STAT240 Problem Set 4

Mengqi Lin

April 6, 2021

Let p be the uniform distribution on $[-1,1]^d$. If p_n is the empirical distribution over n samples, show that $\boldsymbol{E}[W_1(p,p_n)] \geq \Omega(n^{-1/d})$.

solution Consider a partition of $[-1,1]^d$ into n cubes, each with $n^{-1/d}$ length. Then we know that p_n will be in charge of at most n cubes. Denote E to be the event that p_n is in charge of less than $n/2$ of the cubes. then

$$\mathbb{P}(E) = \sum_{i=1}^{n/2} \frac{\binom{n}{i} i!}{n^n} = c$$

So

$$\begin{aligned} \mathbb{E}[W_1(p,p_n)] &\geq \mathbb{E}[W_1(p,p_n)1_E] && (1) \\ &= \mathbb{P}(E)\mathbb{E}[W_1(p,p_n)|E] && (2) \\ &= c\mathbb{E}[W_1(p,p_n)|E] && (3) \end{aligned}$$

However, on the event E, at least $n/2$ of the cubes are not charged by p_n, and so the mass on these cubes will need to match to the n points. But clearly each time a cube is not charged by p_n, a proportion of the mass of p is at distance at least $N^{-1/d}/2$ of the support of p_n. So we must have

$$\mathbb{E}[W_1(p,p_n)|E] \geq c_1 N^{-1/d}$$

0.1 reference: on the rate of convergence in wasserstein distance of the empirical measure

STAT240 Problem Set 4

Due April 6th in class

Challenge problems (turn in as a separate document typset in LaTeX):

6. Consider linear regression with $L(p,\theta) = \boldsymbol{E}_{(X,Y)\sim p}[(Y - \langle \theta, X\rangle)^2]$ (note this is now the *risk* rather than the *excess risk* that we had before). Let $X' = [X, Y]$ be the d-dimensional vector that concatenates X and Y, and define $Z = Y - \langle \theta^*(p), X\rangle$. Suppose the following two conditions hold:

$$\boldsymbol{E}_p[|\langle v, X'\rangle|^3] \leq \kappa^3 \boldsymbol{E}[\langle v, X'\rangle^2]^2 \text{ for all } \|v\|_2 = 1, \tag{1}$$

$$\boldsymbol{E}_p[Z^2] \leq \sigma^2. \tag{2}$$

Assuming that $\kappa\epsilon\psi^{-1}(4\kappa/\epsilon) \leq \frac{1}{8}$, show that this family of distributions has modulus of continuity that is $(\sigma^2 + \epsilon^2)$.

[You may want to do this by first showing that the distribution is resilient in the sense of problem 5.*]*

solution We redefine that $L(p,\theta) = \sup_{f\in\mathcal{F}_\theta} \mathbb{E}_p[f(x)] - L^*(f,\theta)$ with $\mathcal{F}_\theta = \{f : f = (Y - \langle \theta, X\rangle^2)\}$ and that $L^*(f,\theta) = 0$, so then $L(p,\theta) = \mathbb{E}_p[(Y - \langle \theta, X\rangle^2)]$.
And we prove the result by showing that the distribution $p^* \in \mathcal{G}_\downarrow \cap \mathcal{G}_\uparrow$,
To first prove that $p^* \in \mathcal{G}_\downarrow(\rho_1, \epsilon)$, we show that $L(r, \theta^*(p)) \leq \rho_1$, where r is any ϵ friendly perturbation of p under f, i.e

$$\mathbb{E}_r[(Y - \langle \theta^*(p), X\rangle)^2] = \mathbb{E}_r[Z^2] \leq \rho_1$$

