Topics.

Robustness to differences b/t train and test

- Training time. imperfect data collection
  - annotation error
  - measurement error
  - data poisoning

train ≠ test

- Test time.
  - temporal shift ("zoom")
  - sensor failure
  - WSJ → twitter

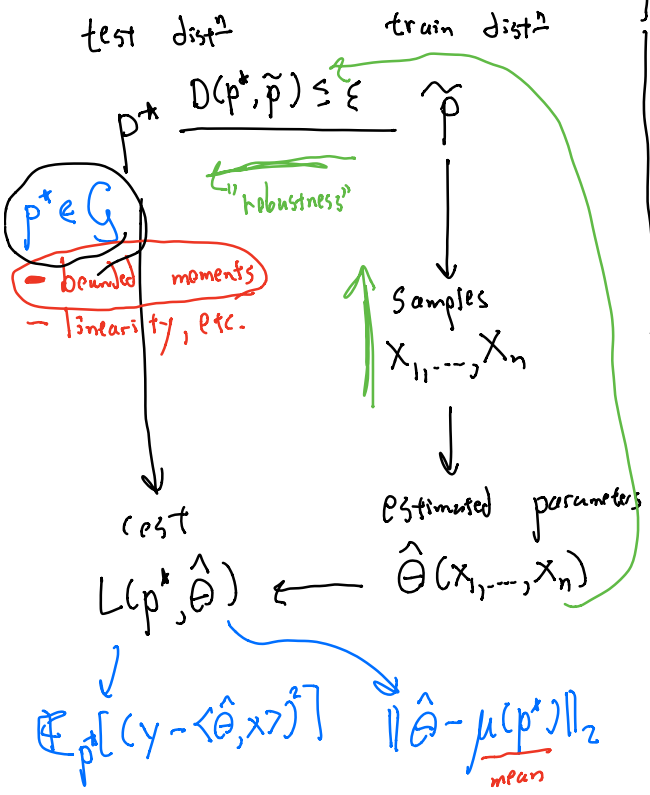
Perspectives. - worst case (vs. avg. case)

- high dimensional setting
  $d \approx 10^3$ or 10^6 or 10^9 vs. 10^1

Two senses:

- $d \gg 1$
- $d/n \gg 1$ ← Lasso ℓ_1 / sparse

non-parametric



regression:

- $\mathbb{E}[(y - \langle \theta^*, x \rangle)^4] \le K^4$
- $\mathbb{E}[\langle x, v \rangle^4] \le C^2 \cdot \mathbb{E}[\langle x, v \rangle^2]^2$
  $\forall\ v \in \mathbb{R}^d$

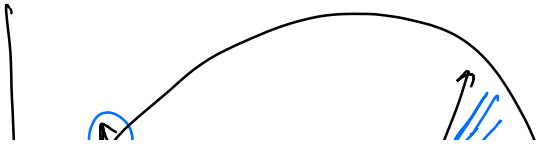
Training: corruptions

Lecs. 1-14

- p^* "nice", $\tilde{p}$ "ugly"
- intuition. "undo" corruption that took $p^* \to \tilde{p}$
- high-dimensional setting. can only undo at level of ysamps of points
- develop geometric + statistical tools
- non-convex, but tractable

Distribution shift.

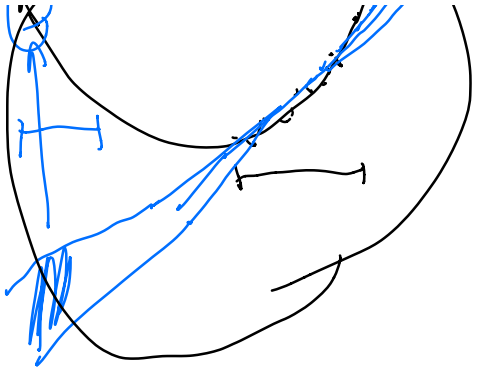
- p^*, $\tilde{p}$ both "nice"



- Idea 1 make $\tilde{p}$ "diverse"
  - find invariant properties
  - hopefully also hold in p^*
- Idea 2 use model uncertainty in $\hat{\theta}$ to notice if have very uncertain predictions on OOD data

Issue: Model mis-specification

- correct invariants might not be in your model
- within-model uncertainty << mis-specification error

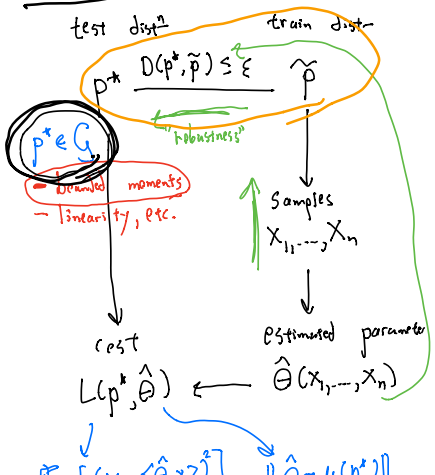


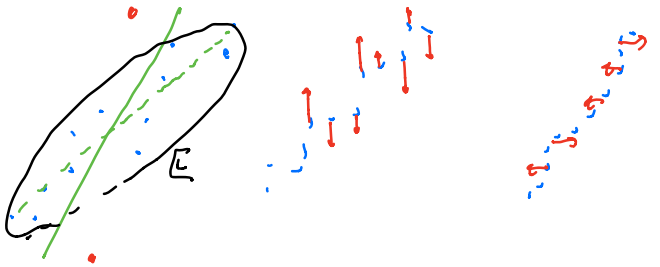
Non-parametric modeling
- kernels, NNs
- non-parametric inference: bootstrap
  - doubly-robust estimators

