LECTURE 6

SVD [Singular Vector Dimension] (efficiently)

$$A = \underbrace{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}_{n \times d} = \underset{u_1 \cdots u_d}{\begin{bmatrix} & U & \end{bmatrix}}_{n \times d} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{bmatrix}_{d \times d} \begin{bmatrix} & V & \end{bmatrix}_{d \times d}$$

U: orthonormal basis for columnspace(A)

V: orthonormal basis for rowspace(A)

$$= \sum_{i=1}^{d} \sigma_i u_i v_i^T$$

σ_i: singular values

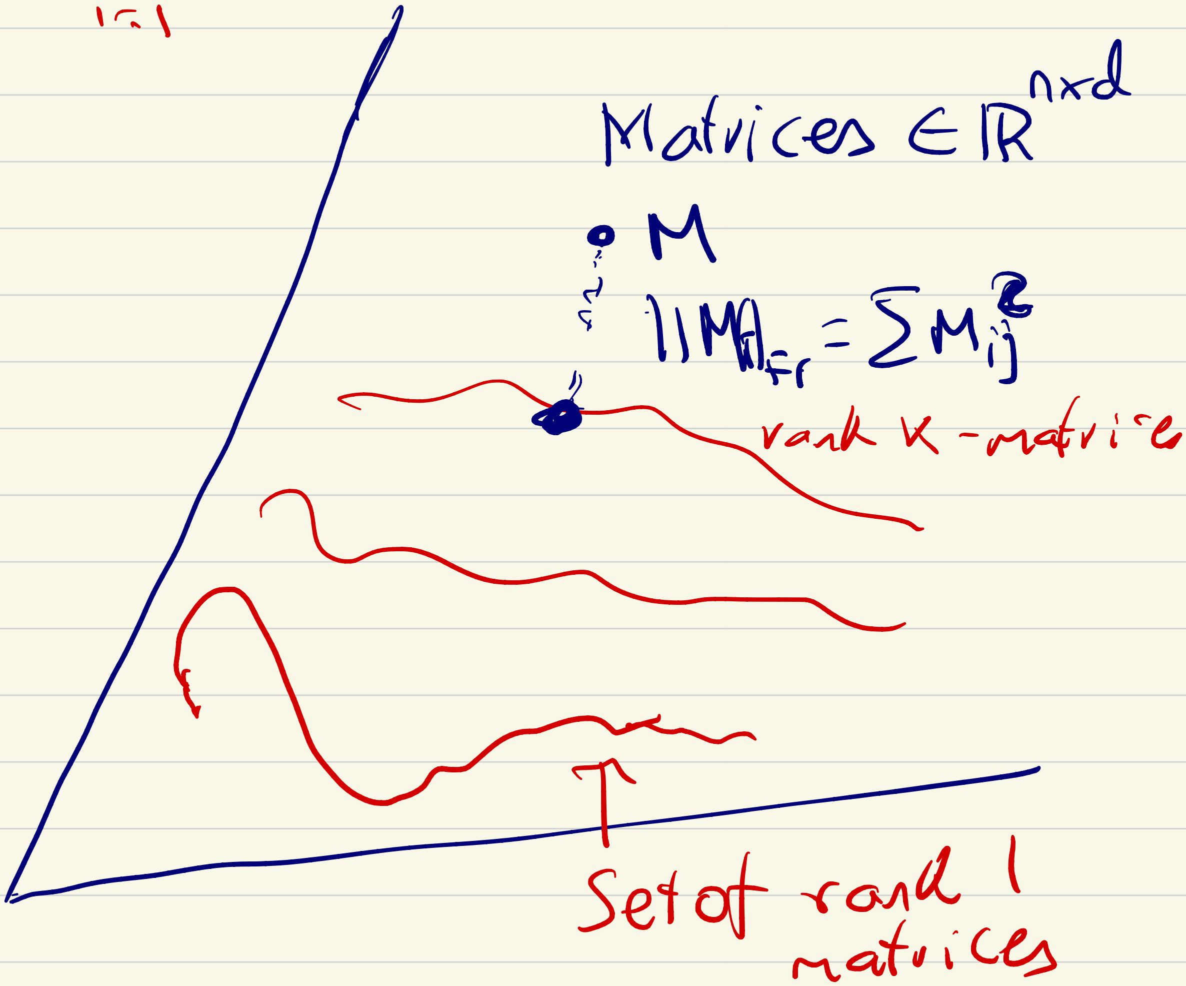
$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$

