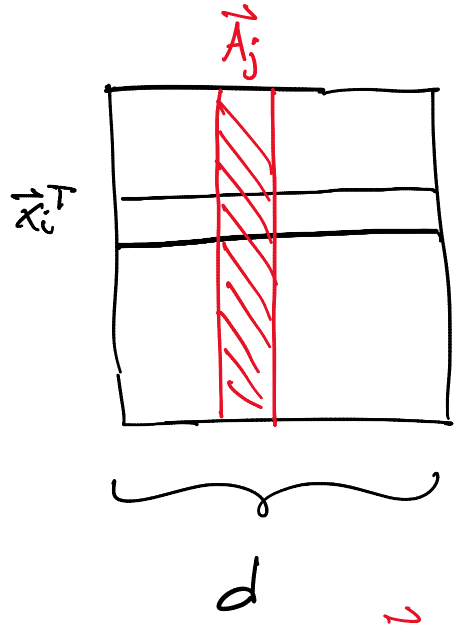
Lecture 2

Thursday, August 31, 2023 9:55 AM

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



$$\vec{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{pmatrix} \in \mathbb{R}^d$$

points/ vectors /
examples
Samples

$\vec{A}_j \in \mathbb{R}^n$

attributes

random variables

$\vec{a}, \vec{b} \in \mathbb{R}^m$

similarity

$$\underbrace{\vec{a}^T \vec{b}}_{\text{Scalar}} \equiv \vec{a} \cdot \vec{b} \equiv \langle \vec{a}, \vec{b} \rangle = \sum_{i=1}^{m} a_i b_i$$

$$\|\vec{a} - \vec{b}\|^2 = \sum_{i=1}^{m} (a_i - b_i)^2$$

distance

squared norm

→ squared error

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

squared norm

$$\|\vec{a}\|^2 = \underline{\underline{\vec{a}^T\vec{a}}}$$ squared norm

$$\|\vec{a}-\vec{b}\|^2 = (\vec{a}-\vec{b})^T(\vec{a}-\vec{b})$$

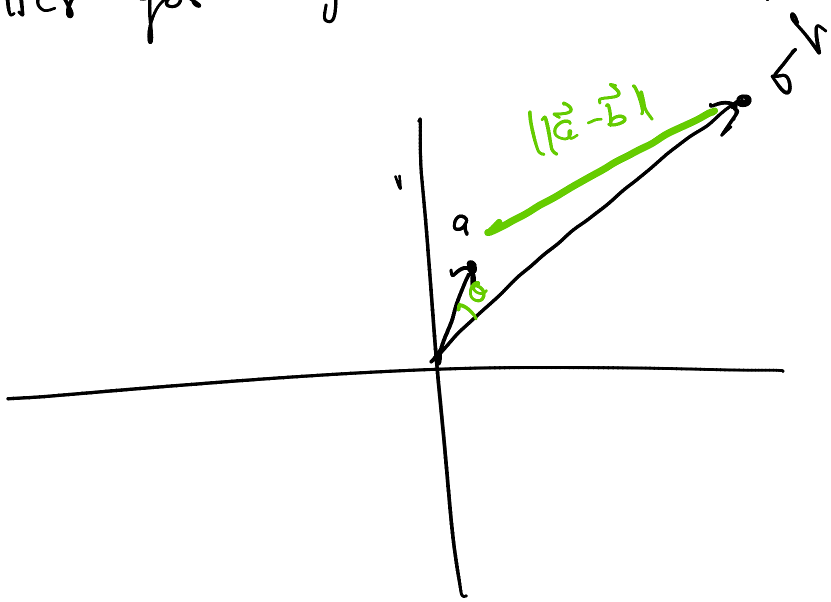
$$= \vec{a}^T\vec{a} + \vec{b}^T\vec{b} - 2\vec{a}^T\vec{b}$$

