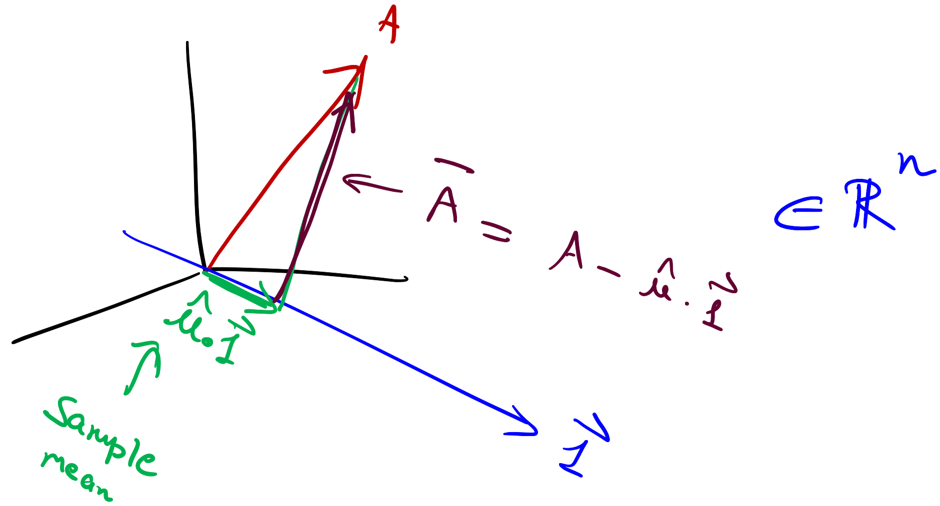
$$A = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$



True Unknown

Variance σ^2 $= E[(A-\mu)^2]$

mean $g(x) = x$

variance $g(x) = (x-\mu)^2$

$$E[g(x)] \to \int_x g(x)\, f(x)\, dx$$

$$E[g(x)] \to \sum_x g(x)\, P(A=x)$$

sqrt $\equiv$ standard deviation (σ)

Sample variance :

$$\boxed{\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2}$$

(biased estimate)

$$E[\hat{\sigma}^2] \neq \sigma^2$$

Unbiased $\hat{\sigma}_u^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})^2$ $\qquad E[\hat{\sigma}_u^2] = \sigma^2$