By the condition (1), choose v to be the unit vector in direction $(-\theta^*(p) \ \ 1)$, and denote $R = \sqrt{\theta^*(p)^2 + 1}$ so then $v = (-\theta^*(p) \ \ 1)^\mathsf{T}/R$. So $\langle v, X'\rangle = Z/R$, and condition (1) implies that

$$\mathbb{E}[|(Z/R)^3|] \leq \kappa^3 \mathbb{E}[(Z/R)^2]^2 \tag{3}$$

$$\mathbb{E}[|Z^3|] \leq \kappa^3 \mathbb{E}[Z^2]^2/R \tag{4}$$

i.e Z has 3rd bounded moment, and bounded by

$$\frac{\sigma'^3}{4} = \kappa^3 \mathbb{E}[Z^2]^2/R$$

Additionally with

$$\mathbb{E}[Z^2] \leq \sigma^2$$

So by Prop 4.7 we have Z is (ρ_1, ϵ) resilient, where $\rho_1 = \max(2\sqrt{2\epsilon/R}\mathbb{E}[Z^2]\kappa^{3/2}, 4\epsilon^2 + 2\epsilon\sigma)$, and by the fact that $\kappa\epsilon\psi^{-1}(4\kappa/\epsilon) \leq \frac{1}{8}$, we further have $\rho_1 = (c\sigma^2, 4\epsilon^2 + 2\epsilon\sigma) = \mathcal{O}(\sigma^2 + \epsilon^2)$. Therefore,

$$\sup |\mathbb{E}_r[Z^2] - \mathbb{E}_p[Z^2]| \leq \rho_1$$

And with $\mathbb{E}_p[Z^2] \leq \sigma^2$, we have

$$|\mathbb{E}_r[Z^2]| \leq \mathcal{O}(\sigma^2 + \epsilon^2)$$

So $p^* \in \mathcal{G}_{\downarrow}(\rho_1, \epsilon)$. Now we prove that $p^* \in \mathcal{G}_{\uparrow}(\rho_2, \epsilon)$, i.e to prove that whenever $L(r, \theta) \leq \rho_1$, we have that $L(p, \theta) \leq \rho_2$:

By condition (1), we know that X' has bounded 3rd moment with

$$\frac{\sigma^3}{4} = \kappa^3 \mathbb{E}[\langle v, X' \rangle^2]^2$$

So p is (ρ, ϵ)-resilient, with $\rho = \max(2\sqrt{2\epsilon}\kappa^{3/2}\mathbb{E}[\langle v, X'\rangle^2], 4\epsilon^2 + 2\epsilon\sqrt{\mathbb{E}[\langle v, X'\rangle^2]}) \leq \max(\sqrt{2}/8\mathbb{E}[\langle v, X'\rangle^2], 4\epsilon^2 + 2\epsilon\sqrt{\mathbb{E}[\langle v, X'\rangle^2]})$.

$$|\mathbb{E}_p[\langle v, X'\rangle^2] - \mathbb{E}_r[\langle v, X'\rangle^2]| \leq \rho \tag{5}$$
$$\mathbb{E}_p[\langle v, X'\rangle^2] \leq c_1 \mathbb{E}_r[\langle v, X'\rangle^2] \tag{6}$$

And we take v in (6) to be $(-\theta, \;\; 1)^\mathsf{T}/\sqrt{\theta^2 + 1}$ So then $\langle v, X'\rangle^2 = (Y - \theta X)^2$, we have

$$L(p, \theta) \leq c_1 L(r, \theta)$$

So whenever $L(r, \theta) \leq \rho_1$, we have $L(r, \theta) \leq \rho_2 = c_1\rho_1 = \mathcal{O}(\sigma^2 + \epsilon^2)$. And so $p^* \in \mathcal{G}_{\uparrow}(\rho_2, \epsilon)$. Hence this family of distributions has modulus of continuity that is $\mathcal{O}(\sigma^2 + \epsilon^2)$.