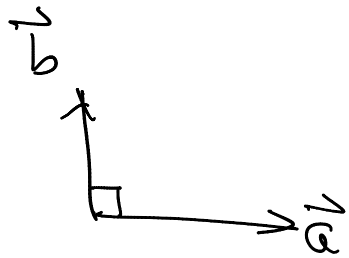
$$\boxed{\underbrace{\|\vec{a}-\vec{b}\|^2}_{\text{distance}} = \underline{\|\vec{a}\|}^2 + \underline{\|\vec{b}\|}^2 - 2\,\underbrace{\vec{a}\cdot\vec{b}}_{\text{similarity}}}$$

$$\cos\theta = \underbrace{\left(\frac{\vec{a}}{\|\vec{a}\|}\right)^T\left(\frac{\vec{b}}{\|\vec{b}\|}\right)}_{\text{dot product}} = \frac{\vec{a}^T\vec{b}}{\|\vec{a}\|\,\|\vec{b}\|}$$

$=$ smallest angle between the two vectors

better for high-dimensional sparse spaces



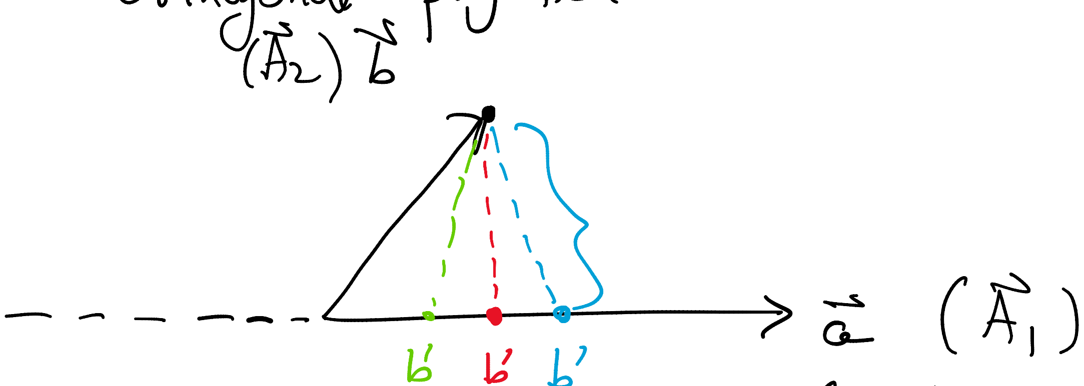


$\underline{\underline{\cos}}\,\theta = 0 \Rightarrow \theta = 90^\circ = \frac{\pi}{2}$

" Correlation "

Orthogonal Vectors

Orthogonal projection



$\vec{b}$ is a vector outside $\vec{a}$

$\underbrace{(\text{ 1d } \underline{\underline{\text{manifold}}})}_{\text{line}}$

Q: what is the $\underline{\underline{\text{'best'}}}$ approximation to $\vec{b}$ on $\vec{a}$

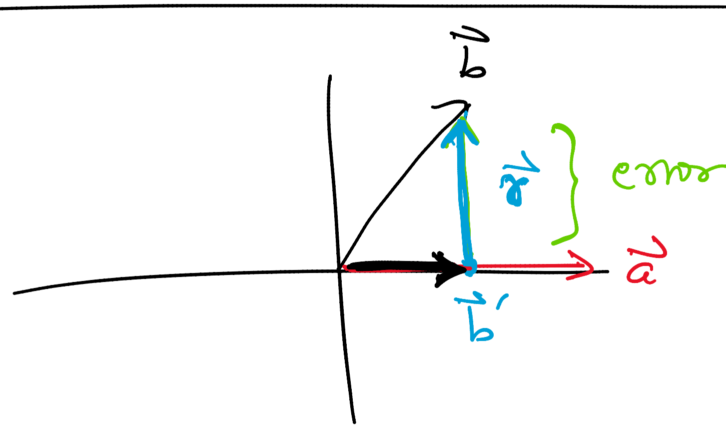
$\vec{b}' =$ point along $\vec{a}$ that is projection of $\vec{b}$

$$\min_{\vec{b}'} \left\{ \| \underbrace{\vec{b} - \vec{b}'}_{\text{error}} \|^2 \right\}$$

