# STAT260 Problem Set 3

Due March 9th in class

**Regular problems:**

1. In class we defined sets $\mathcal{G}^{\downarrow}$ and $\mathcal{G}^{\uparrow}$ for an arbitrary loss $L(p, \theta)$. Here we consider the following generalized construction, that also incorporates a *bridge function* $B(p, \theta)$:

$$\mathcal{G}^{\downarrow}(\rho_1, \epsilon) \stackrel{\text{def}}{=} \{p \mid B(r, \theta^*(p)) \leq \rho_1 \text{ whenever } r \leq \frac{p}{1-\epsilon}\}, \tag{1}$$

$$\mathcal{G}^{\uparrow}(\rho_1, \rho_2, \epsilon) \stackrel{\text{def}}{=} \{p \mid L(p, \theta) \leq \rho_2 \text{ whenever } B(r, \theta) \leq \rho_1 \text{ for some } r \leq \frac{p}{1-\epsilon}, \theta\}. \tag{2}$$

Show that the modulus of continuity of $\mathcal{G}^{\downarrow}(\rho_1, \epsilon) \cap \mathcal{G}^{\uparrow}(\rho_1, \rho_2, \epsilon)$ under $L$ is at most $\rho_2$.

2. Here we bound the modulus of continuity for linear classification. Given $(x, y) \sim p$ where $x \in \mathbb{R}^d$ and $y \in \{\pm 1\}$, Define $L(p, \theta) = \mathbb{P}_{(x,y)\sim p}[y \neq \text{sign}(\langle x, \theta \rangle)]$ and $B(p, \theta) = \mathbb{E}_{(x,y)\sim p}[\max(0, 1 - y\langle x, \theta \rangle)]$, and let $\theta^*(p)$ be the minimizer of $B(p, \theta)$. Suppose that $p$ satisfies the following two properties:

   - $\mathbb{E}_{(x,y)\sim p}[\max(0, 1 - y\langle x, \theta^*(p) \rangle)] \leq (1-\epsilon)\rho_1$
   - Whenever $\mathbb{P}_{(x,y)\sim p}[y\langle x, \theta \rangle \leq \frac{1}{2}] \leq \epsilon + 2(1-\epsilon)\rho_1$ for some $\theta$, then $\mathbb{P}_{(x,y)\sim p}[y\langle x, \theta \rangle \leq 0] \leq \rho_2$.

   Show that $p \in \mathcal{G}^{\downarrow}(\rho_1, \epsilon) \cap \mathcal{G}^{\uparrow}(\rho_2, \epsilon)$.

   *[Remark: Note that the second condition is a type of tail bound, where in every direction where we are somewhat unlikely to be close to the boundary, we are very unlikely to cross the boundary entirely.]*

3. Define the norm $\|x\|_{\mathcal{S}_k} = \max_{\|v\|_2 \leq 1, \|v\|_0 \leq k} \langle x, v \rangle$.

   (a) What is the dual norm to $\|\cdot\|_{\mathcal{S}_k}$? (Note: There isn't a simple closed form; it suffices to provide a description of what the unit norm ball looks like.)

   (b) Show that if $x$ and $y$ both have at most $k$ non-zero entries, then $\|x - y\|_{\mathcal{S}_{2k}} = \|x - y\|_2$.

   (c) Show that any distribution that is $(\rho, \epsilon)$-resilient in the $\ell_2$-norm is also $(\rho, \epsilon)$-resilient in the $\mathcal{S}_k$-norm.

4. Let $Z = Y - \langle \theta^*(p), X \rangle$. Suppose that instead of the bounded noise condition $\mathbb{E}_p[XZ^2Z^\top] \preceq \sigma^2 \mathbb{E}_p[XX^\top]$, we instead have $\mathbb{E}_p[Z^4] \leq \tau^4$. Also assume that the distribution is hypercontractive with $\kappa = \mathcal{O}(1)$. Show that this implies the bounded noise condition for some $\sigma$ that is at most a constant multiple of $\tau$.

**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 $\ell_2$-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 $\mathcal{S}_k$ and $\ell_2$ for some distribution. Note that your construction will likely need to have $d/k$ going to $\infty$ as $k \to \infty$.]*

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

$$\mathbb{E}_p[Z^8] \leq \tau^8, \text{ and } \mathbb{E}_p[\langle X, v\rangle^8] \leq \kappa \mathbb{E}_p[\langle X, v\rangle^2]^4. \tag{3}$$

Show that $p$ is resilient for linear regression with a correspondingly better dependence on $\epsilon$, 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.)