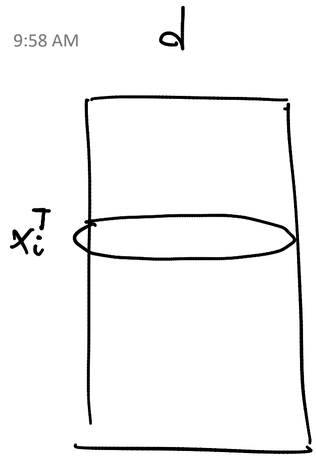
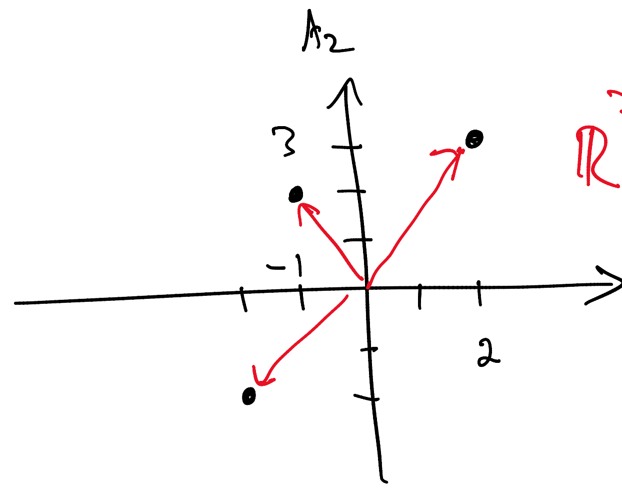
Lecture 3

Tuesday, September 5, 2023 9:58 AM

$D =$ 

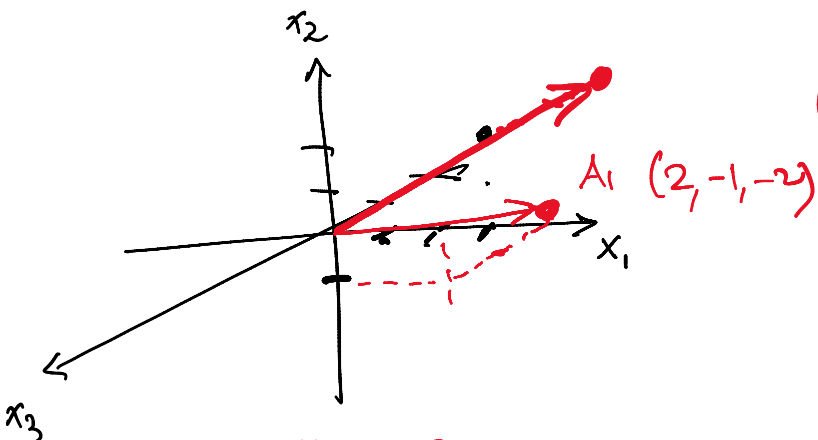
$d = 2$

$$\begin{array}{c|cc} & A_1 & A_2 \\ \hline x_i^T & 2 & 3 \\ & -1 & 2 \\ & -2 & -2 \end{array} \quad n = 3$$



$\mathbb{R}^2 \equiv \mathbb{R}^d$

Point / row view



$(A_1, A_2) \equiv$ basis vectors

$\mathbb{R}^n = \mathbb{R}^3$

$\left\{ \begin{array}{l} \text{Sample space} \\ \text{Col space} \end{array} \right.$

$\hat{\mu} \equiv$ Sample mean

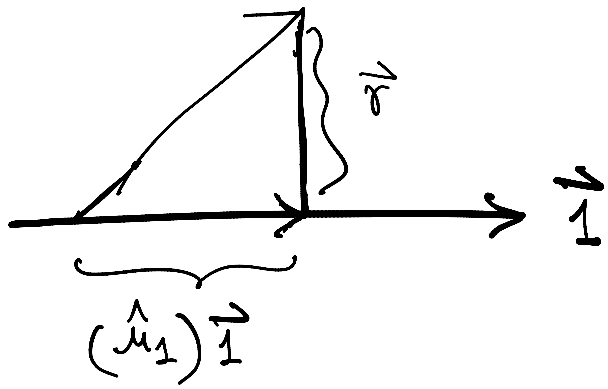
orthogonal projection along $\vec{1}$

(ones vector)