STAT 240 HW1

Mengqi Lin

February 2021

1 Challenge Problem 1

Let $X \sim p^*, X_n \sim p_n^*, \tilde{X_n} \sim \tilde{p_n}$.
WLOG suppose $\sigma = 1$, then $X \sim N(\mu, I)$. So $\langle X, v \rangle$ is a 1-dimensional Gaussian with variance 1 for all unit vector v.
So by problem 3, $\langle X, v \rangle$ is $(\mathcal{O}(\epsilon), \epsilon)$ - stable for $\epsilon \leq 1/4, \forall$ unit vector v. Therefore,

$$X \sim p^* \text{ is } (\mathcal{O}(\epsilon), \epsilon) - \text{ stable .}$$

So by problem 4,

$$X_n \sim p_n^* \text{ is } (\mathcal{O}(\epsilon), \frac{\epsilon}{2} - \mathcal{O}(\sqrt{\frac{d}{n}}) - \text{ stable}$$

around the true mean u with probability at least 1- 2exp(-cϵ^2n). Since $n >> \frac{d}{\epsilon^2}$, X_n is $(\mathcal{O}(\epsilon), c\epsilon)$ - stable. And it's easy to see that the Tukey depth for p_n^* is 1/2, And by $TV(\tilde{p_n}, p_n^*) \leq \epsilon$ and problem 5, we have $\left\|\hat{\theta}_{Tukey}(\tilde{p}_n) - u(p_n^*)\right\|_2 \leq \mathcal{O}(\epsilon)$. And since the samples are sub-Gaussian variables, so by Hoeffding's Inequality, we have $\left\|u(p_n^*) - u(p*)\right\|_2 \leq \mathcal{O}(\epsilon)$, with high probability. And by triangle inequality, we have $\left\|\hat{\theta}_{Tukey}(\tilde{p}_n) - u(p*)\right\|_2 \leq \mathcal{O}(\epsilon)$

2 Challenge Problem 2

WLOG, suppose $\mu = 0$, then $p^* \sim N(0, I)$. Construct the $\tilde{p}$ by putting ϵ-fraction points on $b = (2\sqrt{d}, 0, ..., 0)$, and scale the density in the rest of the points by $\tilde{p} = (1-\epsilon)p^*$. By symmetric, we know that the minimizer $\hat{\theta}_{geom}(\tilde{p})$ must lie on the first axis between the origin point o and b. Now let's suppose the minimizer $\hat{\theta}_{geom}(\tilde{p}) = (\theta, 0, ..., 0)$.

$$\begin{aligned}
\mathbb{E}_{x\sim\tilde{p}}[\|X-\theta\|_2] &= \int_{\bar{b}} \|X-\theta\|_2 d\tilde{P} + \int_b \|X-\theta\|_2 d\tilde{P} && (1)\\
&= \int_{\bar{b}} \|X-\theta\|_2 \phi(x)(1-\epsilon)dx + |2\sqrt{d}-\theta|\epsilon && (2)\\
&= \int_{\bar{b}} \|X-\theta\|_2 \frac{1}{2\pi^{\frac{d}{2}}} \exp\left(\frac{\|X\|_2^2}{2}\right)(1-\epsilon)dx + |2\sqrt{d}-\theta|\epsilon && (3)\\
& && (4)
\end{aligned}$$

Taking derivatite wrt θ:

$$\int_{\bar{b}} \frac{-(X_1-\theta)}{\|X-\theta\|_2} \frac{1}{2\pi^{\frac{d}{2}}} \exp\left(\frac{\|X\|_2^2}{2}\right)(1-\epsilon)dx + \operatorname{sgn}(2\sqrt{d}-\theta)\epsilon$$

Setting the formula to zero and to obtain the optimizer θ We can draw the above formula versus θ, and then find the θ that hit zero, which is around $2\epsilon\sqrt{d}$

STAT260 Problem Set 3

Mengqi Lin

March 9th

Challenge problems (turn in as a separate document typset in LaTeX):

6. Construct a distribution that is $(\sqrt{\epsilon}, \epsilon)$-resilient in the $\mathcal{S}_k$-norm for all $\epsilon < 1/4$, but not $(\rho, 1/10)$-resilient in the ℓ_2-norm for any $\rho < \Omega(k^{0.1})$.

