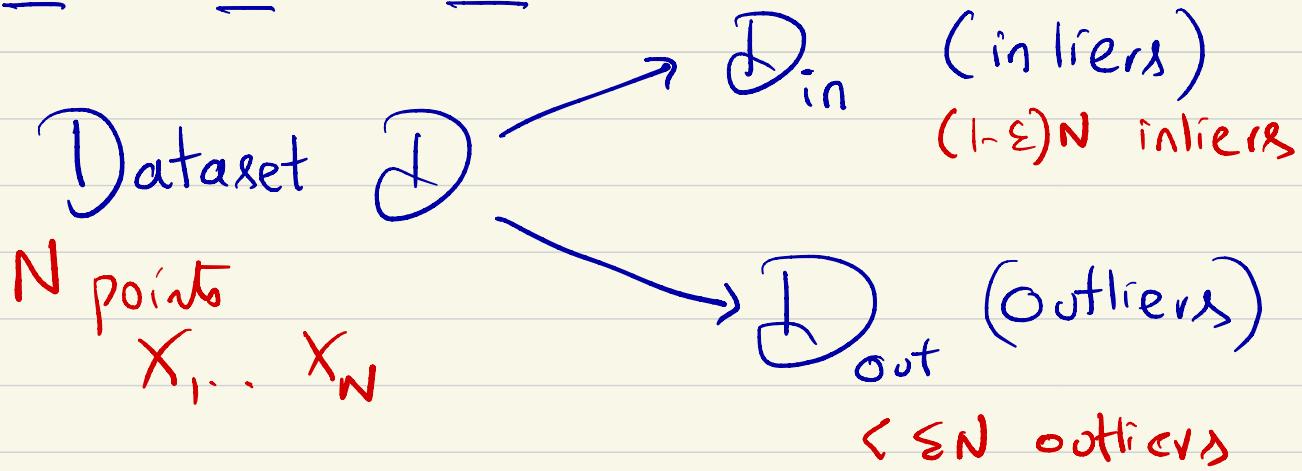


LECTURE 7

Robust Mean Estimation:



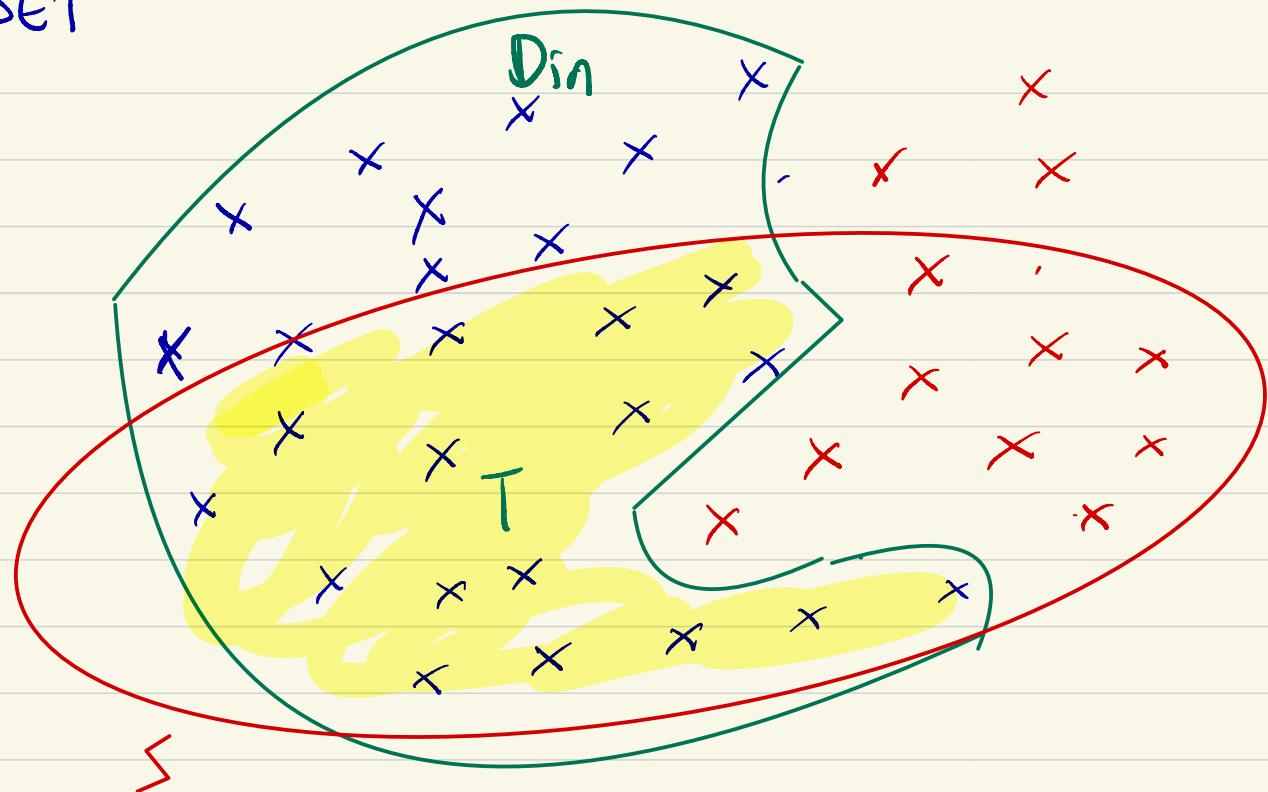
GOAL: Estimate Mean

Defn: A dataset $\mathcal{D} = \{x_1, \dots, x_N\}$ is (ε, Δ) stable if [mean doesn't change on removing εN elements]

$$\forall S \subseteq \mathcal{D} \quad |S| \geq (1 - \varepsilon)N$$

$$\left\| \frac{1}{|S|} \sum_{i \in S} x_i - \frac{1}{|\mathcal{D}|} \sum_{i \in \mathcal{D}} x_i \right\| \leq \Delta$$

DATASET



Assume: D_{in} inliers are (ε, Δ) -stable

Will do: Find subset S which is (ε, Δ) -stable
 $|S| = (1-\varepsilon)N$.

Claim: \Downarrow

$$\|\mu(S) - \mu(D_{in})\| \leq 2\Delta$$

Proof: $T = S \cap D_{in}$

By (ε, Δ) stability of D_{in} : $\|\mu(D_{in}) - \mu(T)\| \leq \Delta$

By (ε, Δ) stability of S : $\|\mu(S) - \mu(T)\| \leq \Delta$

Def: An ε -filtering of a set S

↓

is any set T , $|T| \geq (1-\varepsilon)|S|$

Def: A distribution Θ_{in} is an ε -filtering
of a distribution Θ over \mathbb{R}^n

$$\text{if } \forall x \quad \Theta_{in}(x) \leq \Theta(x) \cdot (1+\varepsilon)$$

Def: A distribution Θ is (ε, Δ) -stable
if $\forall \varepsilon$ -filtering Θ_{in}

$$\|\mu(\Theta) - \mu(\Theta_{in})\| \leq \Delta.$$

LEMMA: A distribution Θ over \mathbb{R}^n

is (ε, Δ) stable for $\Delta = \sqrt{\varepsilon \cdot \|\text{Cov}(\Theta)\|_{\text{op}}}$

- largest eigenvalue

PROOF: Suppose Θ_{in} is an ε -filtering of Θ

To prove $\|\mu(\Theta_{in}) - \mu(\Theta)\| \leq \Delta$.

Claim: $\Theta = (1-\gamma) \cdot \Theta_{in} + \gamma \cdot \Theta_{out}$
for $\gamma \stackrel{\text{def}}{=} \frac{\varepsilon}{1+\varepsilon}$

$$\Theta_{out}(x) \stackrel{\text{def}}{=} \frac{(1+\varepsilon)\Theta(x) - \Theta_{in}(x)}{\varepsilon}$$

then Θ_{out} is a distribution & $\Theta = (1-\gamma)\Theta_{in} + \gamma\Theta_{out}$

Claim: Suppose $\Theta = (1-\gamma) \Theta_{in} + \gamma \cdot \Theta_{out}$

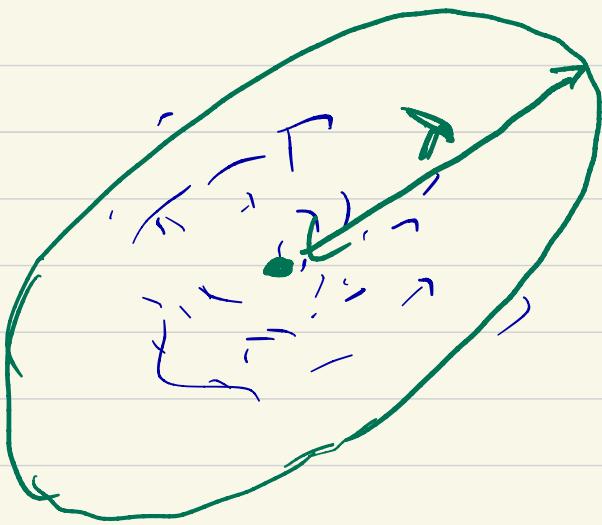
$$\left\{ \begin{array}{l} 1) \underline{\mu(\Theta)} = (1-\gamma) \mu(\Theta_{in}) + \gamma \cdot \mu(\Theta_{out}) \\ 2) \text{Cov}(\Theta) = (1-\gamma)^2 \text{Cov}[\Theta_{in}] + \gamma^2 \cdot \text{Cov}[\Theta_{out}] \\ \quad + 2\gamma(1-\gamma) \left[\frac{(\mu[\Theta_{in}] - \mu[\Theta_{out}])(\mu[\Theta_{in}] - \mu[\Theta_{out}])}{\sqrt{(\mu[\Theta_{in}] - \mu[\Theta_{out}])}} \right] \end{array} \right.$$