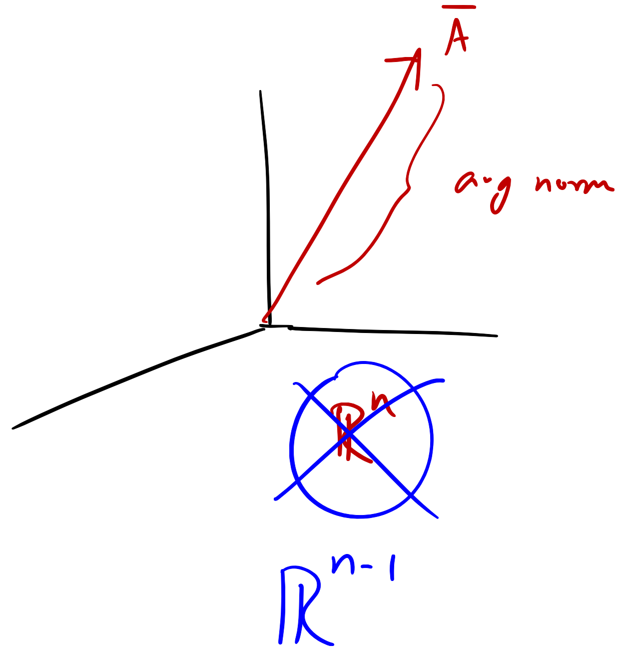
? why !!! degrees of freedom

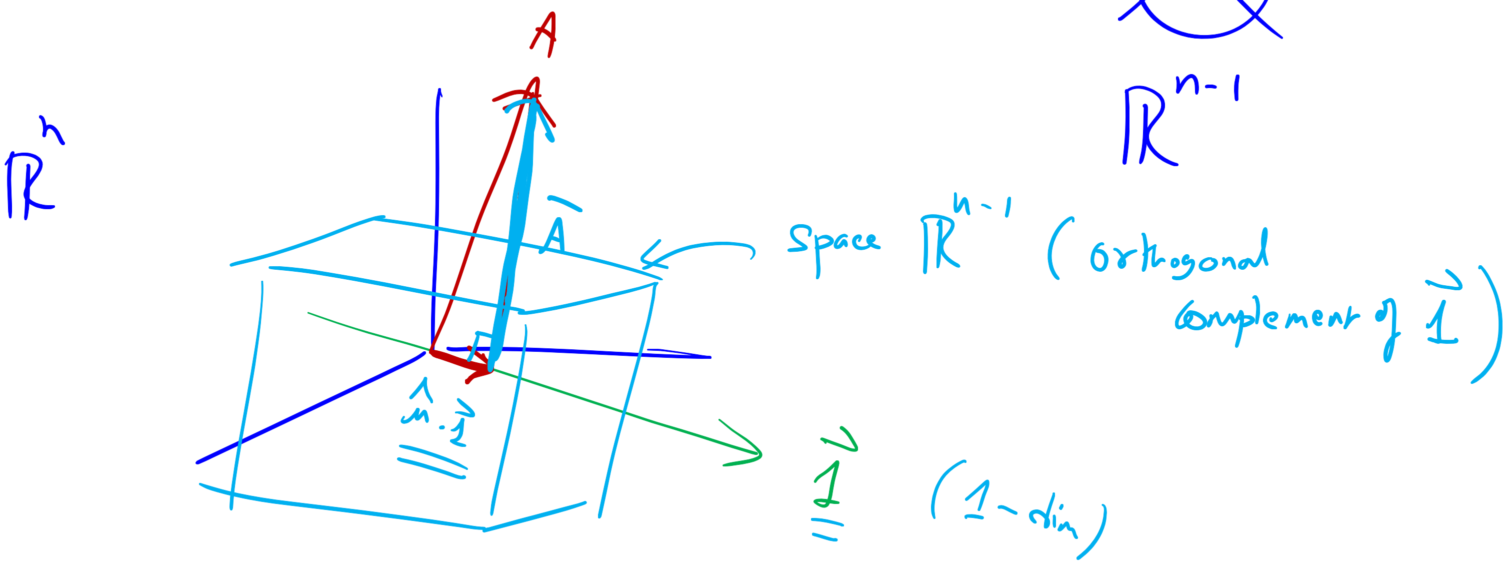
$$\bar{A} = \begin{pmatrix} x_1 - \hat{\mu} \\ x_2 - \hat{\mu} \\ \vdots \\ x_n - \hat{\mu} \end{pmatrix}$$

$$\boxed{\hat{\sigma}^2 = \frac{1}{n} \bar{A}^T \bar{A}} = \frac{\bar{A} \bullet \bar{A}}{n}$$

$$= \frac{1}{n} \|\bar{A}\|^2$$

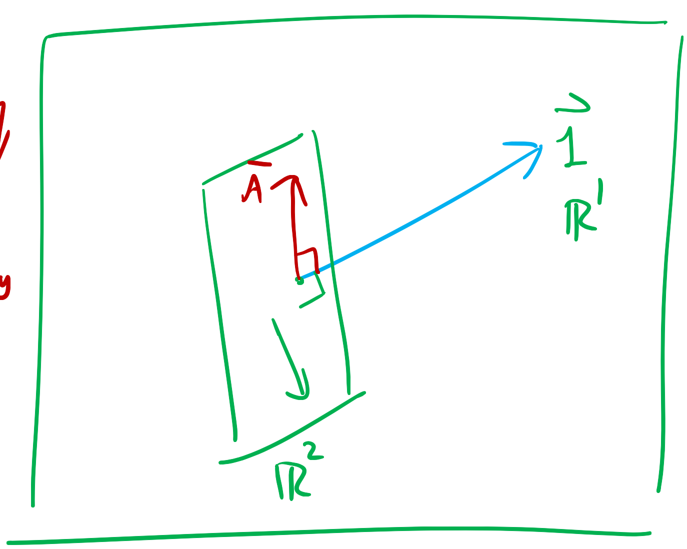
norm squared per dim





dof = degrees of freedom

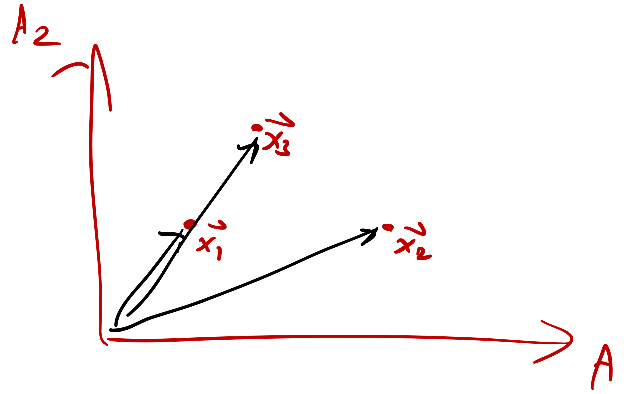
$\equiv$ dimensionality of the subspace where the r.v. lies