A - $n \times n$ matrix

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^T$$

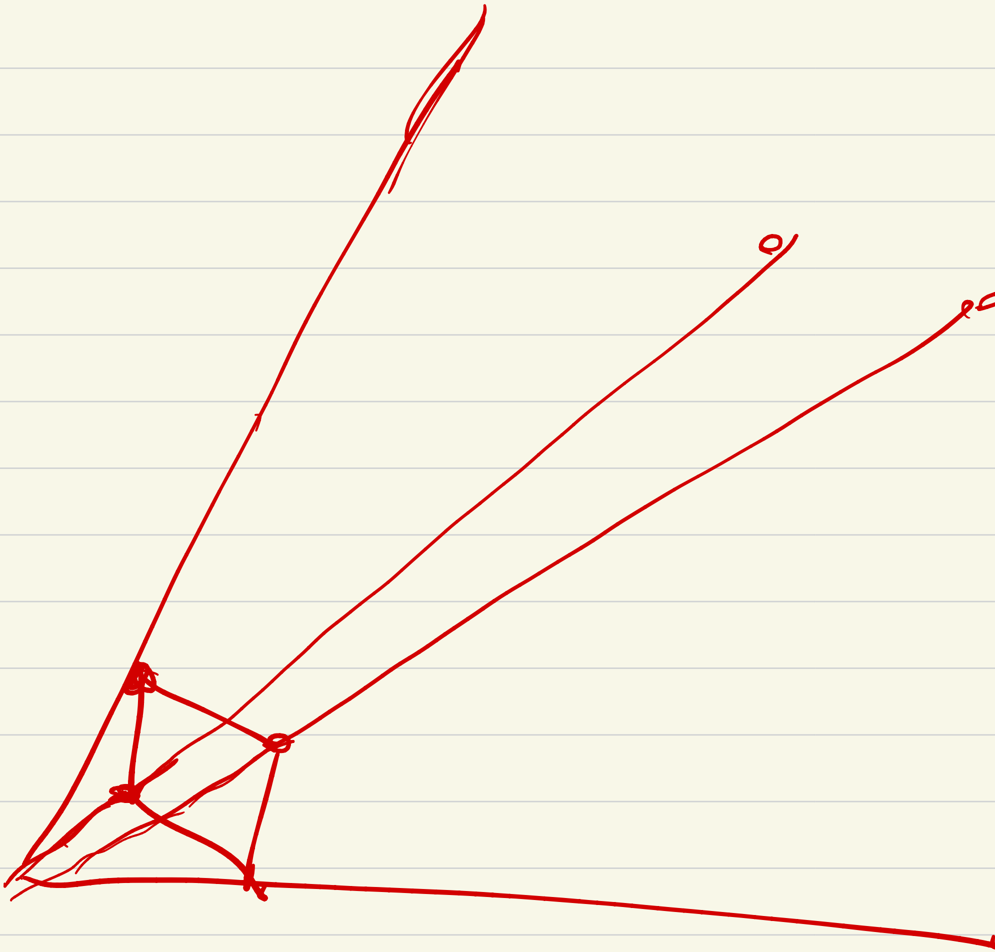
Best k-dimensional approximation A

$$= \sum_{i=1}^{k} \sigma_i u_i v_i^T$$



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

$$\sum_{i=1}^{k} \sigma_i u_i v_i^T = \min_{\substack{\text{rank } k \\ M}} \|M - A\|_{Fr}$$



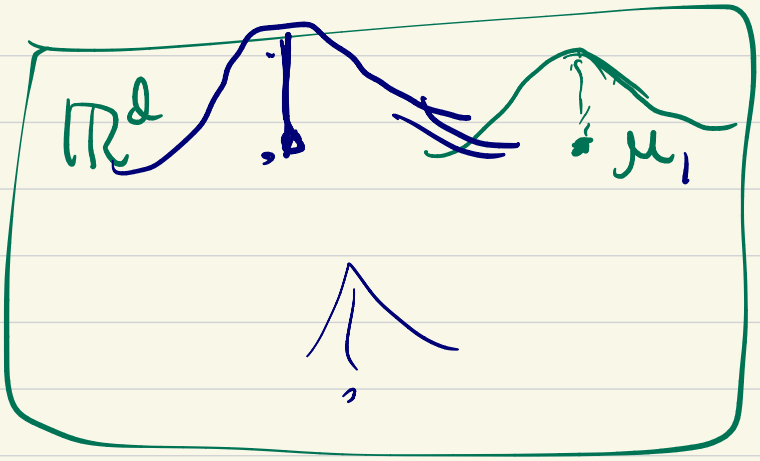
$$\boxed{\sum_{i=1}^{k} \sigma_i u_i v_i^T = \min_{\mathrm{rank}(M)=k} \|M - A\|_{op}}$$