In Eq[2], apply $v^T v$ where $v = \frac{\mu[\Theta_{in}] - \mu[\Theta_{out}]}{\sqrt{(\mu[\Theta_{in}] - \mu[\Theta_{out}])}}$

$$\text{LHS: } v^T \text{Cov}(\Theta) v \leq \|\text{Cov}(\Theta)\|_{op} \rightarrow (1)$$

$$\begin{aligned} \text{RHS: } & v^T \cancel{(1-\gamma^2) \text{Cov} v} + \cancel{v^T \gamma^2 \text{Cov} v} + \\ & + 2\gamma(1-\gamma) \|\mu[\Theta_{in}] - \mu[\Theta_{out}]\|^2 \end{aligned}$$

$$\geq 2\gamma(1-\gamma) \|\mu[\Theta_{in}] - \mu[\Theta_{out}]\|^2 \rightarrow (2)$$



$$\|\underline{\mu}(\underline{\theta}_{in}) - \underline{\mu}(\underline{\theta}_{out})\| \leq \sqrt{\frac{\|\text{Cov}(\underline{\theta})\|_{op}}{2\gamma(1-\gamma)}}$$

$$\|\underline{\mu}(\underline{\theta}) - \underline{\mu}(\underline{\theta}_{in})\|$$

$$= \|\gamma(\underline{\mu}(\underline{\theta}_{in}) - \underline{\mu}(\underline{\theta}_{out}))\|$$

$$= \gamma \cdot \sqrt{\frac{\|\text{Cov}(\underline{\theta})\|_{op}}{2\gamma(1-\gamma)}}$$

$$= \boxed{\sqrt{\epsilon \|\text{Cov}(\underline{\theta})\|_{op}}}$$

Dataset:

$$\mathcal{D} = \{x_1, \dots, x_N\}$$

Goal: Find w_1, \dots, w_N

$$\left\{ \begin{array}{l} w_i = \frac{1}{N} \\ \text{---} \end{array} \right.$$

$\left\{ \begin{array}{l} -0 \leq w_i \leq \frac{(1+\varepsilon)}{N} \\ \text{---} \end{array} \right. \quad \left(\begin{array}{l} w \text{ to be an } \varepsilon\text{-filtering} \\ \text{of } \mathcal{D} \end{array} \right)$

$\left\{ \begin{array}{l} \sum w_i = 1 \\ \text{---} \end{array} \right.$

$\left\{ \begin{array}{l} \text{Cov}[w] < \lambda \cdot \text{Id} \\ \text{---} \end{array} \right.$

$$\sum_{i=1}^N w_i (x_i - \mu(w)) (x_i - \mu(w))^T \leq \lambda \cdot \text{Id}$$

$\downarrow \quad \downarrow$

$\sum w_i x_i$

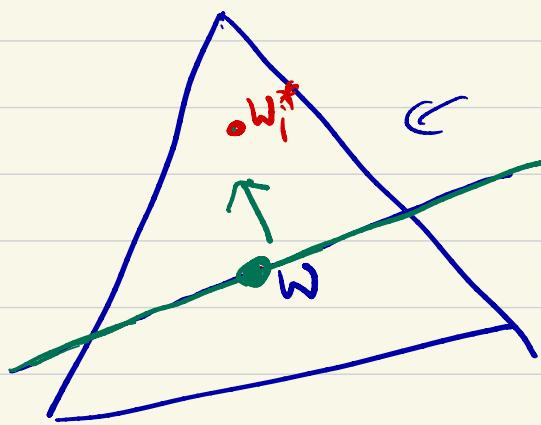
$$x \quad \text{---}$$

Inliers: $w_i^* = \begin{cases} 1/|D_{in}| & \text{if } i \in D_{in} \\ 0 & \text{else} \end{cases}$

w^* satisfies all constraints here.

