = \underset{d\times 1}{\vec{b}}$$



$$D = \begin{array}{c} \\ n \end{array}\overset{d}{\begin{bmatrix} \vec{x}_i : x_{i1}\ x_{i2} \cdots x_{id} \\ \\ \end{bmatrix}} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \vdots \\ \sum_{j=1}^{d} x_{ij} \\ \vdots \end{pmatrix}$$

$n \times d \qquad d \times 1 \qquad n \times 1$

$$\boxed{\Sigma \vec{u} = \alpha \vec{u}}$$

$\vec{u}$ is an eigen-vector

$d \times d \leftarrow$ square

symmetric

Σ is a PSD : positive semi-definite matrix

$$\underbrace{\vec{x}^T \Sigma \vec{x}}_{\text{Scalar}} \geq 0$$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_d^2 \end{pmatrix} = \{\sigma_{ij}\}_{i,j=1\ldots d}$$

$$\sum_i \sum_j x_i \sigma_{ij} x_j \geq 0$$

Σ has (up-to) d non-negative eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_d \geq 0$$

in decreasing order

$$\vec{u}_1 \quad \vec{u}_2 \quad \vec{u}_3 \quad \dots \quad \vec{u}_d$$

$\leftarrow$ eigenvectors (unit vectors)

$\vec{u}_i$ is orthogonal to $\vec{u}_j$ $\forall i, j$

$$\vec{u}_i^T \vec{u}_j = 0 \qquad i \neq j$$

$$\vec{u}_i^T u_i = 1$$

$$\sigma_u^2 = \vec{u}^T \Sigma \vec{u}$$

$$\Sigma \vec{u}_1 = \lambda_1 \vec{u}_1$$

