A = Age

x_1
x_2
$\vdots$
x_n

sample space

$A : \Omega \longrightarrow \mathbb{R}$

$x_i \in \mathbb{R}$

A : random variable

$\{x_1, x_2, \ldots, x_n\}$: random sample

Set of random variables IID with A

IID → independent, identically distributed

"unknown" probability distribution

→ Continuous var: PDF, prob. density function

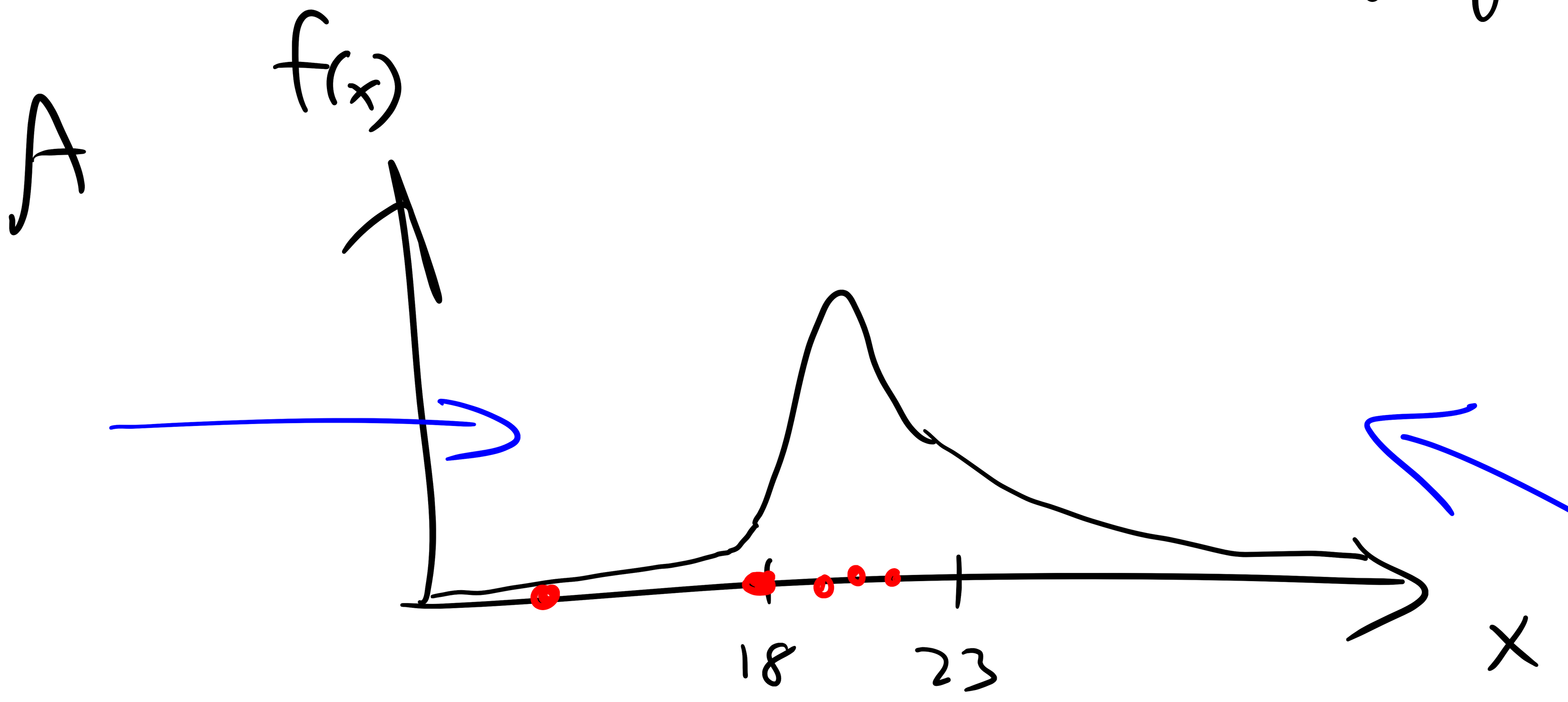
→ discrete: PMF, prob mass function

Prob mass → $P(A=50) = 0$

$P(A = 50) = 0.2$

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

density function



Unknown

Population Parameters

μ mean → location ?

σ^2 variance → spread ?

σ Standard deviation ? (std)

Sample
- mean $\hat{\mu}$
- variance $\hat{\sigma}^2$
- std $\hat{\sigma}$

statistics

$x_1 = 18$
$x_2 = 5$
$x_3 = 20.1$
$x_4 = 19.5$
$\vdots$

D

Salary : $\{ \underbrace{100k, 200k, 400k}, \ldots, 750 \text{ million} \}$

mean

median ← robust

Statistic : $\hat{\theta} : \{x_1, x_2 \dots x_n\} \Rightarrow \mathbb{R}$

$\downarrow$

true but unknown parameter $\rightarrow \theta$

Mean : Expected value of the random variable (rv)