LEARNING MIXTURES OF GAUSSIANS

Standard Gaussian : $N(\mu, Id) =$

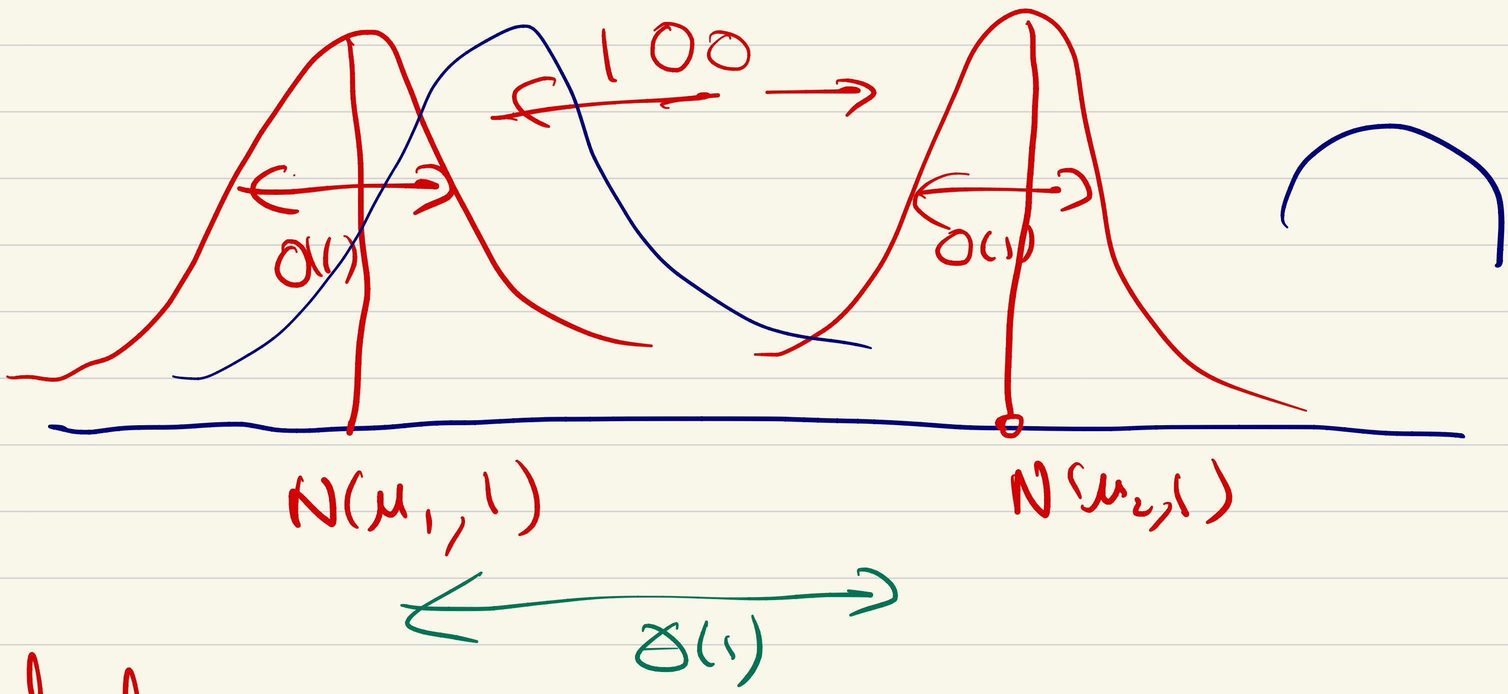
$$\phi(x) \propto e^{-\frac{(x-\mu)^T(x-\mu)}{2}}$$