Projected variance along $\vec{u}_1$

$$\sigma_{u_1}^2 = \vec{u}_1^T \underbrace{\Sigma \vec{u}_1}$$

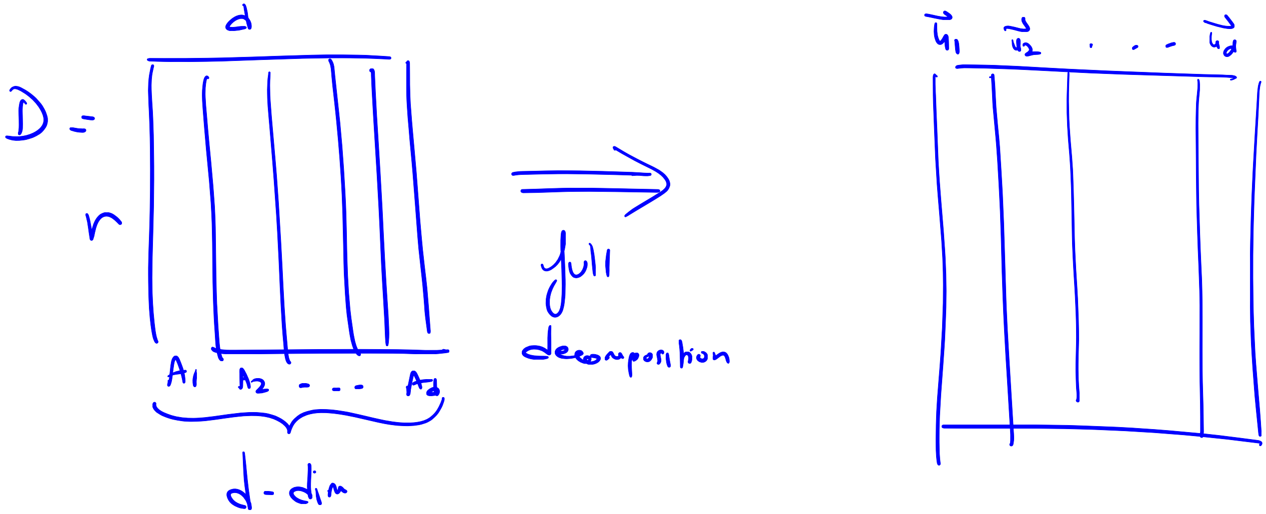
$$= \vec{u}_1^T (\lambda_1 u_1)$$

$$= \lambda_1 \vec{u}_1^T \vec{u}_1$$

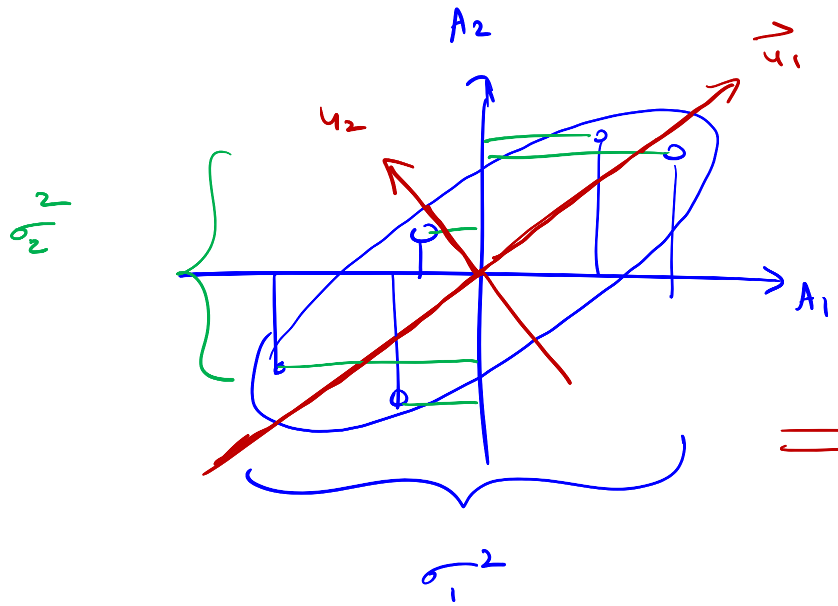
$$= \lambda_1$$

$$\sigma_{u_i}^2 = \lambda_i$$

$\vec{u}_1$ is the direction that maximizes the $\underline{\underline{\text{variance}}}$! λ_1

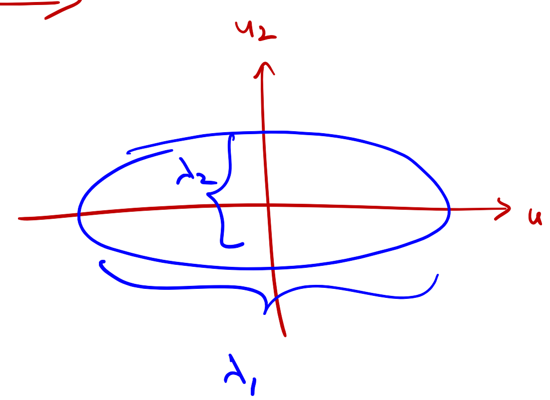


$$(\vec{A}_1\ \vec{A}_2 \dots \vec{A}_d) \xrightarrow{\text{Change of basis}} (\vec{u}_1\ \vec{u}_2 \dots \vec{u}_d)$$

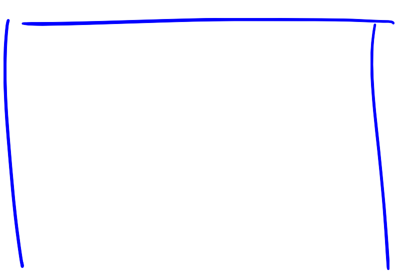


$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

$\Longrightarrow$

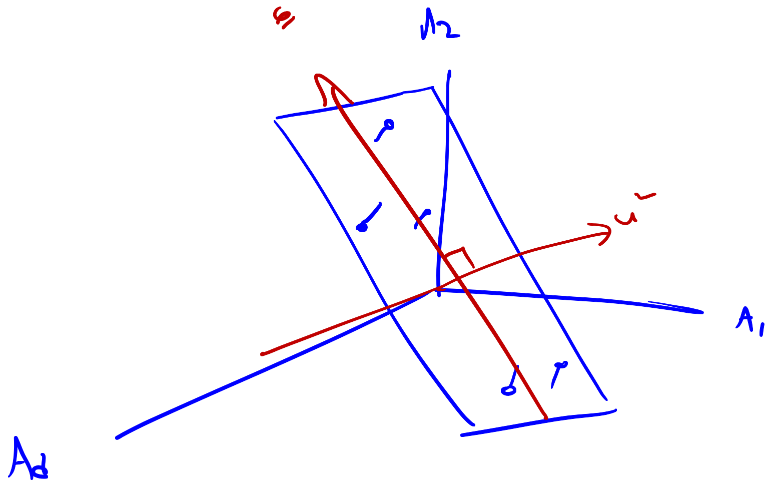


$d = 100$



$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \checkmark$$

$d = 100$



$$\underset{100}{\Sigma} = \underset{100}{\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1d} & \cdots \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_d^2 \end{pmatrix}}$$

$\underline{\underline{\lambda_1 + \lambda_2}}$ capture 99% of the total variance

$d = 2$

$\vec{u}_1$	$\vec{u}_2$

$$\Sigma \vec{u}_1 = \lambda_1 \vec{u}_1$$

known ? ?

