Homework 2
Concentration Bounds, Uniform Laws

CMU 10-716: Advanced Machine Learning (Spring 2019)

OUT: Feb. 5, 2019
DUE: **Feb. 15, 2019, 11:59 PM**.

Instructions:

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. It is also OK to get clarification (but not solutions) from books or online resources, again after you have thought about the problems on your own. There are two requirements: first, cite your collaborators fully and completely (e.g., "Bob explained to me what is asked in Question 4.3"). Second, write your solution *independently*: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only.
- **Submitting your work:** Assignments should be submitted as PDFs using Gradescope unless explicitly stated otherwise. Each derivation/proof should be completed on a separate page. Submissions can be handwritten, but should be labeled and clearly legible. Else, submission can be written in LaTeX. Upon submission, label each question using the tempate provided by Gradescope.
- **Start Early.**

1 Concentration Bounds (Hubert and Karthika)

1. (a) [4pts] Enumerate two different non-negative random variables where Markov's inequality:

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t},$$

is tight i.e. becomes an equality for some values of t.

Hint: Consider discrete random variables.

(b) i. [8pts] Let Z be a zero mean bounded random variable, such that $a \leq Z \leq b$. Then, show that for any t,

$$\mathbb{E}[e^{tZ}] \leq e^{t^2(b-a)^2/8}.$$

Hint:

A. Upper bound e^{tZ} via a convex combination of e^{tb} and e^{ta}, by noting that Z is a convex combination of a and b. Namely, showing that $\mathbb{E}(e^{tZ}) \leq e^{g(u)}$ where $u = t(b-a)$ and $g(u) = -\gamma u + \log(1 - \gamma + \gamma e^u)$ with $\gamma = -\frac{a}{b-a}$.

B. Use the fact that $g''(u) \leq \frac{1}{4}$ and conclude the proof using Taylor's theorem.

ii. [4pts] Show that every bounded random variable ($l \leq X \leq u$) is sub-Gaussian with parameter $\sigma = \frac{u-l}{2}$.
Hint: Use the conclusion from previous question.

(c) Let $\{X_i\}_{i=1}^n$ be a sequence of iid zero-mean sub-Gaussian random variables, each with sub-Gaussian parameter σ. Consider $Z_n = \max_{1 \leq i \leq n} |X_i|$ Prove the following statements.

i. [5pts] $\mathbb{E}[Z_n] < \sigma\sqrt{2\,\log(2n)}$.

Hint: Consider a new random variable without the absolute value.

ii. [5pts] $\mathbb{P}[Z_n > t] < 2ne^{-\frac{t^2}{2\sigma^2}}$

iii. [5pts] Now consider the case where the variables X_i follow a Gaussian distribution $\mathcal{N}(0, \sigma^2)$ instead. Prove that:

$$\mathbb{E}[Z_n] \leq \sqrt{2\sigma^2 \log(n)} + \frac{4\sigma}{\sqrt{2\log(n)}}, \quad \forall n \geq 2.$$

Hint: Use the Bernoulli inequality on the expectation of Z, i.e. for every non-negative integer n, and real nuber $p \geq -1$, $(1+p)^n \geq 1 + np$.

(d) [15pts] Let X be a zero mean sub-Gaussian random variable. Show that there exists $b > 0$ and $c > 0$ s.t.

$$\mathbb{P}(|X| \geq t) \leq c\,\mathbb{P}(|Z| \geq t),$$

for all $t \geq 0$, where $Z \sim N(0, b^2)$.
Hint:

A You can use the fact that for a standard normal $Y \sim N(0,1)$ and any $y \geq 0$, letting $\phi(y)$ denote the standard normal density, we have that:

$$\mathbb{P}(Y \geq y) \geq \phi(y)(\frac{1}{y} - \frac{1}{y^3}).$$

B Bound the ratio $P(X \geq t)/P(Z \geq t)$, consider two cases: (i) $t \leq 2\sigma$ and (ii) $t > 2\sigma$.

2 Continuity of Functionals (Ritika)

Recall that the functional γ is continuous in the sup-norm at F if for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all G such that $\|G - F\|_\infty \le \delta$ implies that $|\gamma(G) - \gamma(F)| \le \epsilon$.