$=$ b' is the orthogonal projection

Optimal Solution

$$\vec{b}' = \underbrace{\boxed{\underbrace{\left(\frac{\vec{b}^T \vec{a}}{\vec{a}^T \vec{a}} \right)}_{\substack{\text{Scalar} \\ \text{projection} \\ \text{of } b \text{ onto } a}} \vec{a}}}_{\substack{\text{projection of} \\ \vec{b} \text{ onto } \vec{a}}} = \underbrace{\left(\frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|} \right)}_{\substack{\text{\# of units} \\ \text{along } \vec{u}_a}} \underbrace{\left(\frac{\vec{a}}{\|\vec{a}\|} \right)}_{\substack{\text{unit vector} \\ \vec{u}_a}}$$



$$\vec{b} = \underset{\substack{\uparrow \\ \text{parallel}}}{\vec{b}'} + \underset{\substack{\nwarrow \\ \text{perpendicular}}}{\vec{\gamma}} \implies \vec{\gamma} = \vec{b} - \vec{b}'$$

$$\underbrace{\text{parallel } \vec{a} \qquad \text{perpendicular to } \vec{a}}_{\text{orthogonal decomposition}}$$

$$D = \begin{array}{c} A_1 \\ \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array}\right] \end{array}$$

$A_1 \equiv$ random variable

x_i is also a random variable from the same distribution as A_1

all x_i's are independent of each other

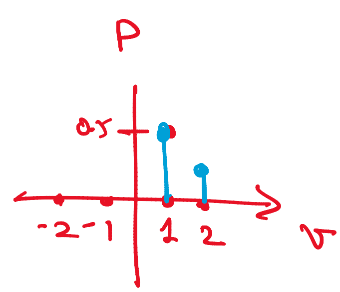
IID : independent & identically distributed

Probability distribution:

$A_1 \equiv$ discrete ($\mathbb{I}$) ← Integers

$A_1 \equiv$ Continuous ($\mathbb{R}$) ← real space

Discrete



$\{P(A_1 = v) \equiv$ probability of value v

$$\left\{\begin{array}{l} P(A_1 = v) \equiv \text{probability of value } v \\ \text{for all } v \end{array}\right. \implies \text{PMF} \quad \text{prob mass function}$$

$f(x) \rightarrow$

Continuous

$$P(A_1 \in [a,b]) = \int_a^b f(x)\, dx$$

prob density function PDF

distribution of A_1 is <u>unknown</u>

Infer:

expected value or mean

Discrete:

$$\mu = \mathbb{E}[A_1] = \sum_v P(A_1 = v) \cdot v$$

A_1
x_1
x_2
,
:
x_n

Sample of n points

true value $\leftarrow$

Unknown!

Continuous:

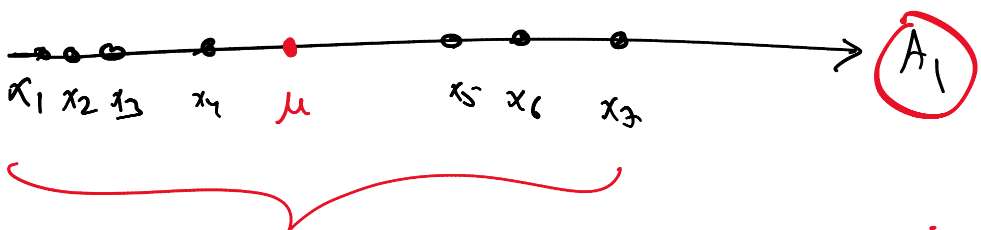
$$\mu = \mathbb{E}[A_1] = \int_v f(x) \cdot x \, dx$$

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

some real valued function of random variables

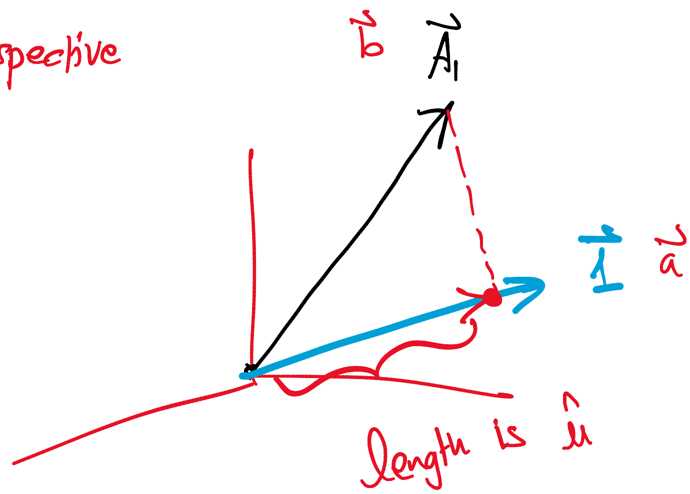
$$\hat{\mu} = f(x_1 x_2 \dots x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\hat{\mu}$ point estimate for the unknown parameter μ