λ_1 largest eigenvalue

corresponding

dominant eigenvector $\vec{u}_1$

$$\vec{x}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

initial guess

$$\boxed{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$$

unit vector

$$\Sigma = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

$\vec{u}_1$

λ_1

$$\vec{x}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\Sigma \vec{x}_0 = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 3 \end{pmatrix} = \vec{x}_1$$

$\lambda = 7$

$\rightarrow$ normalize to unit vector

$\rightarrow$ Scale by dividing by 7

$$\Sigma \vec{x}_1 = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}\underbrace{\begin{pmatrix} 1 \\ 0.429 \end{pmatrix}}_{\vec{x}_1'} = \underbrace{\begin{pmatrix} 5.85 \\ 2.43 \end{pmatrix}}_{\vec{x}_2}$$

$\lambda = 5.85$

$$\Sigma \vec{x}_2 = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}\underbrace{\begin{pmatrix} 1 \\ 0.415 \end{pmatrix}}_{\vec{x}_2'} = \underbrace{\begin{pmatrix} 5.83 \\ 2.42 \end{pmatrix}}_{\vec{x}_3}$$

$\lambda = 5.83$

$\vdots$

few more iterations

$$\| \vec{x}_t - x_{t-1} \| = \left\| \begin{pmatrix} 0.02 \\ 0.01 \end{pmatrix} \right\| = \sqrt{0.02^2 + 0.01^2}$$

$$< \theta = \underline{0.001}$$

tolerance

power-iteration method

$$\Big(\Sigma \underbrace{\big(\Sigma \underbrace{(\Sigma \vec{x}_0)}_{\vec{x}_1}\big)}_{\vec{x}_2}\Big) \simeq \Sigma^t \vec{x}_0$$

$\sqrt{35.85}$

$= 6.312$

$$\lambda_1 = 5.83$$

$$\vec{u}_1 = \begin{pmatrix} 5.83 \\ 2.42 \end{pmatrix} \frac{1}{\sqrt{(5.83)^2 + (2.42^2)}} = \begin{pmatrix} 0.92 \\ 0.38 \end{pmatrix} = \vec{u}_1$$

$$\boxed{\Sigma \vec{u} = \lambda \vec{u}}$$

d equations

$$\left\{\begin{aligned} \Sigma \vec{u}_1 &= \lambda_1 \vec{u}_1 \\ \Sigma \vec{u}_2 &= \lambda_2 \vec{u}_2 \\ &\vdots \\ \Sigma \vec{u}_d &= \lambda_d \vec{u}_d \end{aligned}\right.$$

matrix equation

$$\longrightarrow \quad \Sigma U = U \Lambda$$

$$U = \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_d \\ | & | & & | \end{bmatrix}$$

$d \times d$

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{bmatrix}$$

orthogonal matrix

$$U^{-1} = U^T$$

pair-wise orthogonal columns

& unit length

$$\Sigma \underbrace{U \cancel{U^T}}_{I} = U \Lambda U^T$$

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

eigen decomposition equation

$$\Sigma = \sum_{i=1}^{d} \lambda_i \vec{u}_i \vec{u}_i^T$$

$$\Sigma = \boxed{\lambda_1 \underset{d \times 1}{\vec{u}_1} \underset{1 \times d}{\vec{u}_1^T}} + \underbrace{\lambda_2 \vec{u}_2 \vec{u}_2^T + \cdots + \lambda_d u_d u_d^T}$$

$$\boxed{} = \lambda_1 \boxed{u_1 u_1^T} + \lambda_2 \boxed{u_2 u_2^T} + \cdots$$

$$\Sigma - \lambda_1 u_1 u_1^T = \Sigma' =$$

dominant eigenvector is $\vec{u}_2$ (λ_2)

$\vec{x}_0 \longrightarrow$ random

$\vec{y}_0 \longrightarrow$ random

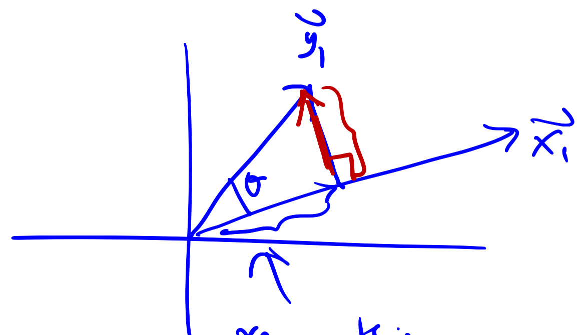
$$\Sigma \vec{x}_0 = \vec{x}_1$$

$$\downarrow$$

$$\vec{y}_1$$

$$\Sigma \vec{y}_0 = \vec{y}_1$$

Orthogonalize $\vec{y}_1$ w.r.t $\vec{x}_1$



remove this

$$\vec{y}_1' = \vec{y}_1 - \left(\frac{y_1^T x_1}{x_1^T x_1}\right) \cdot \vec{x}_1$$