Robust opt estimators:

$$\underset{\theta}{\text{argmin}} \sup_{p:\, D(p,\tilde{p}) \le \varepsilon,\ p \in S} L(p, \theta)$$

Train time robustness





outliers

measurement error

↑ Wasserstein / transportation (later)

$$D(p,q) \approx TV(p,q) = \sup_E |p(E) - q(E)| = \frac{1}{2} \int |p(x) - q(x)|\, dx$$

cost: d^α, $\alpha = 0.1$

mean

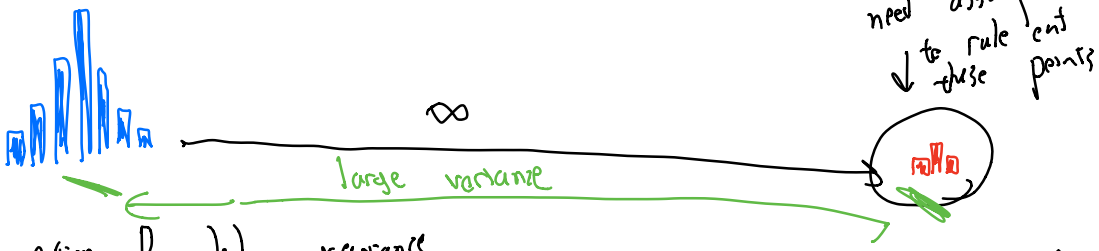
$TV(p,q) \le \varepsilon$: ε-fraction of points are deleted and replaced w/ arbitrary outliers

$E = \{x \mid p(x) > q(x)\}$

$\int \max(p(x) - q(x), 0)$

$= \int \max(q(x) - p(x), 0)$

1D outlier robustness



need assumptions to rule out these points

Assumption: Bounded variance

Assume $\mathbb{E}_{x\sim p}[(x-\mu)^2] \le \sigma^2$

$\Rightarrow G_{var}(\sigma)$

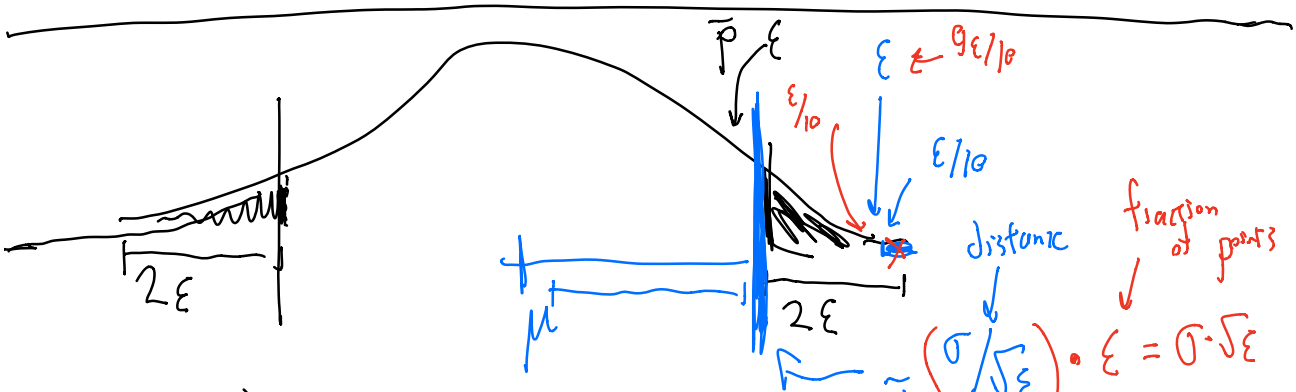
$L(p, \theta) = |\theta - \mu(p)|$ (mean)

Issue: If $\varepsilon \ge \frac{1}{2}$, in trouble.

"breakdown point"

Today: $\varepsilon \le \frac{1}{8}$.

$n = \infty$: you to observe $\tilde{p}$



distance

fraction of points

$\approx (\sigma/\sqrt{\varepsilon}) \cdot \varepsilon = \sigma \cdot \sqrt{\varepsilon}$

$\varepsilon \cdot (\sigma/\sqrt{\varepsilon})^2 = \sigma^2$

Truncated mean:

- Remove left + right (2ε)-tails.
- Take mean $\hat{\mu}$ of remaining points.

Proposition. Assume $TV(p^*, \tilde{p}) \le \varepsilon \le \frac{1}{8}$ and that $p^* \in G_{var}(\sigma)$. Then $|\hat{\mu} - \mu(p^*)| \le q\sigma\sqrt{2\varepsilon}$

$\frac{\varepsilon}{1-q\varepsilon}$ - fraction of points $\Rightarrow 2\varepsilon$ moment bend

$(\varepsilon \le \frac{1}{8})$

$\varepsilon^{3/4}$ under q^{th} moments

Chebyshev's inequality

Unsupervised: X

Supervised: X, y

$X = (X_1, X_2, \ldots, X_p)$

$\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)$

q

update Θ