$$= e^{-\frac{\|x-\mu\|^2}{2}}$$

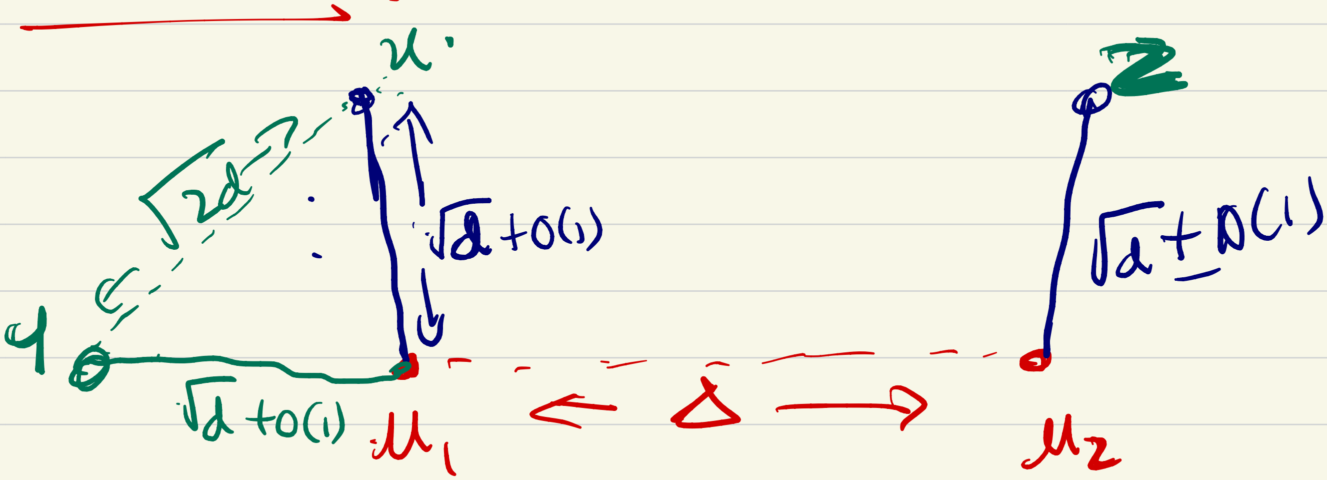


$$\underline{\underline{\theta}} = \frac{1}{2} N(\mu_1, \underline{\underline{Id}}) + \frac{1}{2} N(\mu_2, Id)$$

PROBLEM: 1) Given samples from θ find μ_1 and μ_2



d-dimension:



$$\|x-y\|^2 = (\sqrt{d})^2 + (\sqrt{d})^2 = 2d \pm \sqrt{d}$$

$$\|x-z\|^2 = \|x-\mu_1+\mu_1-\mu_2+\mu_2-z\|^2$$

$$= (\sqrt{d})^2 + \Delta^2 + (\sqrt{d})^2$$

$$= 2d + \Delta^2$$

$$\text{intracluster} = 2d \pm \boxed{\sqrt{d}}$$

$$\wedge\wedge$$

$$\text{intercluster} = 2d + \boxed{\underline{\Delta^2}} \pm \sqrt{d}$$

$$\Rightarrow \boxed{\Delta \gtrsim d^{1/4}}$$

$\underline{\text{PCA}}$: Samples $S \leftarrow \frac{1}{2} N(\underline{\mu_1}, I_d) + \frac{1}{2} N(\underline{\mu_2}, I_d)$

2-dimensional

$\rightarrow$ PCA of S

Find a subspace Π of 2 dimensions

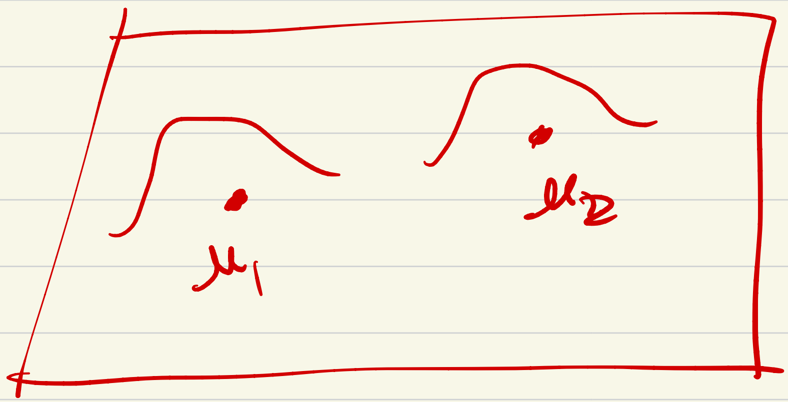
$$\max_{\Pi} \mathop{E}_{x \sim S} \left[\|\Pi x\|^2 \right]$$

$$= \max_{\Pi} \frac{1}{2} \mathop{E}_{x \sim N(\mu_1, I_d)} \left[\|\Pi x\|^2 \right] + \frac{1}{2} \mathop{E}_{x \sim N(\mu_2, I_d)} \left[\|\Pi x\|^2 \right]$$

Claim: $\Pi \in \mathrm{Span}(\underline{\mu_1}, \underline{\mu_2})$

Excercise:

· Project dataset on to Π



2-dim

$$\Pi x \mid x \sim \tfrac{1}{2} N(\mu_1, I_d) + \tfrac{1}{2} N(\mu_2, I_d)$$

$$\wr$$

$$\tfrac{1}{2} N(\mu_1, I_{d_\Pi}) + \tfrac{1}{2} N(\mu_2, I_{d_\Pi})$$

If $\|\mu_1 - \mu_2\| \geq C$ for large C,

then find μ_1, μ_2

$$\underline{\underline{N(0, I_d)}} = (g_1 \cdots g_d)$$

$$\Downarrow$$

$$E[g_i^2] = 1$$

$$\underline{\underline{\sum g_i^2 = d}}$$

$$\sum_{i=1}^{d} g_i^2 =$$ sum d independent variables

with mean $\overline{1}$
$E[g_i^2] = 1$

$$\simeq d + \underline{\underline{\Theta(\sqrt{d})}}$$

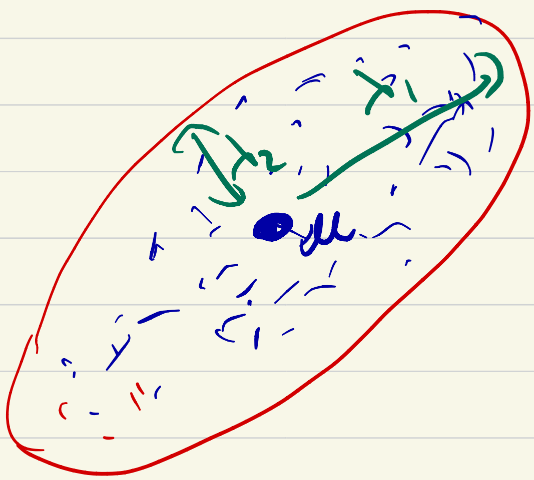
ROBUST MEAN ESTIMATION

$\mathcal{D}$: 1) $\mu(\mathcal{D}) = \mathop{\mathbb{E}}_{x \sim \mu}[x]$

$X_1 \dots X_n \sim D \Leftrightarrow \hat{\mu} = \frac{1}{n} \sum X_i$

2) $\mathrm{Cov}[\mathcal{D}] = \mathop{\mathbb{E}}_{x}\left[(x-\mu)(x-\mu)^T\right]$

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

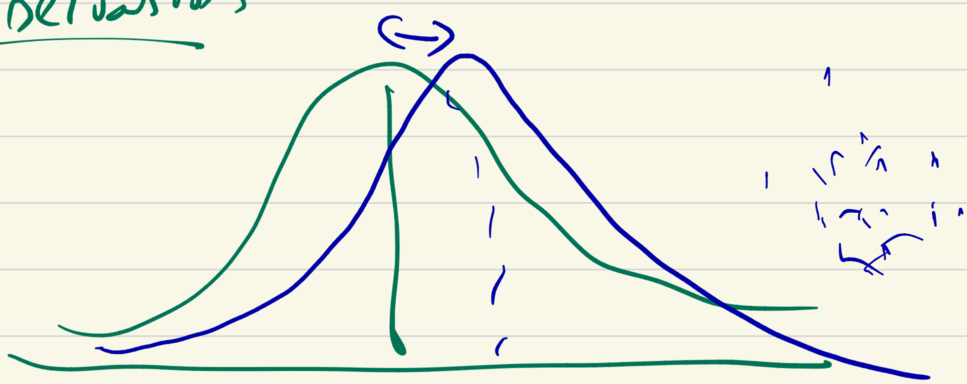


Robust Mean Estimation $\quad \mathcal{D} \approx N(\mu, I_d)$

(inliers)	$(1-\varepsilon)n$ samples from $\mathcal{D}$	Good / G
(outliers)	εn samples are adversarially	Bad / B

Estimate $\mu(\mathcal{D})$

Observations



$\mu \quad N(\mu-\varepsilon, 1)$

$O(\varepsilon)$ error is needed in I_d.