$\mu = E[A]$ $\nearrow$ $\int_x x \cdot f(x)\, dx$ Continuous

x ↑ value, $f(x)$ ↑ density

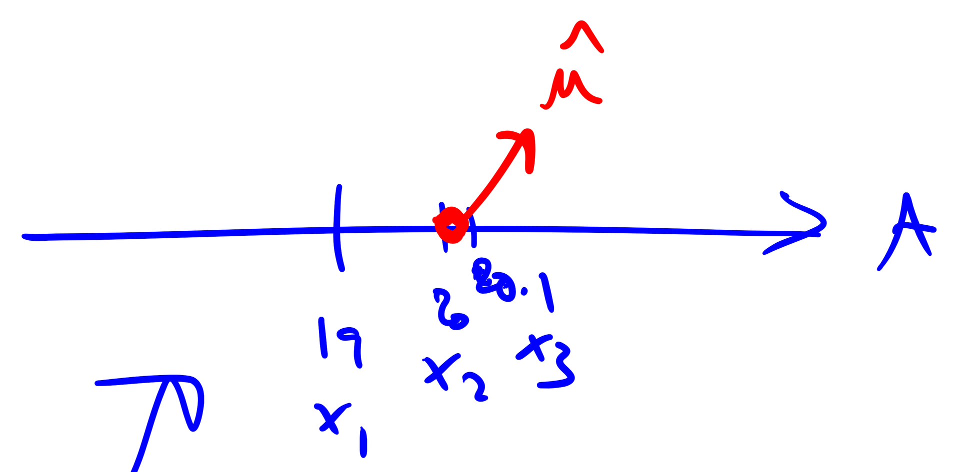
$\searrow$ $\sum_x x \cdot P(A=x)$ discrete

Sample mean

A	Age
x_1	19
x_2	20
$\vdots$	20.1
$\vdots$	22
$\vdots$	18.5
x_n	$\vdots$

random sample

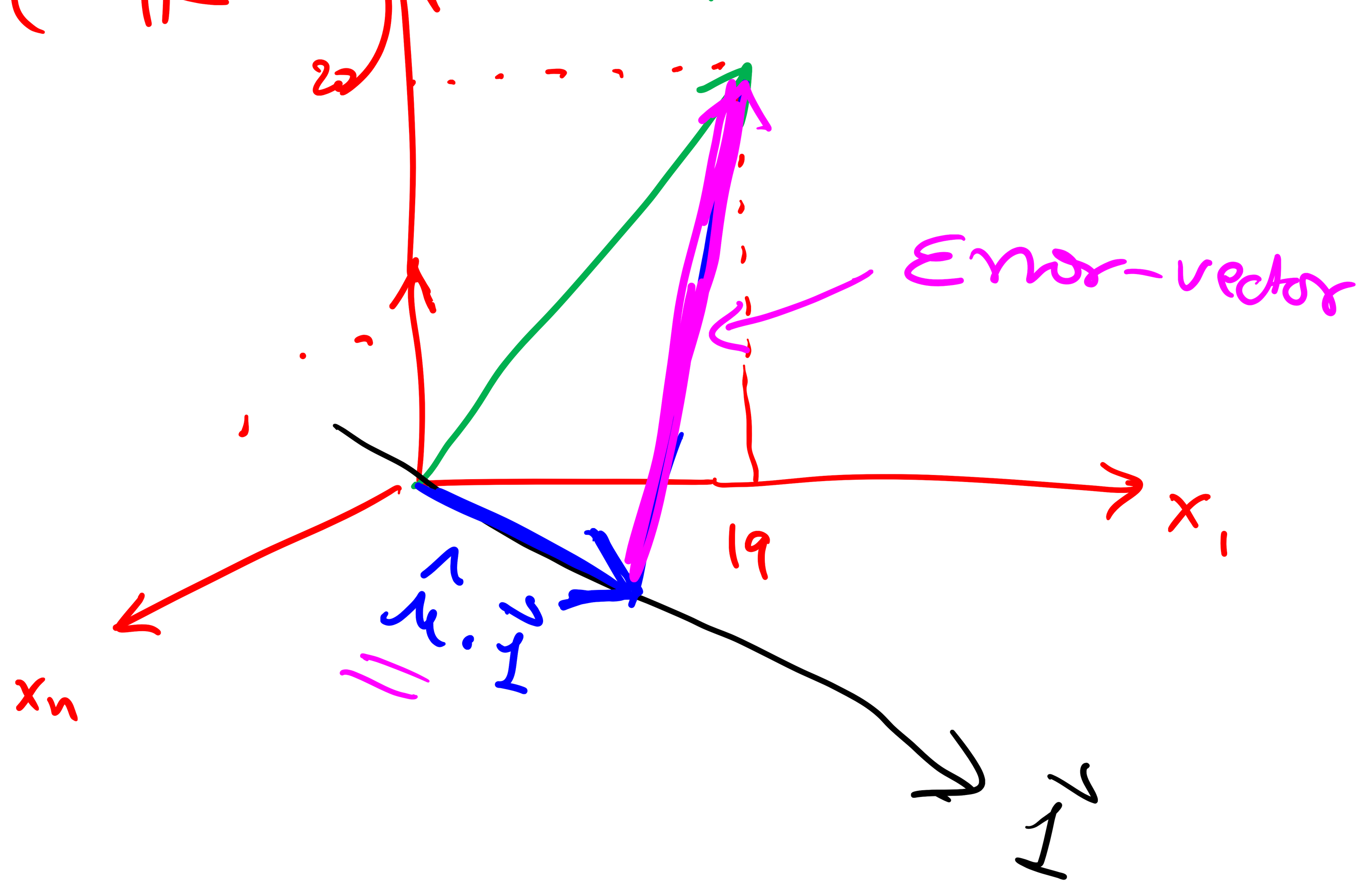
n : sample size



$\rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$ (row-view)

