Lecture 2: Minimum Distance Functionals

Recap:
- Conceptual diagram
- Truncated mean

This time:
- Issues in high dimensions
- Resolving the issues
  - Minimum distance functional
  - Resilience

$$\begin{array}{ccc} \text{test} & & \text{train} \\ p^* \in \mathcal{G} & \xrightarrow[\text{TV}]{D(p^*, \tilde{p}) \le \epsilon} & \tilde{p} \\ \downarrow & & \downarrow \ \text{"population level"} \\ L(p^*, \hat{\theta}) & \longleftarrow & \hat{\theta}(\tilde{p}) \end{array}$$

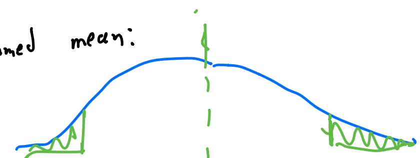
$$\cancel{\begin{array}{c} \downarrow \\ X_1, \dots, X_n \\ \downarrow \\ \hat{\theta}(X_1, \dots, X_n) \end{array}}$$

mean estimation:

$$L(p, \theta) = \|\theta - \mu(p)\|_2$$

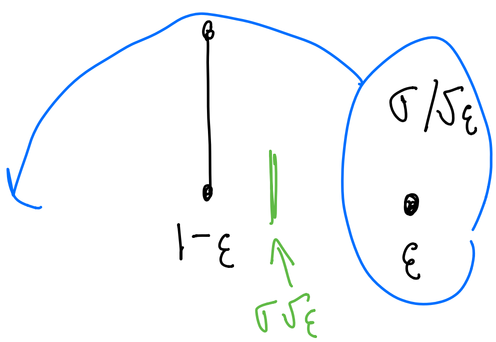
ℓ_2-norm

Trimmed mean:

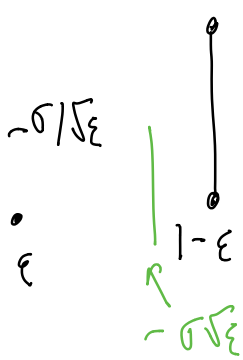


1-dimensional

Prop. If $p^* \in \mathcal{G}_{var}(\sigma)$, then $\hat{\mu}$ has error $O(\sigma\sqrt{\epsilon})$.



$\text{var} \le \sigma^2$



$\text{var} \le \sigma^2$

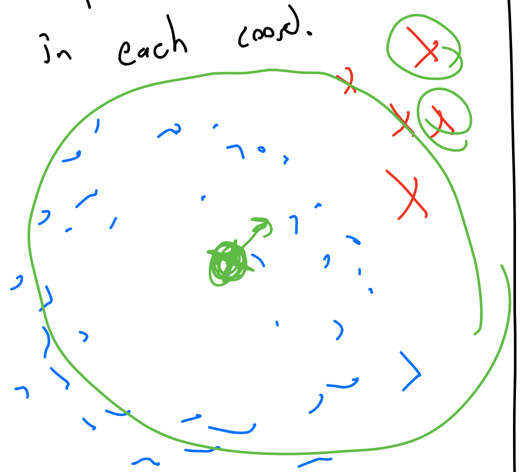
Gaussian distribution

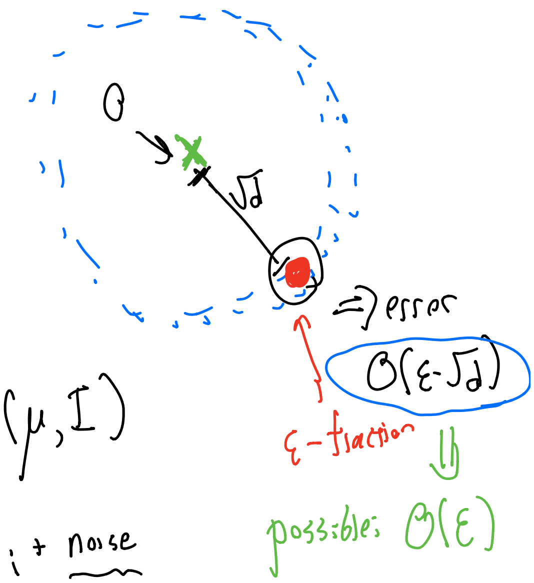
$N(\mu, I)$

mean: μ

covariance: $I_{d \times d}$

↳ independent var. 1 in each coord.