from the sample perspective

$$A_1 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix}$$



$A_1 \in \mathbb{R}^n$

$\vec{1} \in \mathbb{R}^n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

$\mathbb{R}^n$

$\hat{\mu}$ is the scalar projection of $\vec{A_1}$ onto $\vec{1}$

$$\left(\frac{\vec{A_1}^T \vec{1}}{} \right) \vec{1}$$

$$\underbrace{\left(\frac{A_1^T \vec{1}}{\vec{1}^T \vec{1}}\right)}_{\text{Scalar projection}} \vec{1}$$

$$(x_1\ x_2 \dots x_n)\begin{pmatrix}1\\1\\1\\\vdots\\1\end{pmatrix} = \sum x_i$$

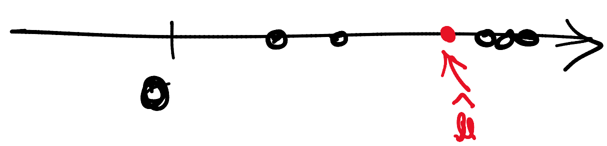
$$(1\,1\,1\,1 \dots 1)\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix} = n$$

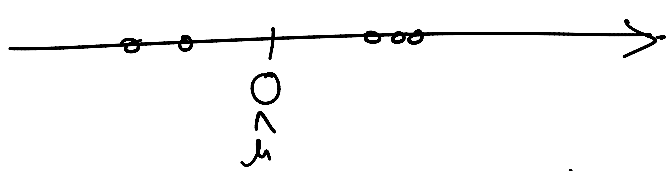
$$\Downarrow$$

$$\frac{\sum x_i}{n} = \hat{\mu}$$

$$D = \begin{array}{c} A_1 \\ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \end{array}$$

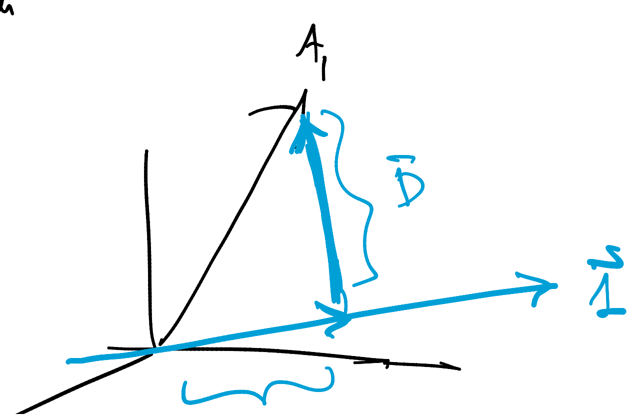
$\hat{\mu}$ = sample mean = $\frac{1}{n} \sum x_i$

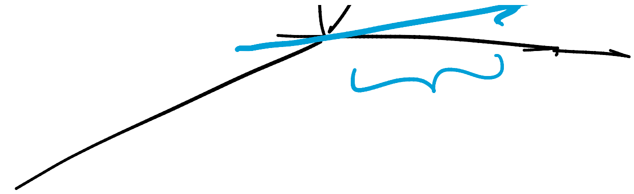




Centered data

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





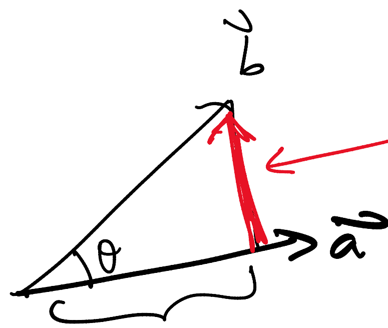
$$A_1 - \hat{\mu}\vec{1} = \vec{D}$$

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

$\vec{D}$ and $\vec{1}$ are orthogonal!