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

$$n \quad \begin{array}{c|c|c|} & A_1 & A_2 \\ \hline \vec{x}_1 & x_{11} & x_{12} \\ \vec{x}_2 & x_{21} & x_{22} \\ \vdots & & \\ \vec{x}_n & & \\ \hline \end{array}$$

$d = 2$



$$\mathbf{A} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \quad \text{bivariate r.v.}$$

$$E[\mathbf{A}] = \begin{pmatrix} E[A_1] \\ E[A_2] \end{pmatrix}$$

Covariance matrix

2×2 square

symmetric ←

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

Population

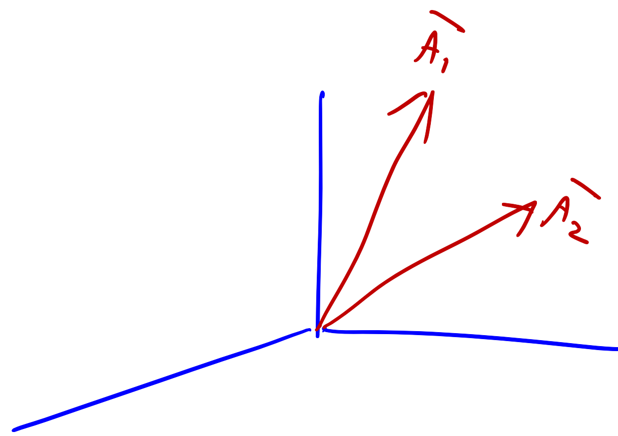
$$\vec{\hat{\mu}} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \qquad \underbrace{\hat{\sigma}_1^2 \ , \ \hat{\sigma}_2^2}_{\text{Variance along each attribute}}$$

Sample Covariance : $\hat{\sigma}_{12}$

$$\hat{\sigma}_{11} = \hat{\sigma}_1^2$$

$$\hat{\sigma}_{22} = \hat{\sigma}_2^2$$

$$\hat{\sigma}_{12} = \frac{1}{n} \sum_{i=1}^{n} (x_{i1} - \hat{\mu}_1)(x_{i2} - \hat{\mu}_2)$$



$\mathbb{R}^n$

$$\hat{\sigma}_{12} = \frac{1}{n} \bar{A}_1^T \bar{A}_2$$

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

$\rho_{12} \equiv$ Correlation between A_1 & A_2

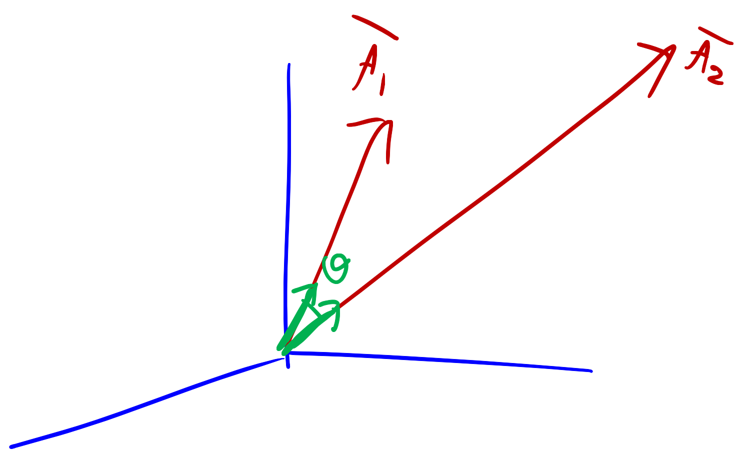
$$= \frac{\sum_{i=1}^{n} (x_{i1} - \hat{\mu}_1)(x_{i2} - \hat{\mu}_2)}{\sqrt{\sum_{i=1}^{n} (x_{i1} - \hat{\mu}_1)^2} \sqrt{\sum_{i=1}^{n} (x_{i2} - \hat{\mu}_2)^2}}$$

$\boxed{\text{row-view}}$

$$= \frac{\sigma_{12}}{\sqrt{\sigma_1^2}\sqrt{\sigma_2^2}} = \frac{\sigma_{12}}{\sigma_1 \cdot \sigma_2} \in [-1, 1]$$