$X \sim N(\mu, I)$

$X_i = \mu_i + \underbrace{\text{noise}}_{\approx 1}$

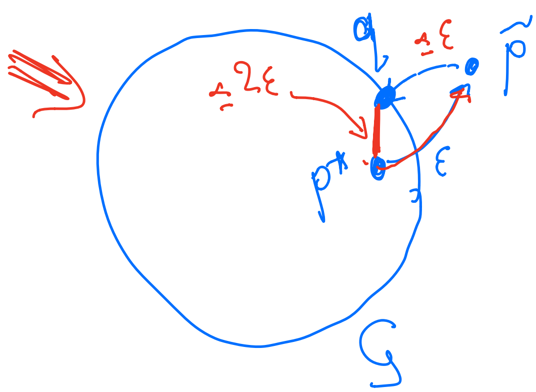
d dimensions

error ≈ 1 in each dimension

ℓ_2-norm: $\sqrt{\underbrace{1^2 + 1^2 + \cdots + 1^2}_{d}} = \sqrt{d}$

D (TV)

L ($\|\cdot\|_2$)



$q = \operatorname{argmin}_{q \in \mathcal{G}} D(q, \tilde{p})$, $\hat{\theta}(\tilde{p}) = \theta^*(q)$

$\theta^*(q) = \operatorname{argmin}_{\theta} L(q, \theta)$

(i.e., mean of q)

Minimum distance functional.

- Project onto $\mathcal{G}$, output opt. params

Proposition. Assume D is a pseudometric. Then:

$$L(p^*, \hat{\theta}(\tilde{p})) \le \underbrace{\sup_{\substack{p,q \in \mathcal{G} \\ D(p,q) \le 2\varepsilon}} L(p, \theta^*(q))}_{\text{modulus of continuity } m(\mathcal{G}, 2\varepsilon)}$$

$D(p^*, q) \le 2\varepsilon$

$D = TV$

$L(p,\theta) = \|\theta - \mu(p)\|_2$

$\mathcal{G} =$ all distributions

$\Rightarrow m(\mathcal{G}, 2\varepsilon) = \infty$

Gaussian:

$\mathcal{G}_{gauss} = \{ N(\mu, I),\ \mu \in \mathbb{R}^d \}$

$\boxed{\text{Claim: } m(\mathcal{G}_{gauss}, 2\varepsilon) = O(\varepsilon)}$

$\rightarrow$ any two Gaussians that are close in TV have similar means.

Lemma. Suppose $\mu, \mu' \in \mathbb{R}^d$ with $\|\mu - \mu'\|_2 = u$. Then

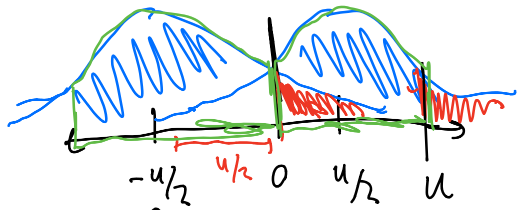
$$TV(p,q) = \int_{\mathbb{R}} |p(x) - q(x)|\, dx$$

$$\frac{1}{2}\min\left(\frac{u}{\sqrt{2\pi}}, 1\right) \le TV(N(\mu, I), N(\mu', I)) \le \min\left(\frac{u}{\sqrt{2\pi}}, 1\right)$$

Pf. Suffices to consider $N(\frac{u}{2}, 1)$, $N(-\frac{u}{2}, 1)$.

TV distance: $\frac{1}{2}\int |p(x) - q(x)|\, dx$

$$= \frac{1}{2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \left| e^{-(t+u/2)^2/2} - e^{-(t-u/2)^2/2} \right| dt$$

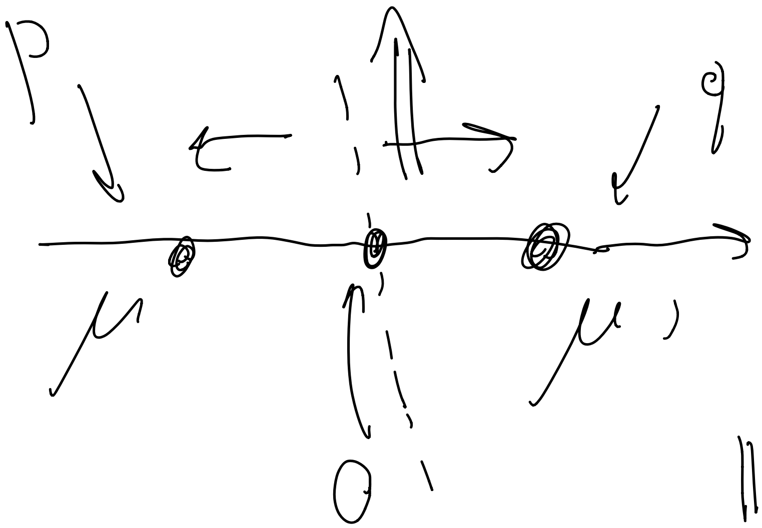


$$\int_{-u/2}^{u/2} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

≤ 1

$$= \int_{-u/2}^{u/2} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt \le \min\left(\frac{u}{\sqrt{2\pi}}, 1\right)$$

$\ge \frac{1}{2}$ if u is small enough entire integral $\gtrsim \frac{1}{2}$ if u is not small enough



$$\|\Sigma^{-1/2}(\hat{\mu} - \mu)\|_2 \leq O(\varepsilon)$$

$$\frac{1}{2}\int_{\mathbb{R}^d} |p(x) - q(x)|\, dx$$

Next. bounded covariance

$$\Sigma = \mathbb{E}[(x-\mu)(x-\mu)^T]$$

$$\Sigma \preceq \sigma^2 \cdot I, \quad \text{equivalently: } \|\Sigma\|_{op} \leq \sigma^3$$

$\mathcal{G}_{cov}(\sigma)$ = all dist^n's w/ cov. $\preceq \sigma^3$.