Orthogonalize one vector w.r.t. another

$\vec{a}$ and $\vec{b}$



new $\vec{r}$

Orthogonalize $\vec{b}$ w.r.t $\vec{a}$

with respect to

and $\vec{r}$ will be orthogonal to $\vec{a}$

$$\vec{b} - \underbrace{\mathrm{proj}_a(\vec{b})}\cdot\vec{a}$$

$$\vec{r} = \vec{b} - \left(\frac{\vec{b}^T\vec{a}}{\vec{a}^T\vec{a}}\right)\vec{a}$$

$\hat{\mu} \equiv$ centrality / location

$E[\hat{\mu}] = \mu$

$$A_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$\hat{\sigma}^2 \equiv$ sample variance / $\hat{\sigma} \equiv$ sample standard deviation

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

$\hat{\sigma} \equiv$ variance / $\hat{\sigma} \equiv$ standard deviation

$\sigma^2 \equiv$ true variance $\quad \sigma \equiv$ std

$$E\left[\left(\underline{x_i - \mu}\right)^2\right]$$

Continuous $\equiv \displaystyle\int_{-\infty}^{\infty} \left(\underline{x_i - \mu}\right)^2 \cdot f(x)\, dx$

$$\boxed{E[g(x)] = \int g(x) \cdot \underline{f(x)}\, dx}$$

pdf

$$\mu = E[x] \Rightarrow g(x) = x$$

$$\sigma^2 = E\left[(x-\mu)^2\right] \Rightarrow g(x) = (x-\mu)^2$$

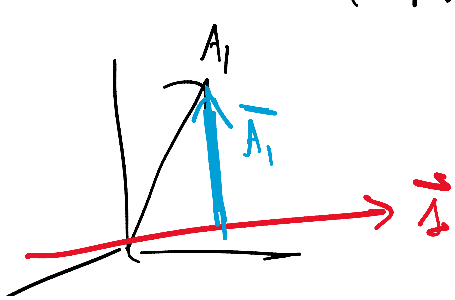
$$\boxed{\hat{\sigma}^2 = \frac{1}{n} \sum \left(\underline{x_i} - \underline{\hat{\mu}}\right)^2}$$

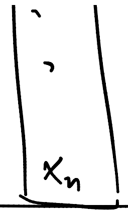
Sample variance

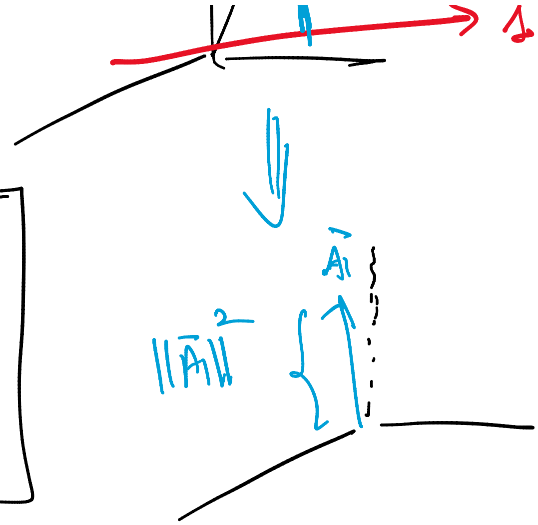
dispersion around the mean.

A_1

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix}$$







biased

$$\boxed{\hat{\sigma}^2 = \frac{\|\vec{A}_1\|^2}{n} = \frac{\|\vec{D}\|^2}{n}}$$

Variance is the "avg" squared norm per dimension

$$\hat{\sigma}_u^2 = \frac{\sum (x_i - \hat{\mu})^2}{n-1}$$

unbiased

bias of an estimator

$$E[\hat{\sigma}^2] \stackrel{?}{=} \sigma^2$$

r.v.