   [The constants 1/4, 1/10, 0.1 are all arbitrarily chosen, the point is to show a polynomial separation between $_k$ and ℓ_2 for some distribution. Note that your construction will likely need to have d/k going to ∞ as $k \to \infty$.]

   Solution. Let d = $k^{1.2}$, note d/k $\to \infty$. denote $e_j = (0, 0, \ldots, d\sqrt{\epsilon/k}, \ldots, 0) \in \mathbb{R}^d$. i.e only the j-th element is non-zero. Suppose X $\sim p = \text{Unif}[\text{e}_j, j \in [d]]$. i.e

   $$\mathbb{P}(X = e_j) = 1/d$$

   then an ϵ deletion of distribution p, i.e distribution r has mass only on (1-ϵ)d points. And WLOG, assume r has mass on the first (1-ϵ)d points. Let Y $\sim r = \text{Unif}\big(\text{e}_j, j \in [(1-\epsilon)\text{d}]\big)$.

   $$\mathbb{P}(Y = e_j) = 1/(1-\epsilon)d$$

   $$u(p) - u(r) = \big(\frac{-\epsilon}{1-\epsilon} 1\{1_{\{(1-\epsilon)d\}}, 0 \ldots, 0\} + 1\{0, \ldots, 0, 1_{\{\epsilon d\}}\}\big)\sqrt{\frac{\epsilon}{k}}$$

   So, for ℓ_2 norm:

   $$\|u(p) - u(r)\|_2 = \frac{\epsilon}{\sqrt{1-\epsilon}}\sqrt{d/k} = \frac{\epsilon}{\sqrt{1-\epsilon}} k^{0.1}$$

   Therefore, it's not $(\rho, 1/10)$-resilient in the ℓ_2-norm for any $\rho < \Omega(k^{0.1})$.

   For $\mathcal{S}_k$ norm, when k is sufficiently large enough, it's obvious that u(p) - u(r) has only two directions, one is $\frac{-\epsilon}{1-\epsilon} 1\{1_{\{(1-\epsilon)d\}}, 0 \ldots, 0\}$, and the other is $1\{0, \ldots, 0, 1_{\{\epsilon d\}}\}$, so the $\mathcal{S}_k$ norm is the maximum inner product of k-nonzeros unit vector in these two directions.

   $$\|u(p) - u(r)\|_{\mathcal{S}k} = \sqrt{\frac{\epsilon}{k}} \max\{k \frac{\epsilon}{1-\epsilon}\frac{1}{\sqrt{k}}, k\frac{1}{\sqrt{k}}\} = \sqrt{\epsilon}$$

   as desired. ■

7. For linear regression, suppose that p satisfies the following higher-order bounded noise and hypercontractivity conditions:

   $$\mathbb{E}_p[Z^8] \leq \tau^8, \text{ and } \mathbb{E}_p[\langle X, v\rangle^8] \leq \kappa \mathbb{E}_p[\langle X, v\rangle^2]^4. \tag{1}$$

   Show that p is resilient for linear regression with a correspondingly better dependence on ϵ, and design a version of the `QuasigradientDescentLinReg` algorithm for this case. (For the algorithm, you may assume that we have an oracle for maximizing over v to compute the bounded noise and hypercontractive quantities, and also take the quasigradient bounds as given; the point is to prove analogs of Lemma 3.10 and Lemma 3.11.)

Solution. The proof in Prop 3.4 involves two key facts: the bounded variance of Z to ensure the resilience of $\mathbb{E}_p[\mathcal{S}_p^{1/2}XZ]$ and the hypercontractivity to ensure the approximation of $S_r = \mathbb{E}_r[XX^\top]$ and $S_p = \mathbb{E}_p[XX^\top]$. Here I claim that both of these two contraints are tighter and hence result in a better error.