Column-view ($\mathbb{R}^n$)

$A = (19, 20, 20.1, 22, 18.5, \dots)^T$



Ones-vector

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

dot-product

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

$\mathbb{R}^n$

$$\vec{a} \cdot \vec{b} = \vec{a}^T \vec{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

$$= \sum_{i=1}^{n} a_i b_i$$

dot product

$\langle \vec{a}, \vec{b} \rangle$

inner-product

$\Downarrow$

leads to the notion of length (norm)

$$\vec{a}^T \vec{a} = \sum_{i=1}^{n} a_i^2 = a_1^2 + a_2^2 + \cdots + a_n^2 = \|\vec{a}\|^2$$

$$\|\vec{a}\|_2 = \sqrt{\|\vec{a}\|_2^2} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

l_2 norm

$$l_p\text{-norm} = \left(\sum_{i=1}^{n} |a_i|^p \right)^{1/p}$$

$$\boxed{\text{as } p \to \infty \quad l_\infty = \max_{i=1}^{n} \{ |a_i| \}}$$

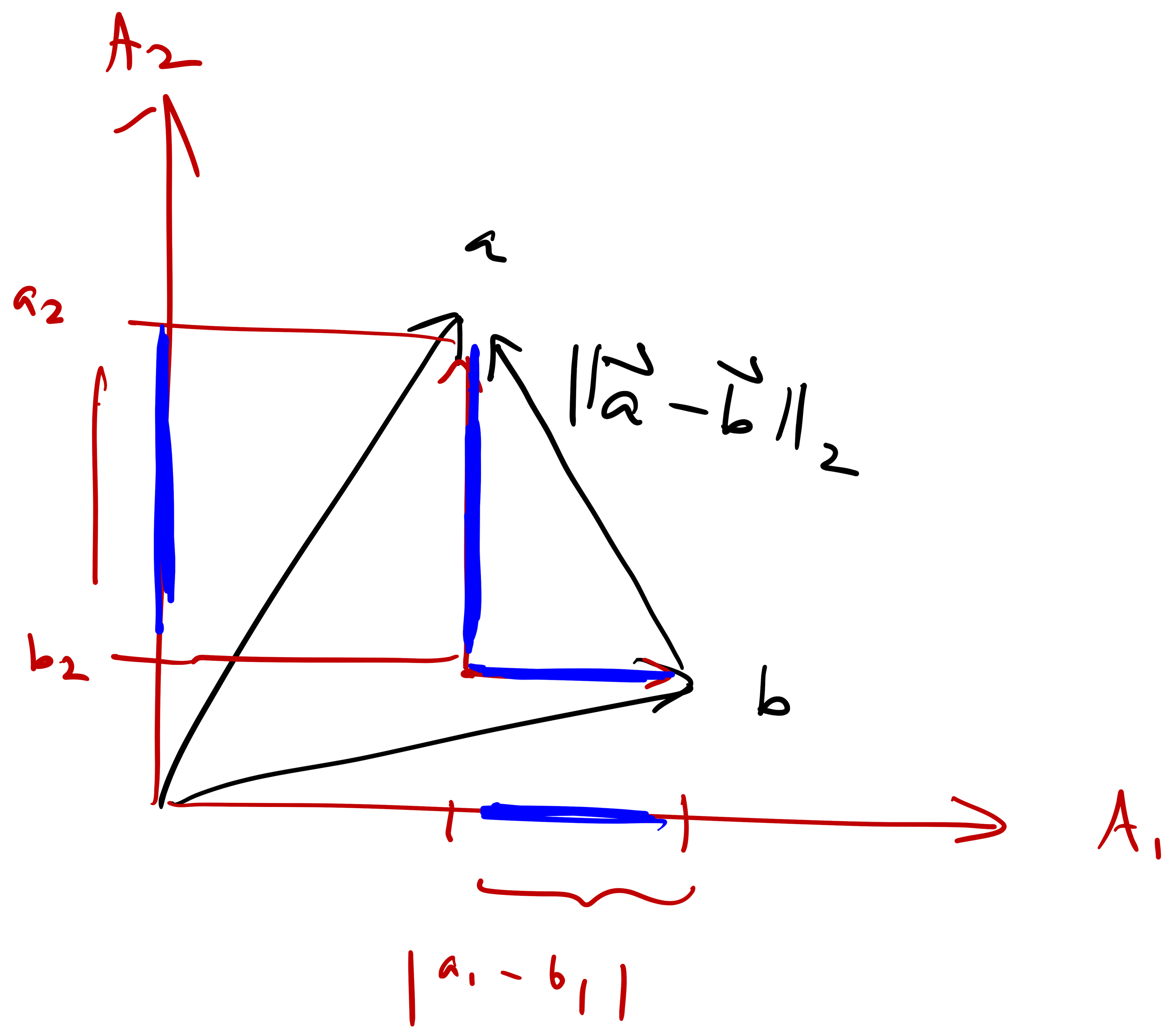
$$l_2 = \left(\sum_{i=1}^{n} a_i^2 \right)^{1/2} \longleftarrow \text{Euclidean length}$$

$$\underline{\underline{l_1}} = \sum_{i=1}^{n} |a_i| = |a_1| + |a_2| + \cdots + |a_n|$$