$\bar{D} \equiv$ Centered data $\equiv$ Orthogonalize wrt $\vec{1}$

$\hookrightarrow \quad x_i^T - \hat{\mu}^T$

A_1



$\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)^T$

$A_1 - (\hat{\mu}_1) 1$

Variance

$\hat{\sigma}_1^2 = \dfrac{\|\bar{A}_1\|^2}{n}$ squared norm / dim

$\hat{\sigma}_1^2 \propto \|\bar{A}_1\|^2$

$\hat{\sigma}_2^2 \propto \|\bar{A}_2\|^2$

$\hat{\sigma}_{12} = \dfrac{\vec{A}_1^T \vec{A}_2}{n}$ dot product between the centered attribute vectors

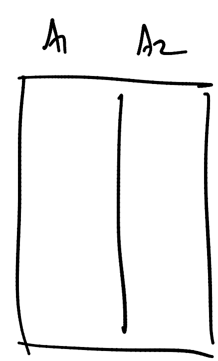
Cov

$\hat{\rho}_{12} = \left(\dfrac{\bar{A}_1}{\|\bar{A}_1\|}\right)^T \left(\dfrac{\bar{A}_2}{\|\bar{A}_2\|}\right) = \cos\theta$

Correlation

$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 \end{bmatrix}$

Cov matrix



$$= \begin{bmatrix} Var(A_1) & Cov(A_1 A_2) \\ Cov(A_1 A_2) & Var(A_2) \end{bmatrix} \qquad d=2$$

Square matrix 2×2, symmetric

$$\hat{\Sigma}^T = \Sigma$$

PSD: positive semi-definite matrix

any vector $\vec{a} \in \mathbb{R}^d$

$$\vec{a}^T \Sigma \vec{a} \geq 0$$

quadratic form

$$\vec{a}^T \vec{a} = ||\vec{a}||^2$$

$$\vec{a}^T I \vec{a}$$

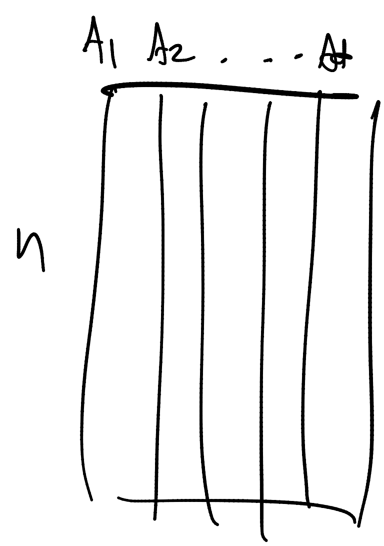
Consequences:

$\hat{\Sigma}$ has d non-negative eigenvalues

$\hat{\Sigma}$ in the full d-dim case

$$= \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \hat{\sigma}_{13} & \cdots & \hat{\sigma}_{1d} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 & \hat{\sigma}_{23} & \cdots & \hat{\sigma}_{2d} \\ \vdots & & \ddots & & \\ \hat{\sigma}_{d1} & \hat{\sigma}_{d2} & \hat{\sigma}_{d3} & \cdots & \hat{\sigma}_d^2 \end{bmatrix}$$

$d \times d$, symmetric, PSD matrix



d non-negative <u>eigenvalues</u>

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

in decreasing order

$$\underbrace{\vec{u}_1 \quad \vec{u}_2 \quad \vec{u}_3 \quad \cdots \quad \vec{u}_d}_{\text{Corresponding eigenvectors}}$$