Claim 1. *p satisfies the first condition (bounded variance of Z) in Prop 3.4 with tighter constants.*

Proof. $\forall$ unit vector v, By Jensen's Inequality, Cauchy Inequality and condition(1) :

$$\mathbb{E}_p^2\big[\langle X, v\rangle^2 Z^2\big] \le \mathbb{E}_p\big[\langle X, v\rangle^4 Z^4\big] \le \sqrt{\mathbb{E}_p\big[\langle X, v\rangle^8\big]\mathbb{E}_p\big[Z^8\big]} \le \sqrt{\tau^8\kappa\mathbb{E}_p[\langle X, v\rangle^2]^4} = \tau^4\kappa^{\frac{1}{2}}\mathbb{E}_p^2\big[\langle X, v\rangle^2\big]$$

Hence, for all v:

$$\mathbb{E}_p\big[\langle X, v\rangle^2 Z^2\big] \le \tau^2\kappa^{\frac{1}{4}}\mathbb{E}_p\big[\langle X, v\rangle^2\big]$$

Therefore,

$$\mathbb{E}_p\big[XZ^2X^\top\big] \preceq \sigma^2\mathbb{E}_p\big[XX^\top\big]$$

with $\sigma^2 = \tau^2\kappa^{\frac{1}{4}}$ Therefore, p satisfies the first condition (bounded variance). ∎

Claim 2. *p satisfies the condition $\mathcal{S}_p \approx \mathcal{S}_r$ with tighter constants.*

Proof. take $S = \langle X, v\rangle^2$, then S has 4-th bounded moments:

$$\begin{aligned}
\mathbb{E}[[S - \mathbb{E}[S]]^4] &= \mathbb{E}[S^4] + \mathbb{E}^4[S] - 4\mathbb{E}[S^3]\mathbb{E}[S] - 4\mathbb{E}^4[S] + 6\mathbb{E}[S^2]\mathbb{E}^2[S] && (2)\\
&\le (\kappa - 3)\mathbb{E}^4[S] - 4\mathbb{E}[S^3]\mathbb{E}[S] + 6\mathbb{E}[S^2]\mathbb{E}^2[S] && (3)\\
&\le (\kappa - 3)\mathbb{E}^4[S] - 4\mathbb{E}^4[S] + 6\mathbb{E}^4[S] && (4)\\
&= (\kappa - 1)\mathbb{E}^4[S] && (5)
\end{aligned}$$

The (3) follows from the given condition, and (4) follows that x^3 is a convex function and using the Jensen's Inequality we have $-\mathbb{E}[S^3] \le -\mathbb{E}^3[S]$. And the fact that $\mathbb{E}[S^2] \le \mathbb{E}^2[S]$.
Therefore S has 4-th bounded moment with $\sigma^4 = (\kappa - 1)\mathbb{E}^4[S]$

We know that distribution with k-th moment is $(\sigma\epsilon^{1-1/k}, \epsilon)$- resilient. So S is $((\kappa - 1)^{1/4}\mathbb{E}[S]\epsilon^{3/4}, \epsilon)$ resilient. Hence,

$$|\mathbb{E}_p[\langle X, v\rangle^2] - \mathbb{E}_r[\langle X, v\rangle^2]| \le (\kappa - 1)^{1/4}\epsilon^{3/4}\mathbb{E}_p[\langle X, v\rangle^2]$$

So $\mathcal{S}_p \approx \mathcal{S}_r$ with a tighter factor than what we have (which is 1/2) in prop 3.4:

$$(1 - (\kappa - 1)^{1/4}\epsilon^{3/4})\mathcal{S}_p \preceq \mathcal{S}_r \preceq (1 + (\kappa - 1)^{1/4}\epsilon^{3/4})\mathcal{S}_p$$

∎

Then follow the similar argument in Prop 3.4 we can have better dependence on ϵ. ∎