$N(\mu, 1) \longrightarrow N(\mu-\varepsilon, 1)$

Median in 1d is within $O(\varepsilon)$ of true mean

- d-dimensions:

  - Geometric median
    - median in each co-ordinate

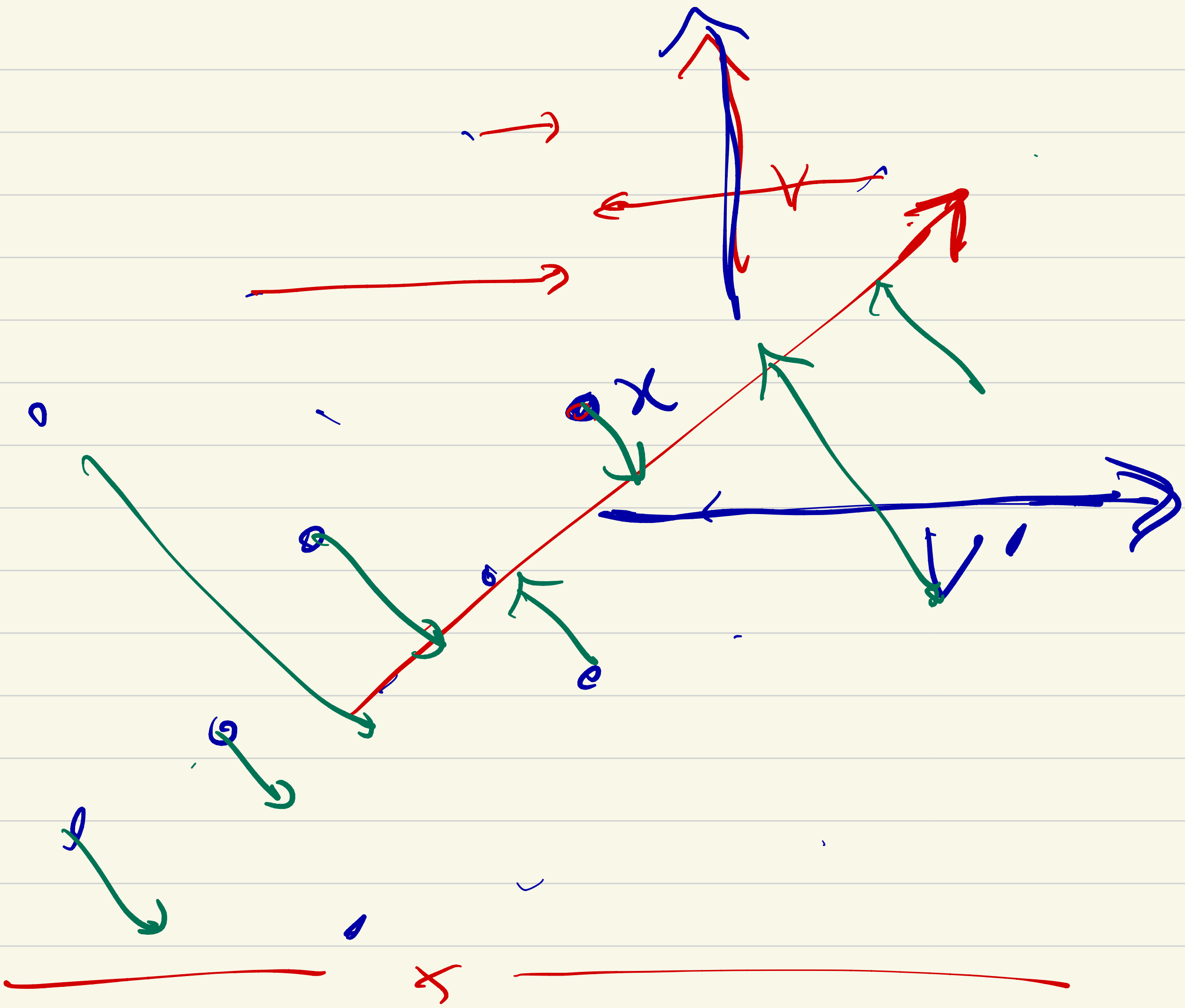
$$\|\hat{\mu} - \underset{\text{true mean}}{\mu}\|^2 \simeq \varepsilon^2 d$$

$$\underline{\|\hat{\mu} - \mu\|} \approx \varepsilon \underline{\sqrt{d}}$$

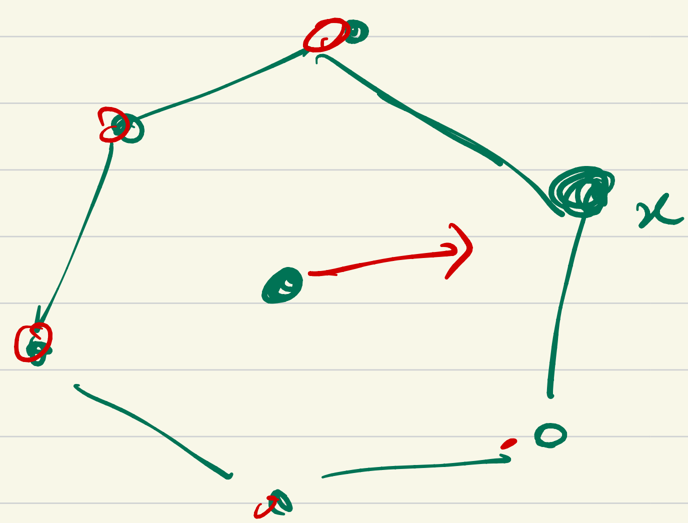
- Tukey median → no efficient algorithms

$$\text{depth}(x) = \min_{v} \underset{\text{in direction } v}{\text{position}(x)}$$