Column-view

$$\rho_{12} = \frac{\cancel{\frac{1}{n}} \bar{A}_1^T \bar{A}_2}{\underbrace{\left(\cancel{\frac{1}{\sqrt{n}}} \|\bar{A}_1\|\right)}_{\sqrt{\sigma_1^2}} \left(\cancel{\frac{1}{\sqrt{n}}} \|\bar{A}_2\|\right)} = \underbrace{\left(\frac{\bar{A}_1}{\|\bar{A}_1\|}\right)^T \left(\frac{\bar{A}_2}{\|\bar{A}_2\|}\right)} = \cos\theta$$



d- dimensions

$$D = \begin{array}{c|cccc|} & A_1 & A_2 & \cdots & A_d \\ \hline \vec{x}_1 & & & & \\ \vec{x}_2 & & & & \\ \vdots & & & & \\ \vec{x}_n & & & & \\ \hline \end{array}$$

$\vec{x}_i \in \mathbb{R}^d \leftarrow$ # of attributes

$A_j \in \mathbb{R}^n \leftarrow$ sample size

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

$\hat{\vec{\mu}} \in \mathbb{R}^d$

$$\boxed{\hat{\vec{\mu}} = \frac{1}{n} D^T \vec{1}}$$

$\vec{1} \in \mathbb{R}^n$

$$\underbrace{\underset{D^T}{(d\times n)}\ (n\times 1)}_{d\times 1}$$

$$= \frac{1}{n} \underbrace{\begin{bmatrix} - & A_1^T & - \\ - & A_2^T & - \\ & \vdots & \\ - & A_d^T & - \end{bmatrix}}_{D^T} \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{\vec{1}}$$

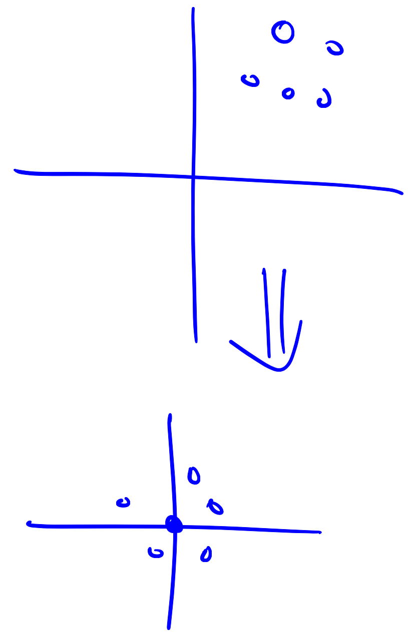
$$\underset{d\times d}{\hat{\Sigma}} = \begin{pmatrix} \hat{\sigma}_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1d} \\ \sigma_{12} & \hat{\sigma}_2^2 & \sigma_{23} & \cdots & \sigma_{2d} \\ \vdots & & \ddots & & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & & \hat{\sigma}_d^2 \end{pmatrix} = \underset{d\times d}{\{\sigma_{ij}\}}_{i,j=1,\dots,d}$$

Inner-product form

$$\bar{D} = \underset{n\times d}{D} - \underbrace{\underset{n\times 1}{\vec{1}} \cdot \underset{1\times d}{\hat{\mu}^T}}_{n\times d}$$

Centered data matrix

$$\hat{\Sigma} = \frac{\underset{d\cdot n}{\bar{D}^T}\ \underset{n\cdot d}{\bar{D}}}{n} \quad \text{(all pairwise dot-product)}$$



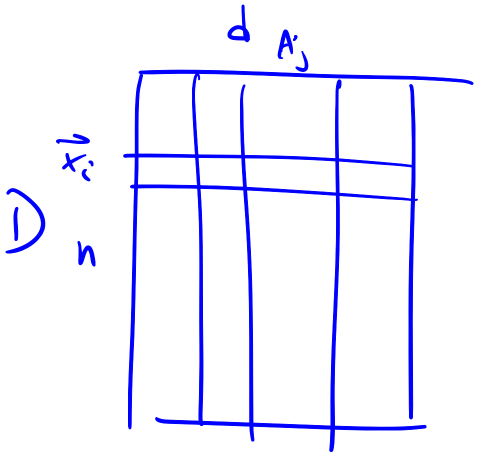
Outer-product form

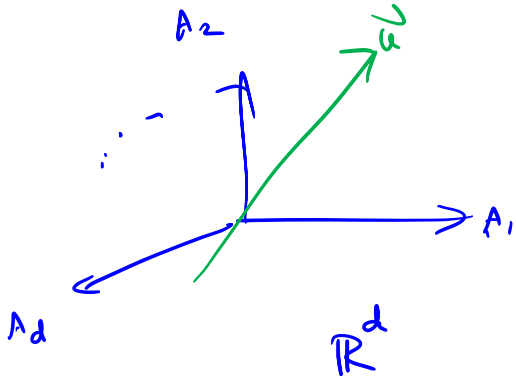
$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\underset{d\times 1}{(\bar{\vec{x}}_i)}\ \underset{1\times d}{(\bar{\vec{x}}_i)^T}}_{d\times d}$$

