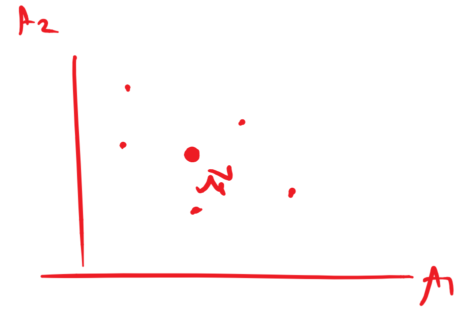
$$\begin{array}{c|c|c|} & A_1 & A_2 \\ \vec{x}_1 & x_{11} & x_{12} \\ \vec{x}_2 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ \vec{x}_n & & \end{array}$$



A_1 : Random variable

A_2 : R.V.

$$X = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \leftarrow \text{vector R.V}$$

$$E[X] = \begin{pmatrix} E[A_1] \\ E[A_2] \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \begin{matrix} \leftarrow \text{mean for } A_1 \\ \leftarrow \text{mean for } A_2 \end{matrix}$$

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} x_{i1} \qquad \vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

Sample mean

$$\hat{\sigma}_1^2 = \frac{1}{n} Z_1^T Z_1$$

var(A_1)

Z_1 : Centered A_1 vector

$$Z_1 = \begin{pmatrix} x_{11} - \mu_1 \\ x_{21} - \mu_1 \\ \vdots \\ x_{n1} - \mu_1 \end{pmatrix}$$

$$\sqrt{var(A_1)} = std(A_1) = \sigma_1$$

$$Cov(A_1, A_2) = \sigma_{12} = E[(A_1 - \mu_1)(A_2 - \mu_2)]$$

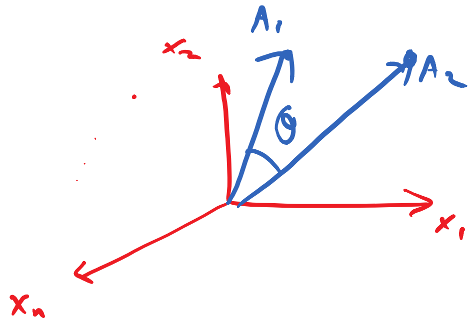
$$= \underbrace{E[A_1 A_2] - E[A_1] \cdot E[A_2]}$$

Sample covariance $\hat{\sigma}_{12}$

$$\boxed{\hat{\sigma}_{12} = \frac{1}{n} \sum_{i=1}^{n} (x_{i1} - \mu_1)(x_{i2} - \mu_2)} \leftarrow$$

$$\boxed{\hat{\sigma}_{12} = \frac{1}{n} \sum_{i=1} \underbrace{(x_{i1} - \mu_1)(x_{i2} - \mu_2)}}$$

point-view



$$\hat{\sigma}_1^2 = \frac{1}{n} \|z_1\|^2 = \frac{1}{n} z_1^T z_1$$

$$\hat{\sigma}_2^2 = \frac{1}{n} \|z_2\|^2 = \frac{1}{n} z_2^T z_2$$

$$\hat{\sigma}_{12} = \frac{1}{n} z_1^T z_2$$

Cov

dot-product of centered attr. vec.

Correlation $\rho_{12} = \dfrac{\sigma_{12}}{\sigma_1 \cdot \sigma_2}$

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_1^2 \cdot \sigma_2^2}}$$

$$\hat{\rho}_{12} = \frac{\cancel{\frac{1}{n}} z_1^T z_2}{\sqrt{\cancel{\frac{1}{n} \cdot \frac{1}{n}} \cdot \|z_1\| \cdot \|z_2\|}} = \left(\frac{z_1}{\|z_1\|}\right)^T \left(\frac{z_2}{\|z_2\|}\right) = \cos\theta$$

Smaller angle

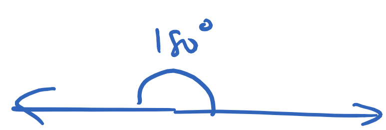
$\theta = [0, 180^\circ]$

$\theta = 90^\circ \; : \; \cos\theta = 0$

orthogonal

$\underline{\underline{\theta = 0^\circ}} \; : \; \cos\theta = 1$

parallel



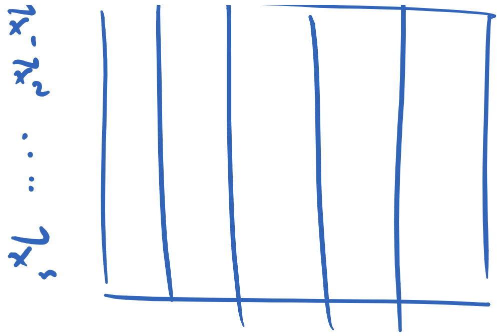
$\theta = 180^\circ \; : \; \cos\theta = -1$