true parameter

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

$$E[\hat{\sigma}_u^2] = \sigma^2 \leftarrow \text{unbiased}$$

from a point perspective $\hat{\mu}$ uses up one degree of freedom out of the n

$$\hat{\mu} = \frac{1}{n}\left(\underbrace{x_1 + x_2 + \dots + x_n}\right)$$

1 scalar

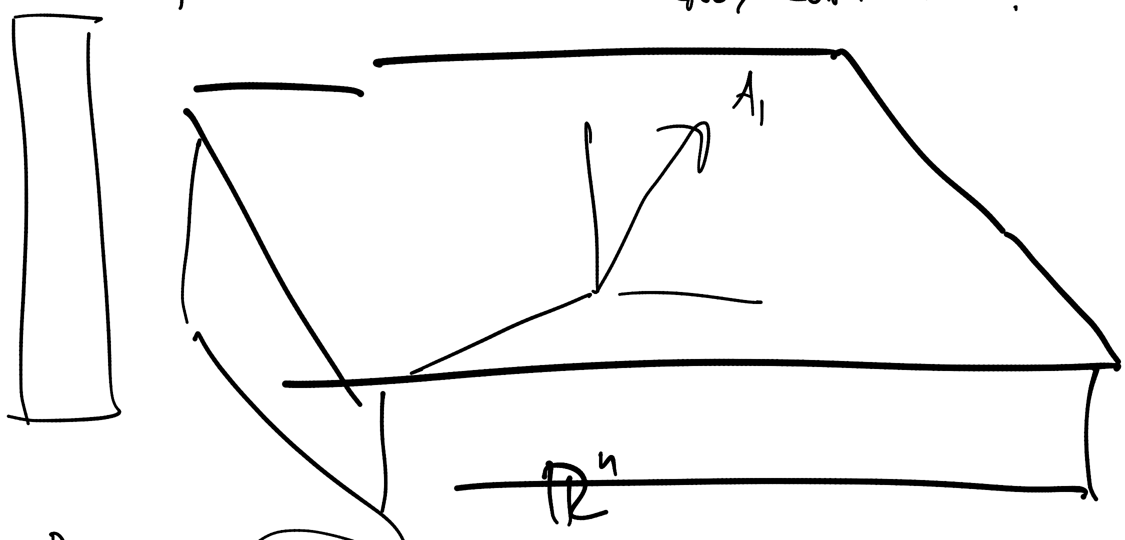
$n-1$ degrees of freedom left

$$\hat{\sigma}_u^2 = \frac{1}{n-1} \sum (x_i - \hat{\mu})$$

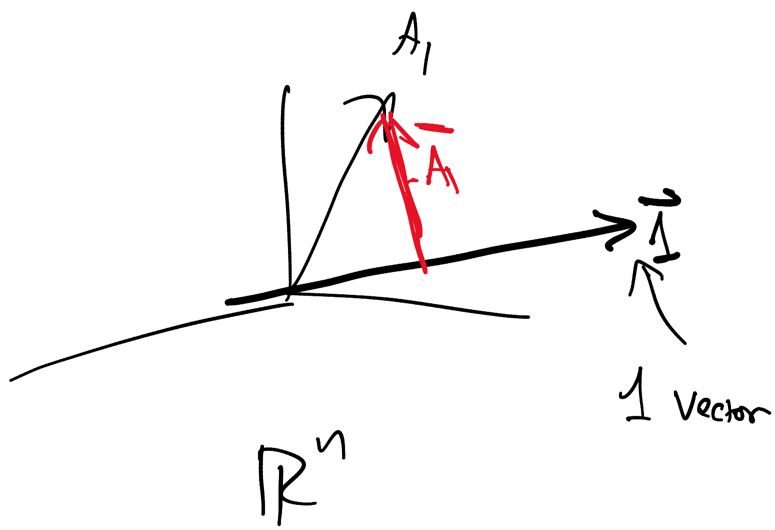
<u>Column perspective:</u>

degrees of freedom $\equiv$ dimensionality of the subspace where the vector/data lives.