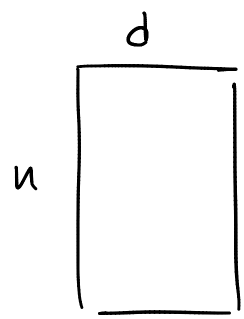
dominant eigenvector

$u_i^T u_j = 0 \quad \forall\, i,j = 1, \dots, d$

$\vec{u}_i$ is orthogonal to $\vec{u}_j$

$u_i^T u_i = 1 \;, \quad \|\vec{u}_i\| = 1$

unit vectors

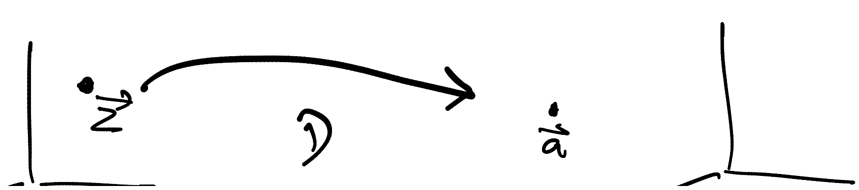
$D =$ 

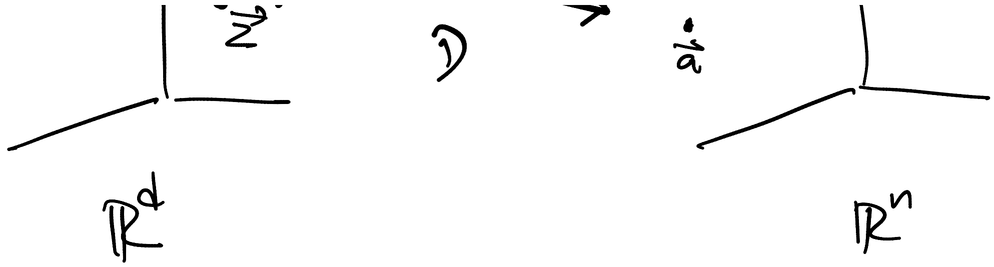
$\vec{z} \in \mathbb{R}^d$

$$\underset{n\times 1}{\vec{a}} = \underset{(n\times d)}{D}\,\underset{(d\times 1)}{\vec{z}}$$

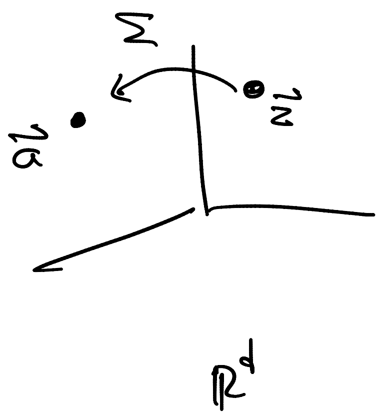
D acts like a function

$\mathbb{R}^d \Rightarrow \mathbb{R}^n$





$$\Sigma \in \mathbb{R}^{d \times d} : \mathbb{R}^d \to \mathbb{R}^d$$



$$\vec a = \Sigma \vec z$$

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

$$\vec z = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\vec a = \Sigma \vec z = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 3 \end{pmatrix}$$

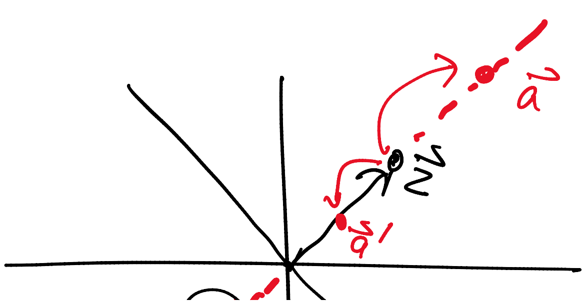
Σ as a transformation

Q: $\exists \underline{\vec u}$, such that $\boxed{\lambda \vec u = \Sigma \vec u}$

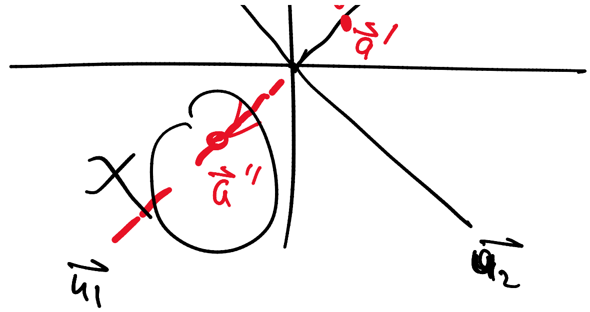
unit vector

eigen value

$\underbrace{\text{eigenvector}}_{\text{invariant direction}}$



for Σ symmetric



for Σ symmetric

$\Sigma \in \mathbb{R}^{2\times 2}$

An eigenvector is a non-zero unit vector $\vec{u} \in \mathbb{R}^d$ that satisfies the equation $\boxed{\Sigma \vec{u} = \lambda \vec{u}}$

eigen value $\lambda \in \mathbb{R}$ (scalar)