$A_1 \; A_2 \; \ldots \; A_d$

$\vec{x}_1$

$\vec{x}_2$

$$\vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \end{pmatrix}$$



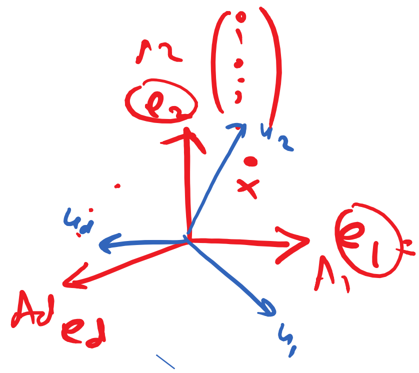
$$\vec{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\\ \vdots \\ \mu_d\end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2d}\\ & \vdots & & \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_d^2 \end{pmatrix}$$

Covariance Matrix

Variance

$\sigma_{ij} = \sigma_{ji}$



$$\vec{x} = x_1\vec{e_1} + x_2\vec{e_2} + \cdots + x_d\vec{e_d}$$

$$= \begin{pmatrix}x_1\\ x_2\\ \vdots \\ x_d\end{pmatrix} = \underbrace{c_1\vec{u_1} + c_2\vec{u_2} + \cdots + c_d\vec{u_d}}$$

$d\times 1$

(1) $d\times d$: square

(2) Symmetric matrix

(3) positive semi-definite (PSD)

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

real, non-negative eigenvalue

Matrix Σ is PSD

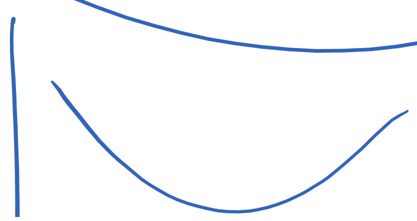
for all vectors $\underline{x} \in \mathbb{R}^d$

$$\underbrace{x^T \Sigma x}_{\text{quadratic form}} \geq \underline{0}$$

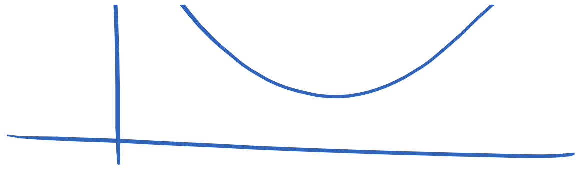
$$x^T y \geq 0 \qquad y = \Sigma x$$

$$x^T x = \underbrace{x^T I x}_{\text{Scalar}}$$

$\det(\Sigma) \equiv$ generalized variance



$$\det(\Sigma) = \prod^{d} \lambda_i$$



$$\det(\Sigma) = \prod_{i=1}^{d} \lambda_i$$

if some eigenvalue $\lambda_i = 0$

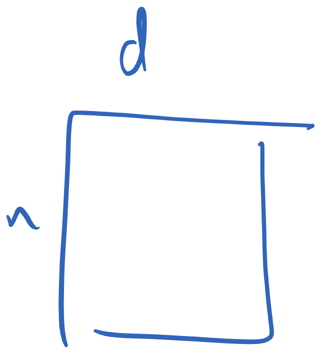
$$\det(\Sigma) = 0$$

$$\Downarrow$$

Σ is singular

it doesn't have an inverse

$D =$



$$\vec{X} = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_d \end{pmatrix}$$

vec. r.v.