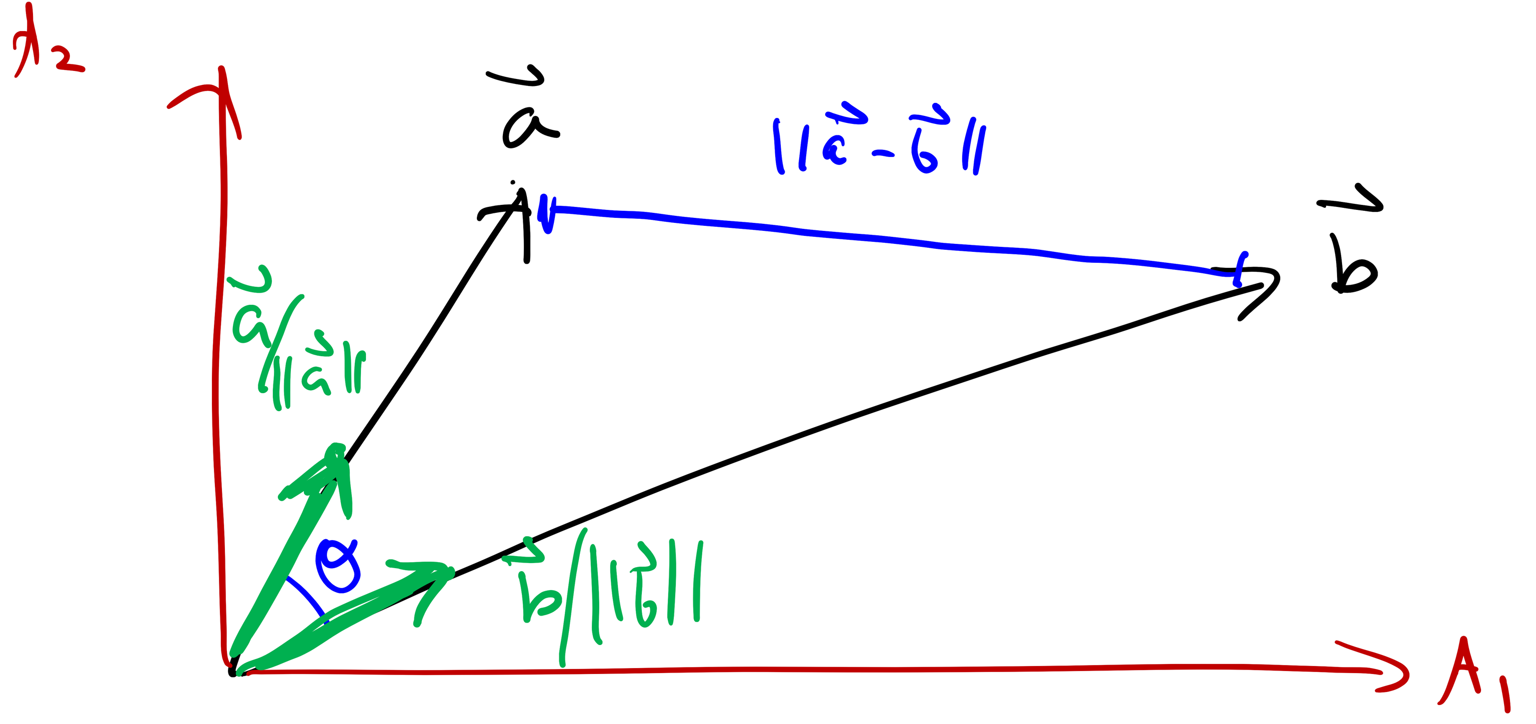
norm $\Longrightarrow$ distance

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

$$\|\vec{a} - \vec{b}\|_2 = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

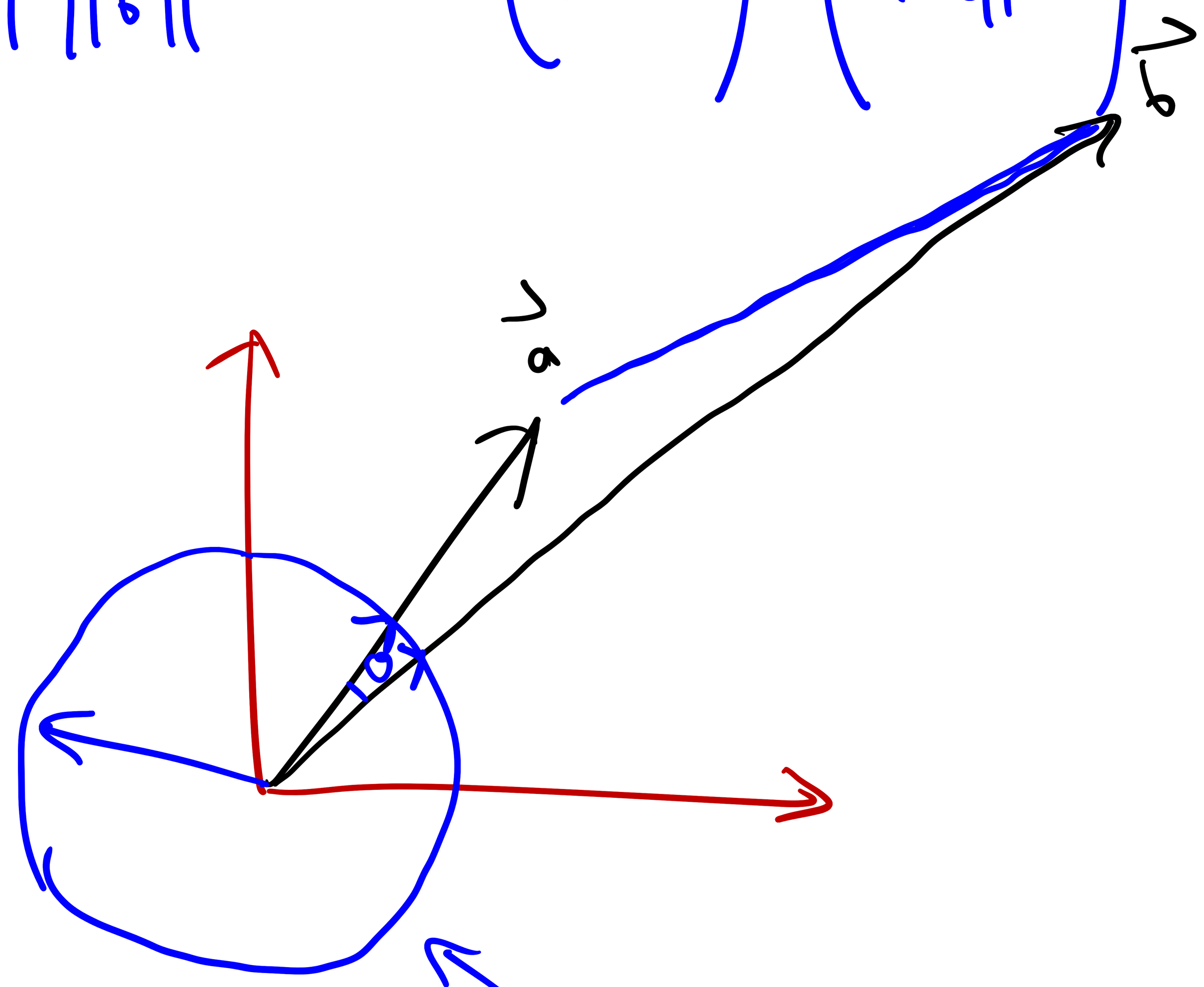