1. [**13 pts**] Given n i.i.d. samples with law specified by F, let $\widehat{F}_n$ be the empirical CDF. Show that if γ is continuous in the sup-norm at F, then $\gamma(\widehat{F}_n) \xrightarrow{\text{prob}} \gamma(F)$.

2. Which of the following functionals are continuous with respect to the sup-norm? Prove or disprove.

   (a) [**4 pts**] The mean functional $F \mapsto \int x dF(x)$.

   (b) [**4 pts**] The Cramér-von Mises functional $F \mapsto \int [F(x) - F_0(x)]^2 dF_0(x)$.

   (c) [**4 pts**] The quantile functional $Q_\alpha(F) = \inf\{t \in \mathbb{R} \mid F(t) \ge \alpha\}$.

3 Rademacher Complexity (Biswajit)

Consider a class $\mathscr{F}$ of functions uniformly bounded by b. Let $\{X_1, \ldots, X_n\}$ be a set of samples from a probability distribution $\mathbb{P}$ over a sample space $\mathcal{X}$. Define

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}} := \sup_{f \in \mathscr{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X)] \right|. \tag{1}$$

As usual the Rademacher complexity of $\mathscr{F}$ is defined as,

$$\mathcal{R}_n(\mathscr{F}) = \mathbb{E}_{X,\varepsilon} \left[\sup_{f \in \mathscr{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(X_i) \right| \right]. \tag{2}$$

We have already seen the upper bound, that with probability at least $1 - \exp\left(-\frac{n\delta^2}{2b^2}\right)$,

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}} \le 2\mathcal{R}_n(\mathscr{F}) + \delta. \tag{3}$$

In this question we will prove the following lower bound: with probability at least $1 - \exp\left(-\frac{n\delta^2}{2b^2}\right)$

$$\|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}} \ge \frac{1}{2}\mathcal{R}_n(\mathscr{F}) - \frac{\sup_{f \in \mathscr{F}} |\mathbb{E}[f]|}{2\sqrt{n}} - \delta. \tag{4}$$

1. [**2 pts**] Show the following one-sided inequality: with probability at least $1 - \exp\left(-\frac{n\delta^2}{2b^2}\right)$,

   $$\|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}} - \mathbb{E}_X \|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}} \ge -\delta \tag{5}$$

   Hint: Use the bounded difference inequality (BDI): `https://people.eecs.berkeley.edu/~bartlett/courses/281b-sp06/bdddiff.pdf`. Assume it holds without proof.

2. [**10 pts**] Consider the class of centered functions defined as

   $$\overline{\mathscr{F}} = \{f - \mathbb{E}_X[f] \mid f \in \mathscr{F}\}. \tag{6}$$

   Show that

   $$\mathcal{R}_n(\overline{\mathscr{F}}) \ge \mathcal{R}_n(\mathscr{F}) - \frac{\sup_{f \in \mathscr{F}} |\mathbb{E}[f]|}{\sqrt{n}}. \tag{7}$$

3. [**8 pts**] Show that $\mathcal{R}_n(\overline{\mathscr{F}}) \leq 2\mathbb{E}_X\|\mathbb{P}_n - \mathbb{P}\|_{\mathscr{F}}$ and use the results from the previous subproblems to prove the inequality.

   Hint: Start with $\mathcal{R}_n(\overline{\mathscr{F}})$ and introduce a ghost sample to get rid of $\mathbb{E}[f]$. Get rid of ε_i from the resulting expression.

4. [**5 pts**] Consider the function class $\mathscr{F} = \{\text{sign}(\langle\theta, x\rangle) \mid \theta \in \mathbb{R}^d, \|\theta\|_2 = 1\}$. Assume that for iid samples X_i drawn from some distribution, $\{X_1, \dots, X_n\}$ are linearly independent with probability 1 for $d \geq n$. Compute $\mathcal{R}_n(\mathscr{F})$ when $d \geq n$.

   Hint: You can assume without proof that a linear classifier can shatter any set of linearly independent vectors $\{X_1, \dots, X_n\} \subset \mathbb{R}^d$, where $d \geq n$.