$\vec{X}$: Unknown

Parametric statistics

vs.

non-parametric statistics

assume $\vec{X}$ is a multivariate Normal distribution

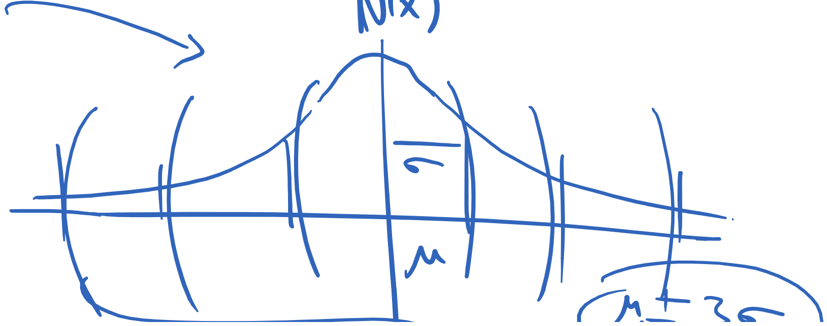
infer μ_i, σ_{ij}, ...

density - estimation

$$\boxed{A_1 \sim N(x \mid \mu, \sigma^2)}$$

distributed normally

$N(x \mid 0, 1)$



$N(x \mid 0, 1)$ ... $|\mu|$ ... $\mu \pm 3\sigma$

Standard normal

$= 99.99\%$

z-score : $\dfrac{x - \mu}{\sigma}$

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

threshold

1-dim

exponential decay as you move away from μ.

Multivariate Normal (MVN)

$$N(\vec{x} \mid \vec{\mu}, \Sigma) = \frac{1}{(\sqrt{2\pi})^d \det(\Sigma)} \exp\left\{\frac{(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}{2}\right\}$$

Parameters of $\underbrace{N}_{\text{infer !!!}}$

Standard MVN : $\vec{\mu} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \vec{0}$

$\Sigma = I$

$\Sigma^{-1} = I$

$\vec{z} = \vec{x} - \vec{\mu}$

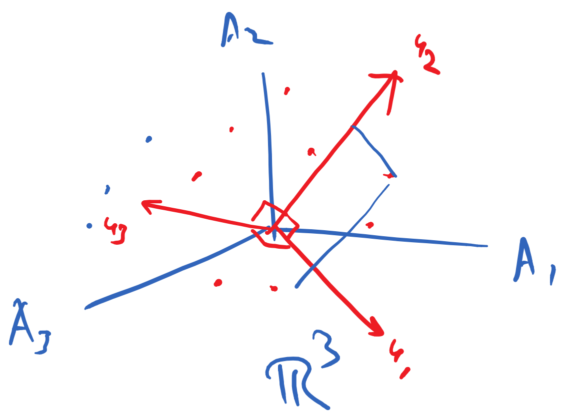
$z^T \Sigma^{-1} z \leftarrow$ quadratic form.

$\hat{\Sigma}$: eigenvalues

eigenvectors

Az

PCA as a change of basis vectors



$$\hat{\Sigma} = \begin{pmatrix} \sigma_1^2 & & \sigma_{ij} & \\ & \sigma_2^2 & & \\ \sigma_{ji} & & \ddots & \\ & & & \sigma_d^2 \end{pmatrix}$$

$\sigma_{ij} \neq 0$

Original basis:

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \longleftarrow \text{standard basis matrix in } \mathbb{R}^3$$

New basis

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

$$V = \begin{bmatrix} | & | & | \\ \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \\ | & | & | \end{bmatrix} \longrightarrow \Sigma_u = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_d^2 \end{bmatrix}$$

all covariances in the new basis are 0

$$\boxed{\Sigma \vec{u}_i = \lambda_i \vec{u}_i}$$

eigenvector/value equation

d eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \geq 0$

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

$$\vec{u}_i^T \vec{u}_j = 0 \longleftarrow \text{orthogonal eigenvectors}$$

u_d

$\vec{u}_i^T \vec{u}_j = 0$ ← orthogonal eigenvectors

$U = \begin{bmatrix} \text{matrix of eigenvector as col's} \\ \end{bmatrix}$

→ Orthogonal matrix

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

$U^T U = I$

$\Sigma \vec{u}_i = \lambda_i \vec{u}_i$ →

$\Sigma \vec{u}_1 = \lambda_1 \vec{u}_1$
$\vdots$
$\Sigma \vec{u}_d = \lambda_d \vec{u}_d$

$\Sigma U = U \Lambda$

$\Sigma = U \Lambda U^T$

Cov in original basis

Cov in new basis

$\Lambda = \underbrace{\begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{pmatrix}}_{\text{diagonal matrix}}$