$$\|\vec{a} - \vec{b}\|_1 = |a_1 - b_1| + |a_2 - b_2|$$

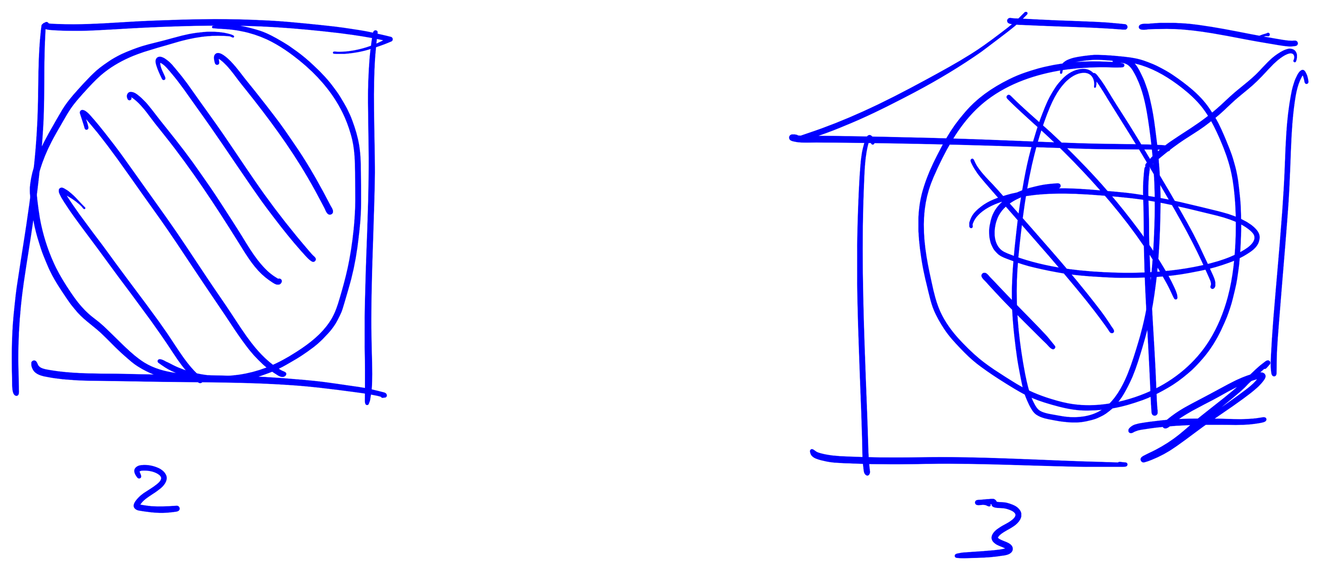


$$\cos\theta = \frac{\vec a^T \vec b}{\|\vec a\|\,\|\vec b\|} = \left(\frac{\vec a}{\|\vec a\|}\right)^T \left(\frac{\vec b}{\|\vec b\|}\right)$$