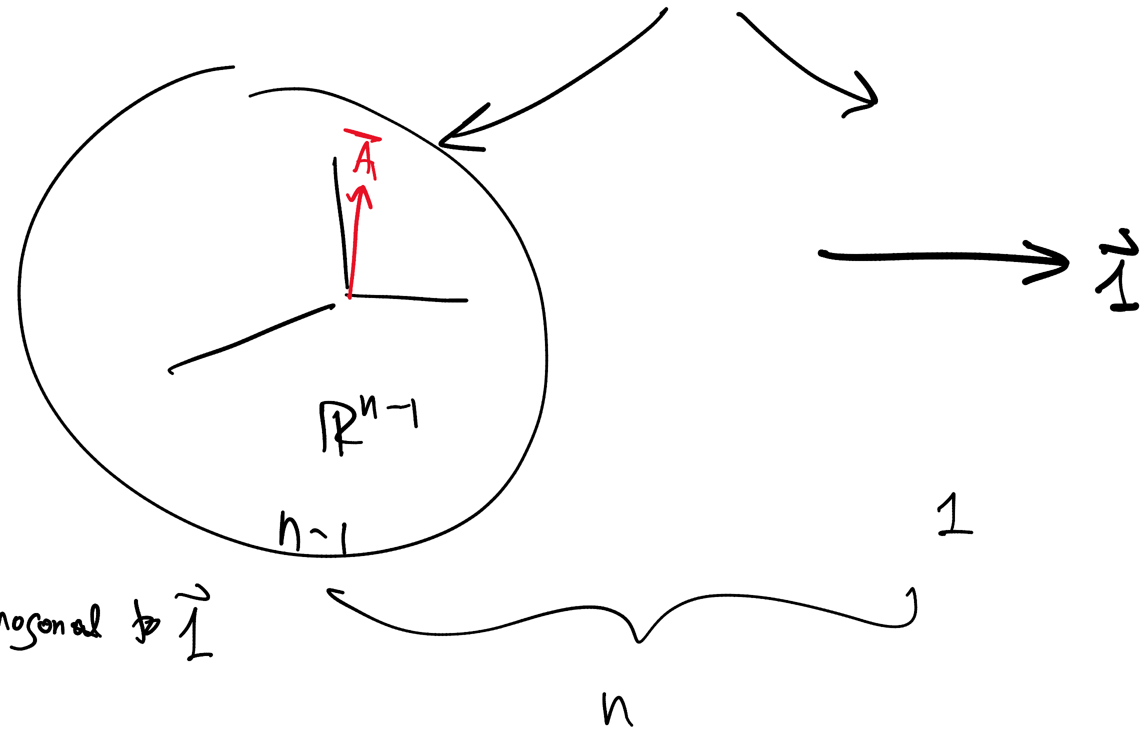
$A_1 \in \mathbb{R}^n$



$$\frac{\|\bar{A}_1\|^2}{n} \quad \text{or} \quad \frac{\|\bar{A}_1\|^2}{n-1}$$



$\text{dof} = \underline{\underline{n-1}}$



Orthogonal to $\vec{1}$

$$D = \begin{array}{c|cc} & A_1 & A_2 \\ \hline \vec{x}_1^T & x_{11} & x_{12} \\ \vec{x}_2^T & x_{21} & x_{22} \\ & \vdots & \\ \vec{x}_n^T & x_{n1} & x_{n2} \end{array} \quad n$$

$d = 2$

$A_1, A_2 \in \mathbb{R}^n$

$\vec{x}_i^T \in \mathbb{R}^2$

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

Sample mean

$$D = \begin{array}{c|cc} & A_1 & A_2 \\ \hline x_1^T & 1 & 0 \\ x_2^T & 0 & 1 \\ x_3^T & 1 & 1 \\ x_4^T & 2 & 3 \end{array}$$