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

Power-Iteration to find the dominant eigenvector / eigenvalue

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

$\hat{\lambda}_1 = 7$

$$\vec{x}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \longrightarrow \Sigma \vec{x}_0 = \begin{pmatrix} 7 \\ 3 \end{pmatrix} = \vec{x}_1'$$

$\hat{\lambda}_1 = 5.86$

$$\vec{x}_1' \xrightarrow[\text{by } 1/7]{\text{scale}} \begin{pmatrix} 1 \\ 3/7 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 \\ 0.43 \end{pmatrix}}_{\vec{x}_1} = \begin{pmatrix} 5.86 \\ 2.43 \end{pmatrix} = \vec{x}_2' \xrightarrow[\text{scale}]{} \begin{pmatrix} 1 \\ 2.43/5.86 \end{pmatrix} = \begin{pmatrix} 1 \\ 0.42 \end{pmatrix}$$

$\vec{x}_1$

$\lambda_1 = 5.84$

$$\vec{x}_2 = \begin{pmatrix} 1 \\ 0.42 \end{pmatrix} \longrightarrow \Sigma \vec{x}_2 = \begin{pmatrix} 5.84 \\ 2.42 \end{pmatrix} \xrightarrow[\text{scale}]{} \begin{pmatrix} 1 \\ 2.42/5.84 \end{pmatrix} = \begin{pmatrix} 1 \\ 0.42 \end{pmatrix}$$

$$\vec{u}_1 = \begin{pmatrix} 1 \\ 0.42 \end{pmatrix} \xrightarrow[\text{unit vector}]{} \frac{1}{\sqrt{1^2 + 0.42^2}} \begin{pmatrix} 1 \\ 0.42 \end{pmatrix} = \frac{1}{\sqrt{1.18}} \begin{pmatrix} 1 \\ 0.42 \end{pmatrix}$$

$$= \frac{1}{1.07} \begin{pmatrix} 1 \\ 0.42 \end{pmatrix}$$

$$\boxed{\text{dominant eigenvector} \longrightarrow \vec{u}_1 = \begin{pmatrix} 0.93 \\ 0.39 \end{pmatrix} \quad \lambda_1 = 5.84}$$

do one more matrix multiplication and find λ

$$\Sigma u_1 = (\lambda_1) u_1$$

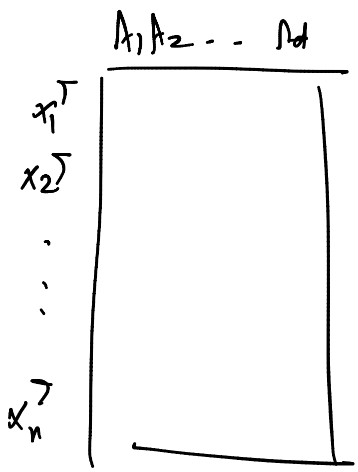
Dominant $\rightarrow$ Power Iteration $\rightarrow$ λ_1, $\vec{u}_1$

Q: how to find λ_2, $\vec{u}_2$, and the rest?

"Approximation"

$$D = \mathbb{R}^{n \times d}$$

$A_1 A_2 \dots A_d$



Q: what single point best approximates D?

$\vec{z}$ should be close to all other points

$\vec{z} \in \mathbb{R}^d$

$$\text{square dist}(\vec{z}, \vec{x}_i) = \|\vec{z} - \vec{x}_i\|^2$$

$\vec{x}_i \in D$

$\vec{z}$?