Procedure: Apply min. dist. functional

Want to show:

If $p, q \in \mathcal{G}_{cov}(\sigma)$ and $TV(p,q) \le 2\varepsilon$, then $\|\mu(p) - \mu(q)\|_2 \le ??$

$O(\sigma\sqrt{\varepsilon})$.

Pf. If $TV(p,q) \le \varepsilon$, we can get from p to q by deleting ε-fraction of mass and moving it elsewhere.

$$p \underset{\text{add } \varepsilon}{\overset{\text{delete } \varepsilon}{\rightleftarrows}} r \underset{\text{delete } \varepsilon}{\overset{\text{add } \varepsilon}{\rightleftarrows}} q$$

There exists r that can be obtained from both p and q by deleting ε-fraction + re-normalizing.

Lemma. $\underbrace{\|\mu(p)-\mu(r)\|_2}$ $\leq O(\sigma\sqrt{\varepsilon})$

$\|\mu(q)-\mu(r)\|_2 \leq O(\sigma\sqrt{\varepsilon})$

$\Rightarrow \|\mu(p)-\mu(q)\|_2 \leq O(\sigma\sqrt{\varepsilon})$.

If $\|Cov_p[X]\| \leq \sigma^2$

and E is an event w.p.

$1-\varepsilon$, then

$\|\underbrace{\mathbb{E}[X|E]}_{\mu(r)} - \underbrace{\mu}_{\mu(p)}\|_2 \leq O(\sigma\sqrt{\varepsilon})$.

Chebyshev's inequality.

If $cov \leq \sigma^2$ and $p(E) \geq \delta$

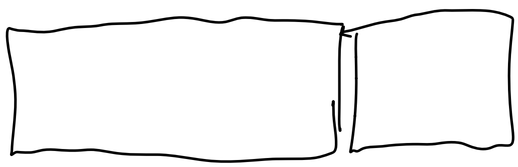
then $|\mathbb{E}[X-\mu|E]| \leq \sigma\sqrt{\frac{2(1-\delta)}{\delta}} \approx \sigma/\sqrt{\delta}$.

$\delta \approx 1 \Rightarrow \sigma\sqrt{1-\delta}$

---

pf. ① Reduce to $\delta \le \frac{1}{2}$

$\delta \ge \frac{1}{2}$ E $\neg E$ $\delta \le \frac{1}{2}$



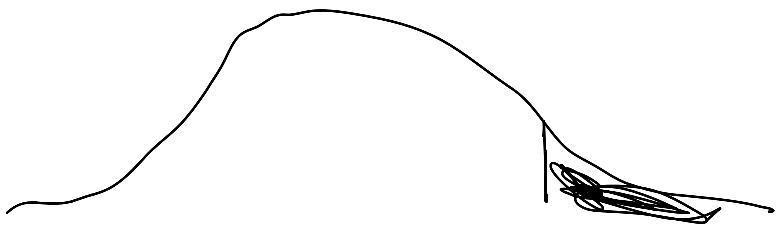
$$p(E)\,\mathbb{E}[X-\mu \mid E] + p(\neg E)\,\mathbb{E}[X-\mu \mid \neg E]$$

$$\mathbb{E}[X-\mu] = 0$$

$$\mathbb{E}[X-\mu \mid E] = \frac{p(\neg E)}{p(E)}\,\mathbb{E}[X-\mu \mid \neg E].$$

② Cauchy-Schwarz

Take-away: Bounded cov. dist^{n}s can't shift their mean much if we delete ϵ-fraction.



Recap-

- Use midpoint property of TV
- Just need to show that deleting ϵ-fraction can't change mean by much (tail bound)

Def^n Resilience

p is (ρ, ε)-resilient if

$$|\mathbb{E}[X - \mu \mid E]| \leq \rho$$

whenever $p(E) \geq 1 - \varepsilon$.

$\mathrm{Cov.} \leq \sigma \Rightarrow (\Theta(\sigma\sqrt{\varepsilon}), \varepsilon)$-resilient.

Min. dist. for $\mathcal{G}_{\mathrm{cov}}$ works

$\mathcal{G} = \{ p \mid p$ is (ρ, ε)-resilient$\}$

Min. dist. for $\mathcal{G}$