Ellipsoid Algorithm

Find $\{\omega_i\}$ using ellipsoid.



$$a) \sum \omega_i = 1$$

$$b) \omega_i \leq (1+\varepsilon)/N$$

Given ω :

$$\text{Cov}(\omega)$$

$$\left\| \sum \omega_i [x_i - \mu(\omega)] [x_i - \mu(\omega)]^T \right\|_{op} = \lambda$$

let $\|\text{Cov}(\omega)\|_{op} = \lambda$ and let v be top

eigenvector

$$\boxed{v^T \text{Cov}(\omega) v = \lambda}$$

Fixed x_i, ω, v ,

$$L(y) = v^T \left[\sum y_i [x_i - \mu(\omega)] [x_i - \mu(\omega)]^T \right] v$$

For some λ

$$\rightarrow \boxed{L[\omega] \stackrel{\text{def}}{=} \lambda}$$

Show:

$$\rightarrow \boxed{L[\omega^*] < O(\epsilon\lambda)}$$

then

$$\boxed{L[y] \leq \lambda}$$

is a hyperplane.

Lemma:

$$L[\omega^*] = \sqrt{\sum_i w_i^* [x_i - \underline{\mu}(\omega)] [x_i - \underline{\mu}(\omega)]^T}$$

$$= \sqrt{\left[\sum_i w_i (x_i - \underline{\mu}[\omega^*]) \cdot (x_i - \underline{\mu}[\omega^*])^T \right]}$$

$$+ \sqrt{\left[(\underline{\mu}[\omega^*] - \underline{\mu}[\omega]) \cdot (\underline{\mu}[\omega^*] - \underline{\mu}[\omega])^T \right]}$$

$$\leq \sqrt{\text{Cov}[\omega^*]} + \|\underline{\mu}[\omega^*] - \underline{\mu}[\omega]\|^2$$

in red

$$\leq C_1 +$$

$$\|\mu(\omega^*) - \mu(\omega)\|^2$$

↓ ↓
 filter candidate

$$\begin{aligned}
 &\leq 2 \|\mu(\omega^*) - \mu(\omega \cap D_{in})\|^2 \\
 &\quad + 2 \|\mu(\omega \cap D_{in}) - \mu(\omega)\|^2 \\
 &\quad \quad \quad \overbrace{\quad \quad \quad}^{\pi}
 \end{aligned}$$

$$\lesssim O(\varepsilon\lambda) + O(\varepsilon\lambda) = O(\varepsilon\lambda).$$

Note 1) $\omega \cap D_{in}$ is an $O(\varepsilon)$ filtering
of ω^*

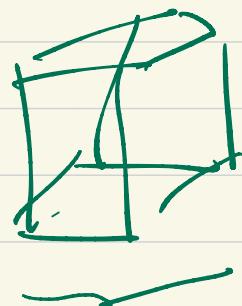
2) $\omega \cap D_{in}$ is an $O(\varepsilon)$ filtering
of ω

TENSORS

$T \in$ higher dimensional array of numbers
3-dimensional tensor / 3 modes

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

$$\mathbb{R}^{m \times n \times p}$$



Matrix M :

$$\rightarrow M(x, y) = \sum M_{ij} \cdot x_i y_j$$

↑
bilinear form

$$M: [\text{vector}] \times [\text{vector}] \rightarrow \mathbb{R}$$

$$\rightarrow M: [\text{vector}] \rightarrow [\text{vector}]$$

↓
linear
transformation

T_3 $(\text{vector}) \times (\text{vector}) \times (\text{vector}) \rightarrow \mathbb{R}$

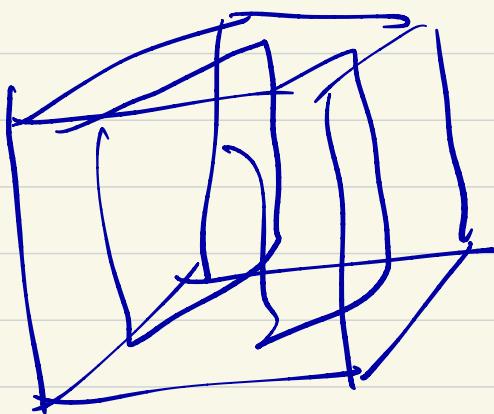
$$T[x, y, z] = \sum_{i,j,k} T_{ijk} x_i y_j z_k$$

x_1, y_1, z_1
 \vdots
 x_n, y_n, z_n

$[\text{vector}] \times [\text{vector}] \rightarrow [\text{vector}]$

$$\hat{T}[x, y] = \begin{bmatrix} \sum T_{1ij} x_i y_j \\ \sum T_{2ij} x_i y_j \\ \vdots \end{bmatrix}$$

$[\text{vector}] \rightarrow [\text{vector}] \times [\text{vector}]$



$$x \sim \mathbb{R}^n \quad D$$

Moments: $\mu = \underset{x \sim D}{E}(x)$

$$\text{Cov} = \underline{\underline{E}}(x - \mu)(x - \mu)^T$$

$$T_{ijk} = E[(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)]$$

$$E[x_i x_j x_k]$$

Tensors \in used to encode higher order correlations of random variables.