dot product between unit vectors

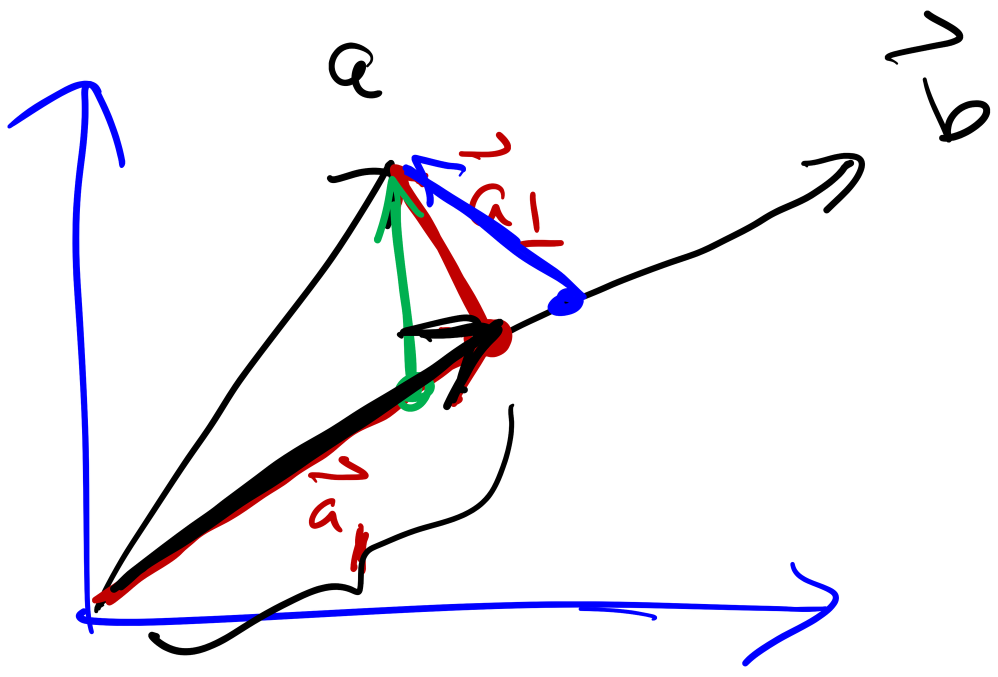


unit hyper-sphere



4 5 . . –

Projection of $\vec a$ onto $\vec b$



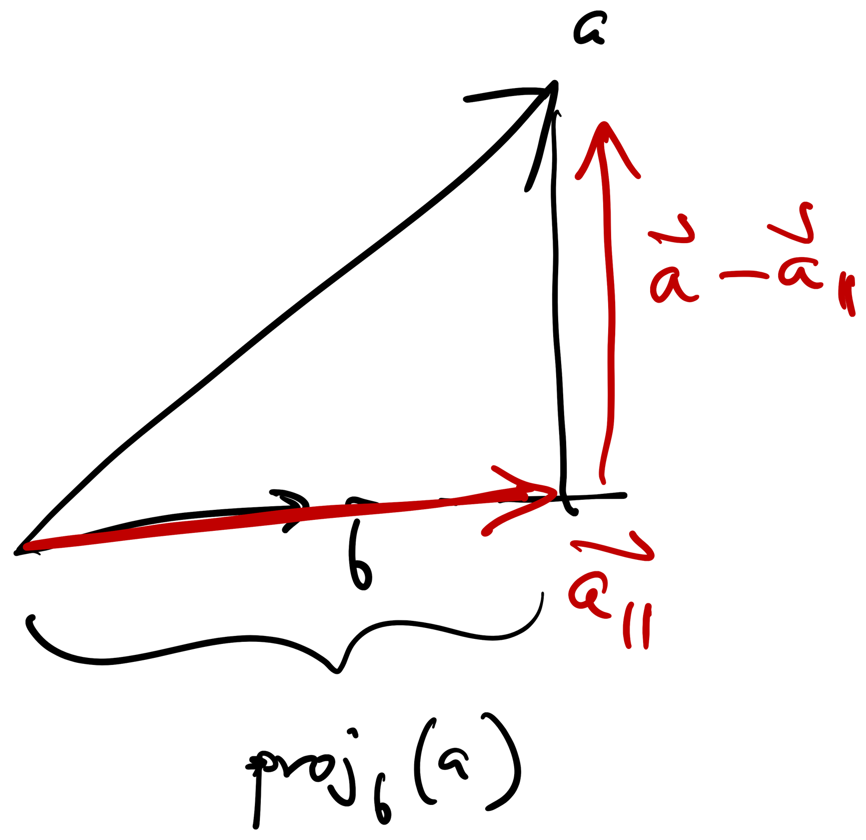
$\vec a_\parallel$ ← approximation to $\vec a$ in the $\vec b$ direction