$$\downarrow$$

$$\vec{u}_2$$

PCA: Input: D, frac of variance $\gamma = 95\% = 0.95$

1) $\bar{D} \leftarrow$ centered data

2) $\Sigma \leftarrow$ covariance matrix

3) Solve $\Sigma \vec{u} = \lambda \vec{u}$

$$\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_d$$
$$\vec{u_1} \quad \vec{u_2} \quad \quad \vec{u_d}$$

4) total variance of the data $= \sum \lambda_i$

5) find the smaller k such that

variance from 1^{st} k dim $\leftarrow$

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{d} \lambda_j} \geq \gamma$$

total

6) Create a new dataset of k dims

$\vec{x_i} \rightarrow$ project onto $\vec{u_1}, \vec{u_2}, \dots, \vec{u_k}$

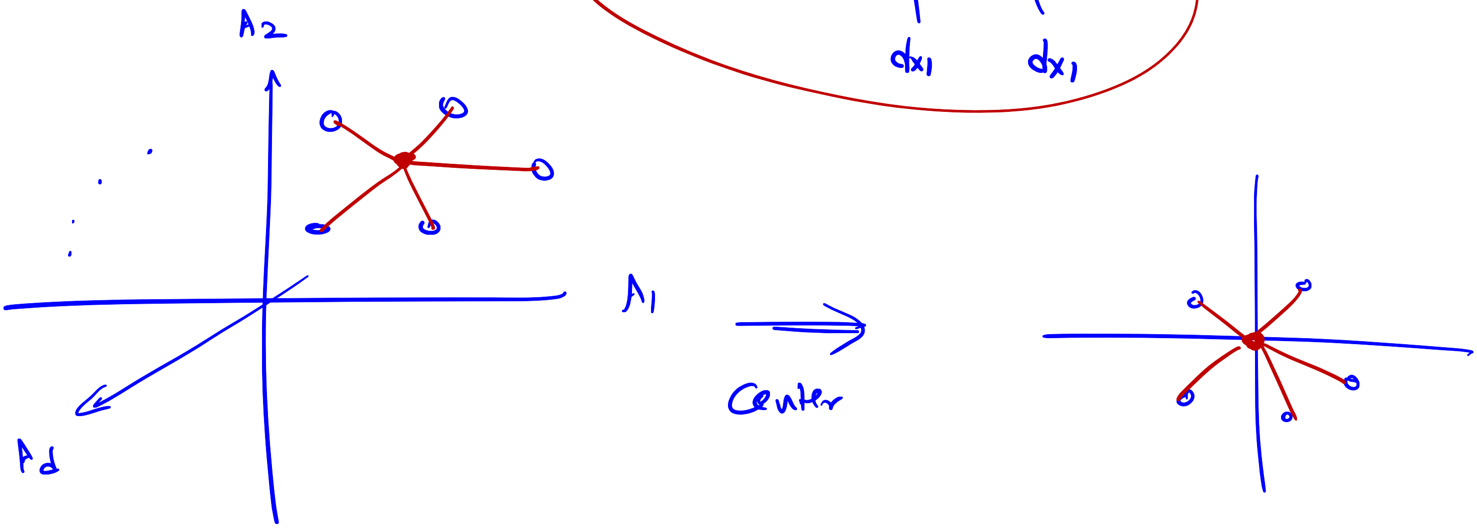
$$\text{proj}_{u_j}(\vec{x_i}) = \left(\frac{\vec{x_i}^T u_j}{\underbrace{u_j^T u_j}_{=1}} \right)$$

new point $\vec{x_i}' \rightarrow \underbrace{\left(\qquad \qquad \right)}$

projected onto k-dim space

$k \ll d$

total variance $= \frac{1}{n} \sum \| \underset{d\times 1}{\vec{x}_i} - \underset{d\times 1}{\hat{\mu}} \|^2$



$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \cdots \\ & \ddots & & \\ & & \cdots & \sigma_d^2 \end{pmatrix}$$

$$\boxed{\text{tot var} = \underbrace{\sum_{i=1}^{d} \sigma_i^2} = \underbrace{\sum_{i=1}^{d} \lambda_i}}$$