$$\vec{x}_i \longrightarrow \bar{\vec{x}}_i = \vec{x}_i - \hat{\vec{\mu}}$$

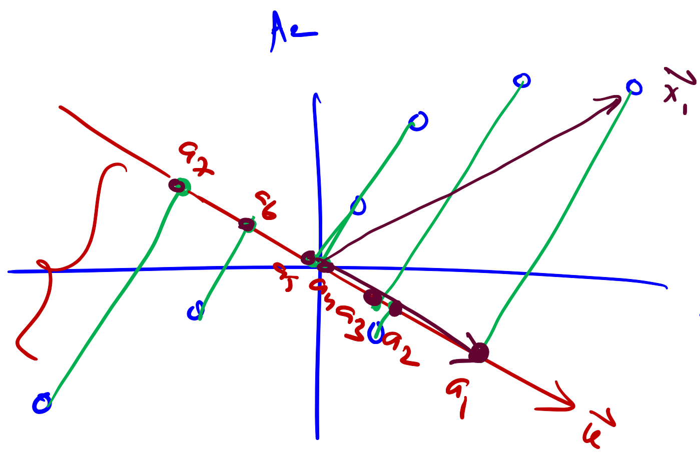
centered point

$$\underset{d\times d}{\hat{\Sigma}} = \underset{\bar{x}_1\bar{x}_1^T}{\Box} + \underset{\bar{x}_2\bar{x}_2^T}{\Box} + \Box + \cdots + \Box$$





"good" direction / attribute (constructed)



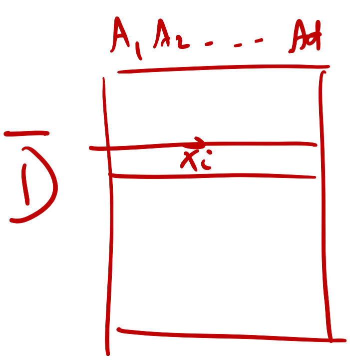
1) center data $\bar{D}$

2) $$\underbrace{\text{Proj}_{\vec{u}}(\vec{x}_1)}\cdot\vec{u} = \left(\frac{\vec{x}_1^T\vec{u}}{\underbrace{\vec{u}^T\vec{u}}}\right)\vec{u}$$

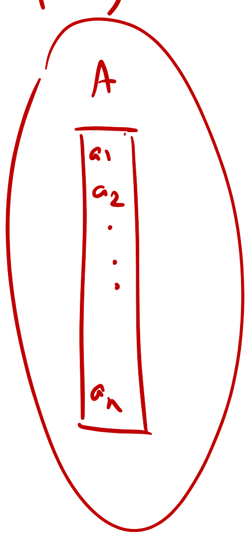
Scalar projection

all centered already ($\vec{x}_i = \bar{\vec{x}}_i$)