$\|\vec a_\perp\|^2$ ← error

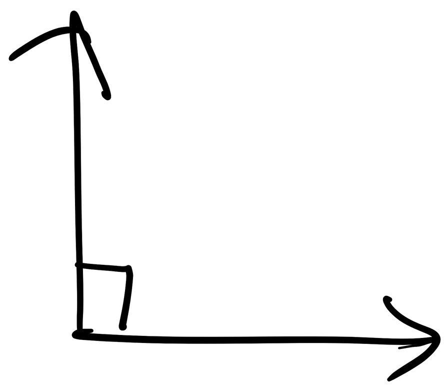
$$\vec a_\parallel = \underbrace{\mathrm{proj}_b(a)}_{\text{scalar offset}} \cdot \underbrace{\vec b}_{\text{vector direction}}$$

$$\vec{a}_{\parallel} = \underbrace{\left(\frac{a^T b}{b^T b}\right)}_{\mathrm{proj}_{\vec b}(\vec a)} \vec{b} \quad \leftarrow \text{direction}$$

Scalar

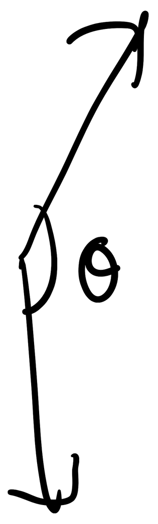


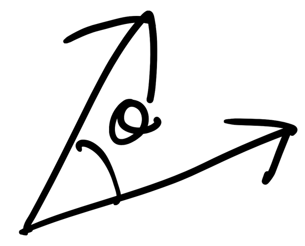
$$\text{error} = \vec a - \vec a_{\parallel}$$