$$\boxed{\min_{\vec{z}} J \equiv \sum_{i=1}^{n} \|\vec{z} - \vec{x}_i\|^2}$$

$$\frac{\partial J}{\partial \vec{z}} = \vec{0}$$

partial derivative, set to zero, solve (if possible)

$$\sum \|\vec{z} - \vec{x}_i\|^2 = \sum_{i=1}^{n} (\vec{z} - \vec{x}_i)^T (\vec{z} - \vec{x}_i)$$

$$= \sum_{i=1}^{n} \left(\vec{z}^T z + x_i^T x_i - 2 z^T x_i \right)$$

$$\frac{\partial J}{\partial \vec{z}} = \sum_{i=1}^{n} 2\vec{z} - 2\vec{x}_i = \vec{0}$$

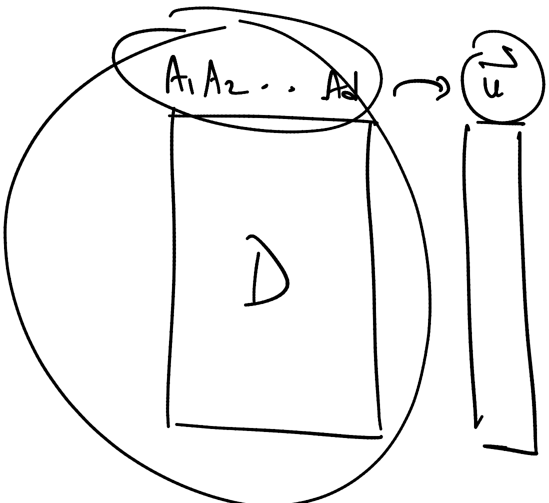
$$= \underbrace{\sum_{i=1}^{n} \vec{z}} = \sum_{i=1}^{n} \vec{x_i}$$

$$\Rightarrow \quad n.\vec{z} = \sum_{i=1}^{n} \vec{x_i}$$

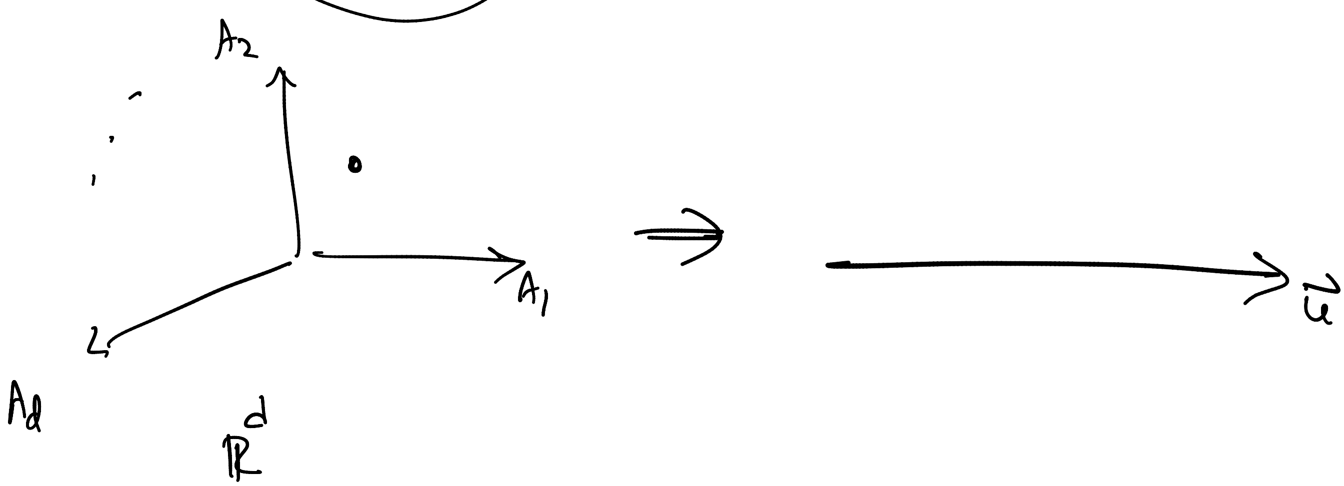
$$\Rightarrow \quad \vec{z} = \frac{1}{n} \sum_{i=1}^{n} \vec{x_i} = \vec{\mu}$$

mean value!

Q. what is the best 1d approximation to D



$\leftarrow$ find a new axis!