→ Eigenvalues $\underline{\underline{\text{are}}}$ the variance in U basis eigen

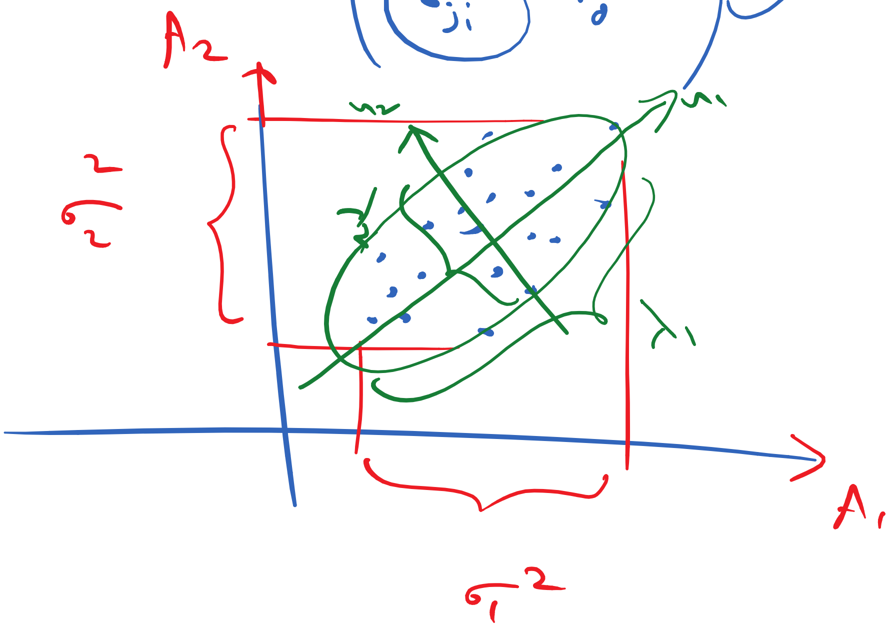
$\underline{\underline{\text{totvar}}}(D) = \frac{1}{n} \sum \|\vec{x} - \vec{\mu}\|^2$

$= tr(\Sigma) = \sigma_1^2 + \sigma_2^2 + \dots \sigma_d^2 = tr(\Lambda)$

$$= tr(\Sigma) = \sigma_1^2 + \sigma_2^2 + \dots \sigma_d^2 = tr(\Lambda)$$

$$= \lambda_1 + \lambda_2 + \dots + \lambda_d$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & \sigma_{ij} \\ & \sigma_2^2 & \\ \sigma_{ji} & & \ddots \ \sigma_d^2 \end{pmatrix}$$



σ_{12} is positive

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

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

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

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

$r << d$

10   1000

keep r dimensions

throw away remaining $d-r$ dimensions

Choose r based on "approximation", e.g. 90% of total variance

$$\min_r \left\{ \frac{\lambda_1 + \lambda_2 + \dots + \lambda_r}{totvar(D)} \geq \theta \right\}$$

L   totvar(D)   threshold

PCA as dimensionality reduction

d is large

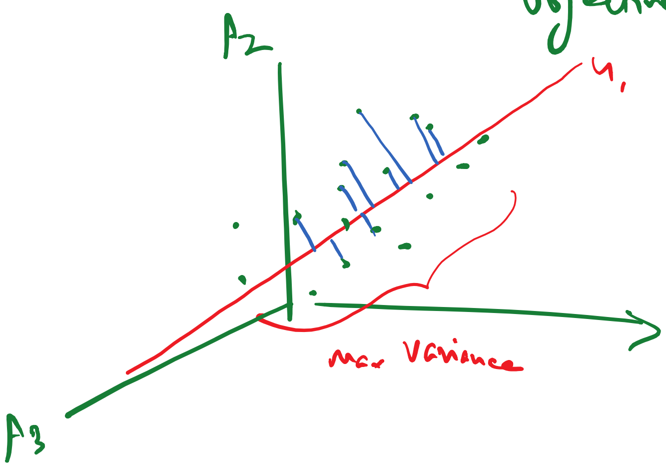
0-dimensional approx : mean point

1-dimensional approx: find the best line

Objective: maximize projected variance (variance)

|||

minimize error (sum of error)



u_1 ?

$\|\vec{\varepsilon}\|^2$

best 2D

min error

$\vec{x} - \vec{x}_p = \vec{\varepsilon}$

error vector