Orthogonal means $a^T b = 0$

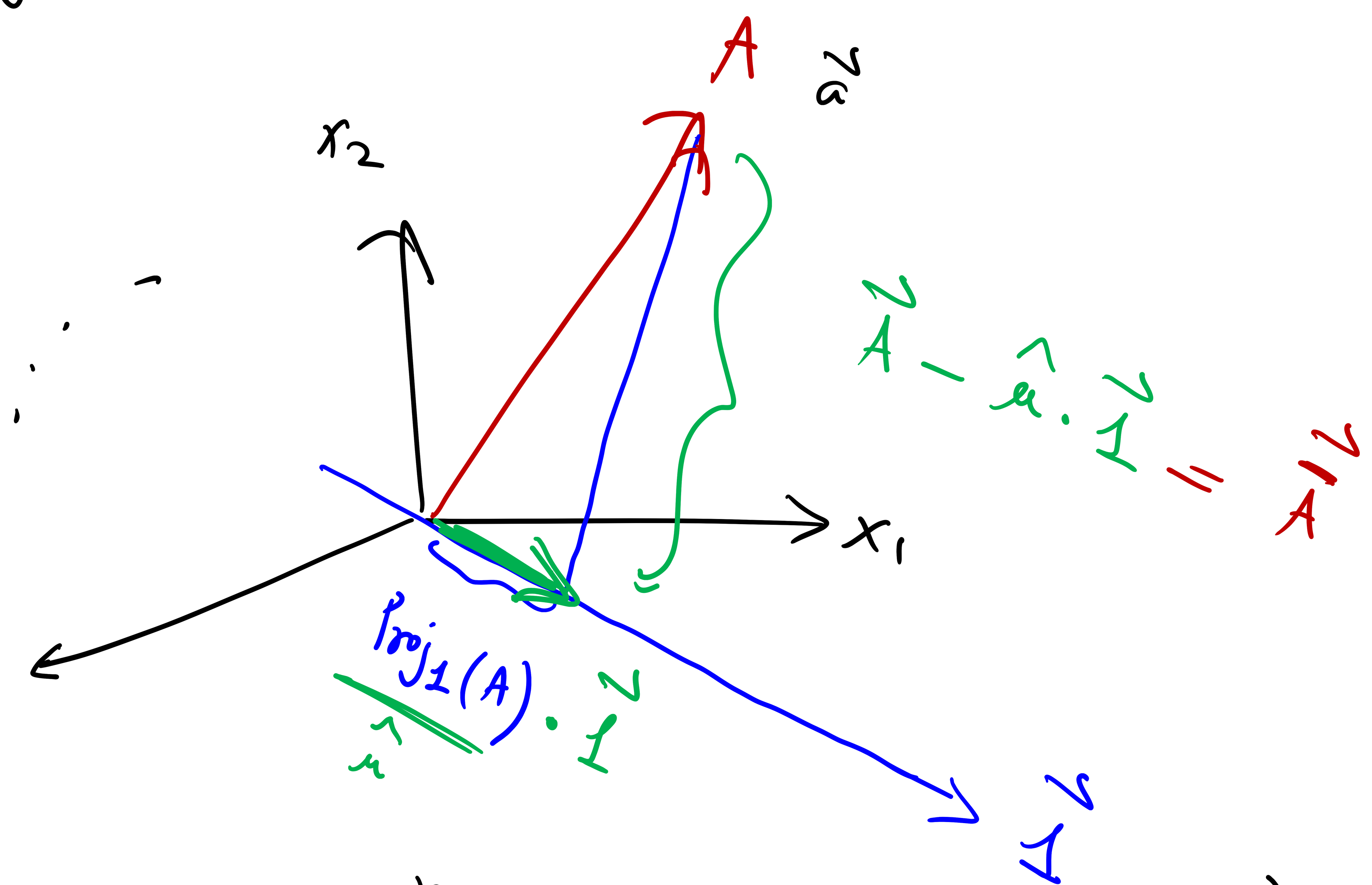
or $\cos\theta = 90^\circ$





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

feature vector $\vec A \in \mathbb{R}^n$



$\mathbb{R}^n$

$$\mathrm{Proj}_{\vec 1}(A) = \frac{A^T 1}{1^T 1}$$

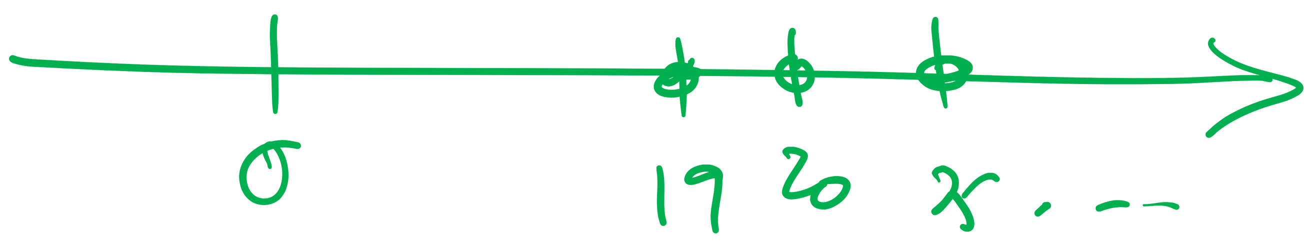
$$\vec A = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \vec 1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\hat\mu = \frac{x_1 + x_2 + \cdots + x_n}{1 + 1 + \cdots + 1} = \frac{\sum x_i}{n}$$

error: $\vec{A} - \hat{\mu}\cdot\vec{1}$ $\qquad \hat{\mu}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}$

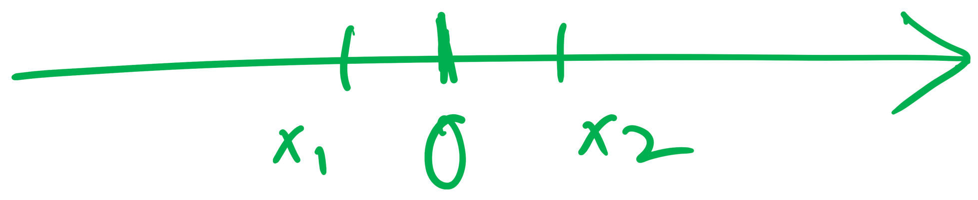
$$= \begin{pmatrix} x_1 - \hat{\mu} \\ x_2 - \hat{\mu} \\ \vdots \\ x_n - \hat{\mu} \end{pmatrix}$$

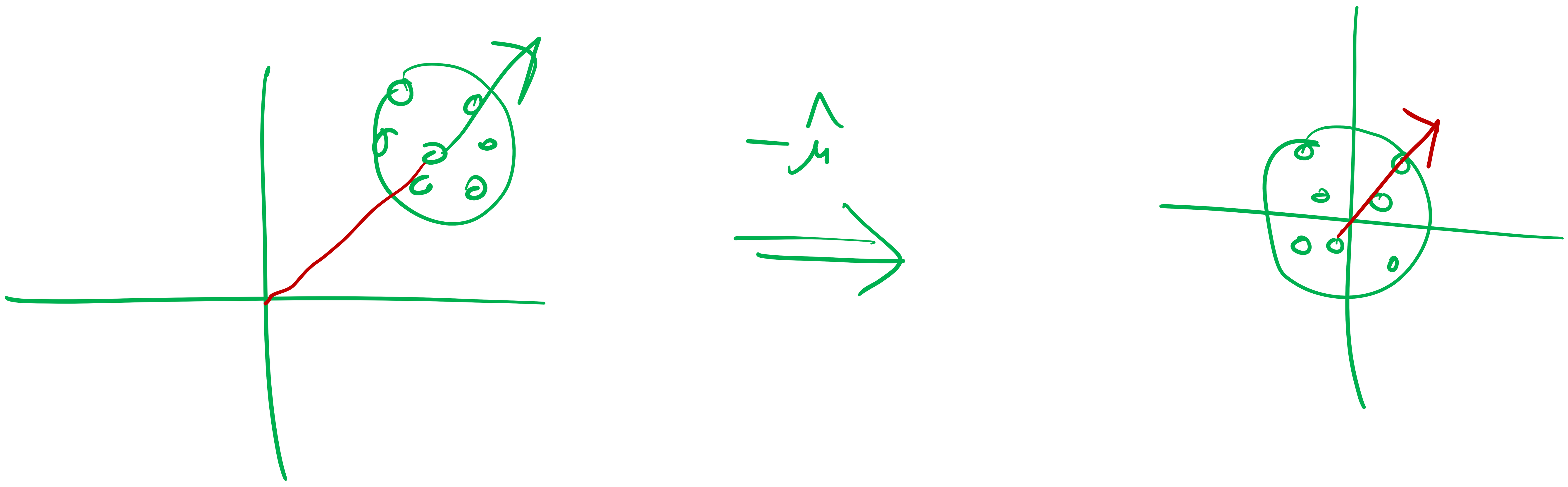
$= \bar{A} \leftarrow$ centered vector



Data centering!

$\Downarrow -\hat{\mu}$ translation





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

$$\text{Variance} = E\big[\underbrace{(A - \mu)^2}\big]$$

$$\int_x 2 -$$