$$a_i = \text{Proj}_{\vec{u}}(\vec{x}_i)$$



$\Rightarrow$



3) $\vec{u}$ is unit vector

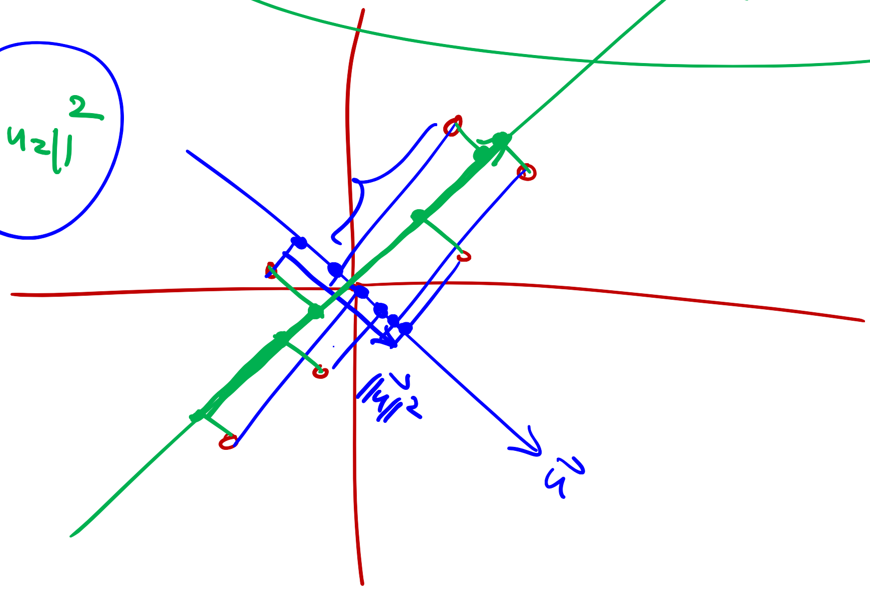
$$\|\vec{u}\|^2 = 1$$

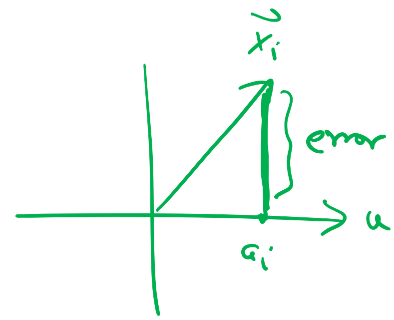
$$\vec{u}^T\vec{u} = 1$$

Objective: 1) maximize variance

2) minimize the squared error (sum over all points)

$\|u_2\|^2$





find $\vec{u}$ that maximize projected variance

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

2) $a_i = proj_{\vec{u}}(\vec{x}_i)$
$= \dfrac{\vec{u}^T\vec{x}_i}{\cancel{\vec{u}^T\vec{u}}\; 1}$

$$\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - 0)^2$$

$\hat{\mu}$ along $\vec{u}$ direction
= zero

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

all projected points (along $\vec{u}$)

find $\vec{u}$ $\sigma_u^2 = \frac{1}{n}\sum_{i=1}^{n} a_i^2$ to maximize

$$\boxed{J = \max_{\vec{u}} \sigma_{\vec{u}}^2 \; ?}$$

← Scalar value

$$\sigma_{\vec{u}}^2 = \frac{1}{n}\sum_{i=1}^{n} a_i^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\vec{u}^T\vec{x}_i\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\vec{u}^T\vec{x}_i\right)\left(\vec{u}^T\vec{x}_i\right)^T$$

$$(AB)^T = B^T A^T$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\vec{u}^T\vec{x}_i\right)\left(\vec{x}_i^T\vec{u}\right)$$

$$= \vec{u}^T\left(\frac{1}{n}\sum_{i=1}^{n}\vec{x}_i\vec{x}_i^T\right)\vec{u}$$

$$\boxed{\sigma_{\vec{u}}^2 = \underbrace{\vec{u}^T\Sigma\vec{u}}_{\text{quadratic form}}}$$

Σ = Covariance matrix of D

$\vec{u}$ is unknown

infinite choice

"random projections"

$$J = \max_{\vec{u}} \underbrace{\vec{u}^T\Sigma\vec{u}}_{\text{Constrained optimization}} \quad \boxed{\text{subject to } \vec{u}^T\vec{u} = 1}$$

$\vec{u}^T\vec{u} - 1 = 0$

$\vec{u} = \mathbb{R}^d = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{pmatrix}$

$$\nabla_{\vec{u}} = \boxed{\frac{\partial J}{\partial \vec{u}} = 0}$$

try to solve for $\vec{u}$

- closed form
- iterative gradient method

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

← unconstrained

α: Lagrange multiplier

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

$$\left[\begin{array}{l} \vec{u}^T\vec{u} = (\vec{u})^2 \\ \Sigma \vec{u}^T u = \vec{u}^T \Sigma \vec{u} \\ \Sigma \vec{u}^2 \end{array}\right]$$

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

special equation : eigen value (α) & eigen vectors ($\vec{u}$) of Σ

$\vec{u}$ & α

the direction that maximize the proj variance is $\vec{u}$, the eigenvector corresponding to the largest eigenvalue α of Σ

$$\vec{u}^T \Sigma \vec{u}$$

$$(u_1\ u_2 \cdots u_d) \begin{bmatrix} d \times d \\ \Sigma \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{pmatrix} = \sum_{i=1}^{d} \sum_{j=1}^{d} u_i \sigma_{ij} u_j$$