$\frac{4}{4} \quad \frac{5}{4}$

$\hat{\mu} = \left(1, \frac{5}{4}\right)^T \leftarrow$ location

$\wedge 2 \qquad \wedge 2$

$\frac{4}{4}$ $\frac{5}{4}$

$\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$

variance for A_1 for A_2

true Covariance $\equiv E\left[(A_1-\mu_1)(A_2-\mu_2)\right]$

Unknown $= E\left[\bar{A}_1 \cdot \bar{A}_2\right]$

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

Sample Covariance between A_1 and A_2

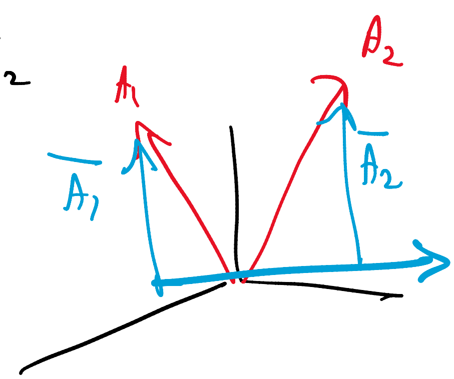
$\vec{x}_i^T = (x_{i1}, x_{i2})$

A_1 A_2

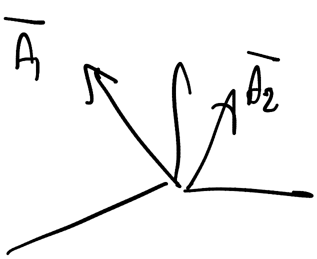
asymptotically unbiased

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

Col



$\mathbb{R}^n$



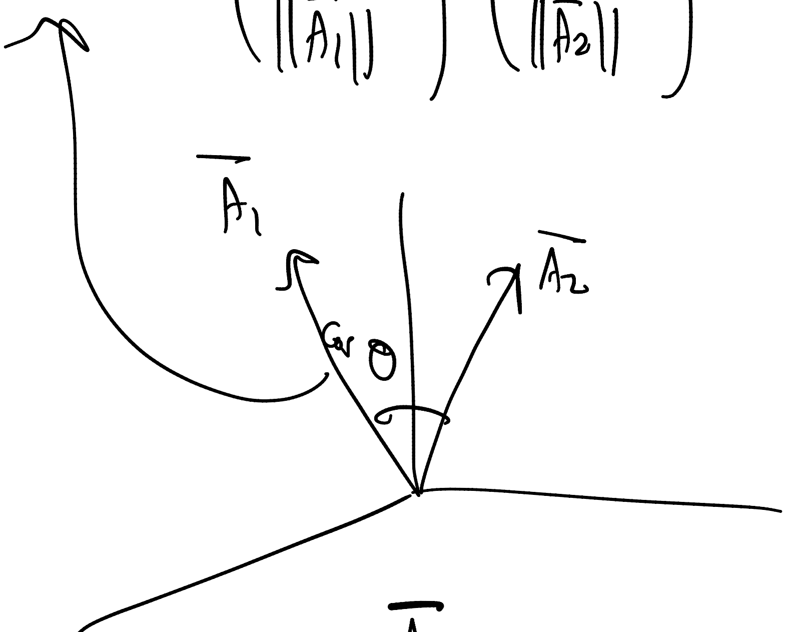
$\mathbb{R}^{n-1}$

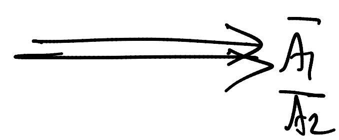
$$\boxed{E\left[\hat{\sigma}_{12}\right] = \sigma_{12} \quad \text{as} \quad n \to \infty}$$

$$\text{Correlation} = \frac{\hat{\sigma}_{12}}{\sqrt{\sigma_1^2 \cdot \sigma_2^2}} = \frac{\sum (x_{i1} - \hat{\mu}_1)(x_{i2} - \hat{\mu}_2)}{\sqrt{\sum (x_{i1} - \mu_1)^2 \sum (x_{i2} - \mu_2)^2}}$$

$$= \frac{\bar{A}_1^T \bar{A}_2}{\|\bar{A}_1\| \, \|\bar{A}_2\|}$$

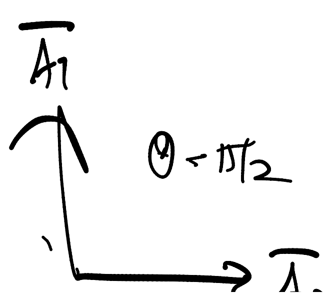
$$\text{Correlation} = \left(\frac{\bar{A}_1}{\|\bar{A}_1\|}\right)^T \left(\frac{\bar{A}_2}{\|\bar{A}_2\|}\right) = \cos\theta_{(A_1 \,\&\, A_2)}$$



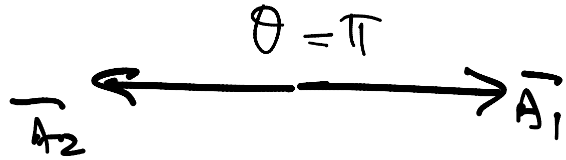


$\cos\theta = 1$

$\theta = 0$



$\cos\theta = 0$



$\cos\theta = -1$