Project all points onto the $\vec{u}$ direction $\vec{u}$ should be a

Project all points onto the $\vec{u}$ direction, $\vec{u}$ should be a unit vector

$x_i \in \mathbb{R}^d$ $\quad\quad$ $\vec{u} \in \mathbb{R}^d$

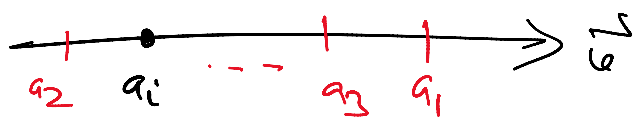
$\vec{u}^T\vec{u} = 1$

$\|\vec{u}\| = 1$

$$\underbrace{\left(\frac{\vec{x}_i^T\vec{u}}{\vec{u}^T\vec{u}}\right)}_{\text{scalar projection}}\vec{u}$$

$$a_i = \vec{x}_i^T\vec{u}$$

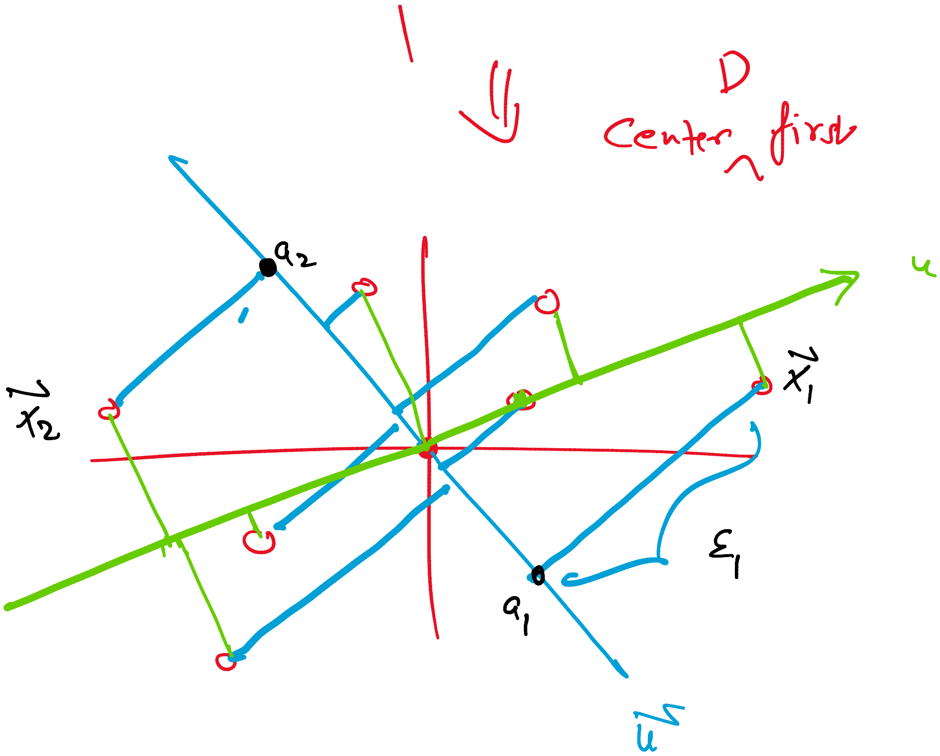
projected scalar value along $\vec{u}$

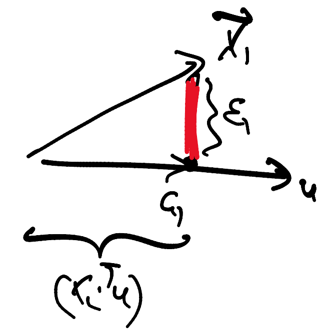


I Objective: maximize the projected variance

$$\frac{1}{n}\sum_{i=1}^{n}\left(a_i - \mu_a\right)^2$$

II: Minimize error (distance squared)





$$\min_{\vec{u}} \underbrace{\frac{1}{n}\sum_{i=1}^{n} \|\varepsilon_i\|^2}_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n} \left\| \vec{x}_i - (x_i^T \vec{u})\,\vec{u} \right\|^2$$

mean squared error

maximization of projected variance $\equiv$ minimization of MSE

$\vec{u}$ is infact the dominant eigenvector of Σ

Cov matrix!

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

λ_1 $\leftarrow$ variance along the new axis

eigenvalue