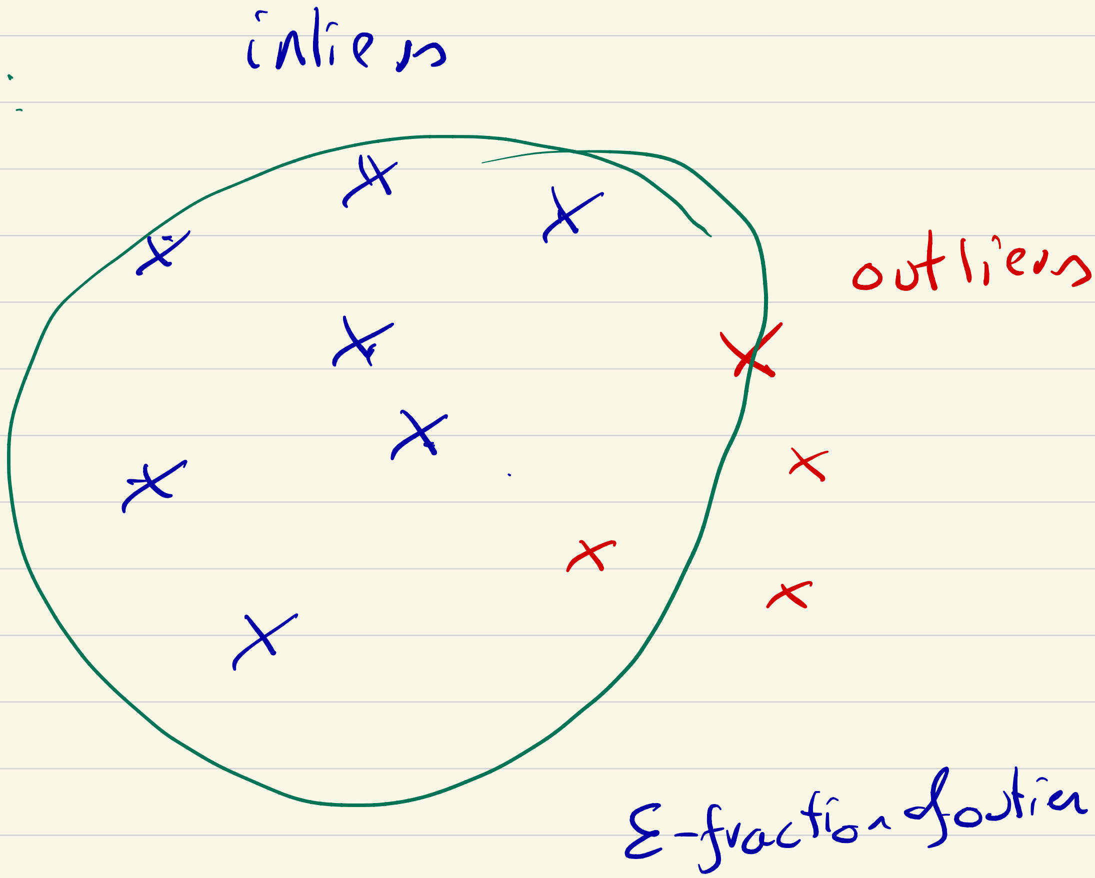
$$x^* = \arg\max \text{depth}(x)$$

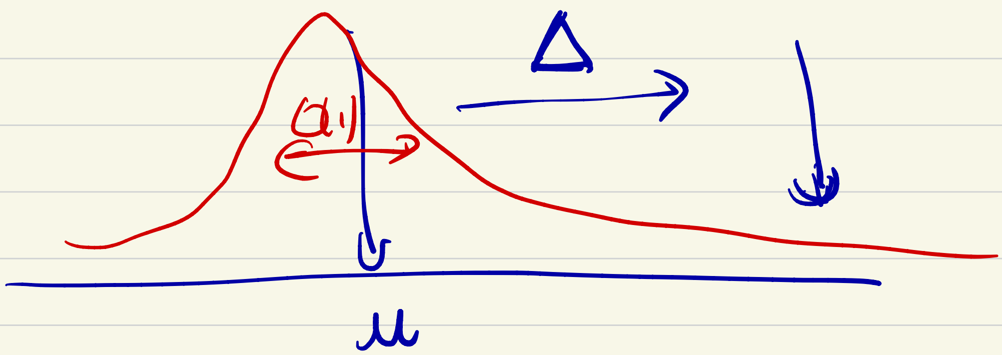


$\langle v, x_1 \rangle \;,\; \ldots \;\; \langle v, x_n \rangle$



Intuition:

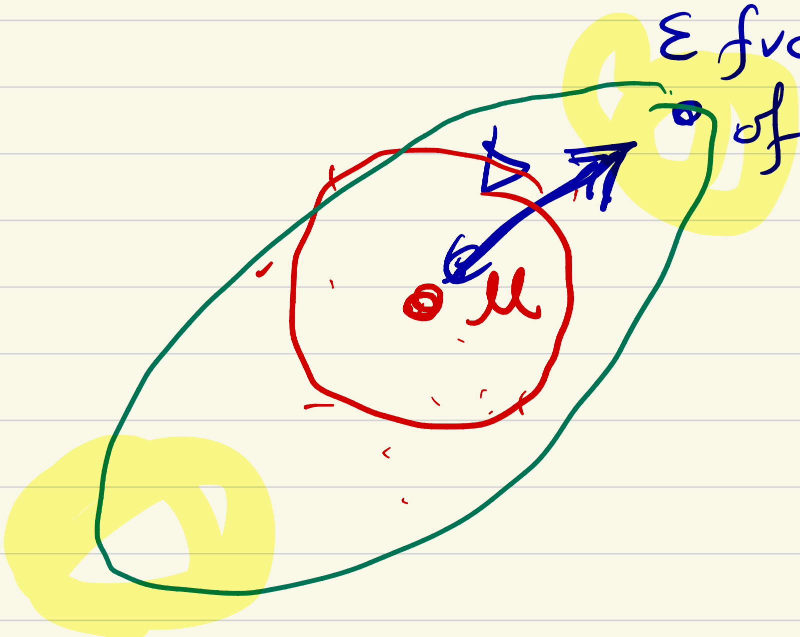




To shift the mean by Δ

$\Rightarrow$ add points $\underline{\underline{\dfrac{\Delta}{\varepsilon}}}$ away

$\Rightarrow$ Variance: $\varepsilon \cdot \left(\dfrac{\Delta}{\varepsilon}\right)^2 = \underline{\underline{\dfrac{\Delta^2}{\varepsilon}}}$ to the variance



ε fraction of points at distance Δ/ε

$x_1 \ldots x_n$ from a distribution D on $\mathbb{R}$

estimate mean of D

$$\hat{\mu} = \frac{1}{n} \sum_i x_i$$

$\mathcal{D}$ is Gaussian

$\forall \varepsilon, \delta$

$$\boxed{\begin{aligned} &\Pr\left[\,|\hat{\mu} - \mu| > \varepsilon\,\right] \leq \delta = e^{-\varepsilon^2 n/2} \\ &\text{if} \quad n = \boxed{\frac{2 \log 1/\delta}{\varepsilon^2}} \end{aligned}}$$

Power law distribution:

$$\phi(x) \propto x^{-c}$$

for some c

Empirical mean does not give exponential convergence

Median of Means $\left(\quad \vdots \qquad\qquad \vdots \quad \right)$

Median of Means [80s]

1) Split dataset into k buckets

let 2) $\mu_i = \text{Mean}(\text{Bucket } i)$

3) $\hat{\mu} \leftarrow \underset{\text{1 dim.}}{\overset{\text{Tukey}}{\text{Median}}}(\mu_1 \ldots \mu_k)$

$\Longrightarrow$ For confidence δ, accuracy ε,

$$n = \boxed{O}\left(\frac{\log 1/\delta}{\varepsilon^2}\right) \quad \boxed{k \approx \log 1/\delta}$$

Sept [Valiant et. al] an algorithm to compute 1 dimensional mean

$$n = 2\frac{\log 1/\delta}{\varepsilon^2}(1+o(1))$$

Compute d- dimensional mean

with same guarantee as Gaussian

up to const factors

$$n \approx O\left(\frac{\log 1/\delta}{\varepsilon^2}\right)$$

Heavy tailed Statistics