High Dimensional Statistics

Spring 2025

Sharif University of Technology

Dr. Amir Najafi

Homework 2 — Martingale Based Inequalities and Concentration of Measure — Due: 1404/1/24

Problem 1: Bennett's Inequality *[10 points]*

(a) Consider a zero-mean random variable such that $|X_i| \leq b$ for some $b > 0$. Prove that

$$\log \mathbb{E}\big[e^{\lambda X_i}\big] \leq \sigma_i^2 \lambda^2 \left\{ \frac{e^{\lambda b} - 1 - \lambda b}{(\lambda b)^2} \right\} \quad \text{for all } \lambda \in \mathbb{R},$$

where $\sigma^2 = \text{Var}(X_i)$.

(b) Given independent random variables $X_1, \ldots, X_n$ satisfying the condition of part (a), let

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \text{Var}(X_i)$$

be the *average variance*. Prove *Bennett's inequality*:

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i \geq \delta\Big] \leq \exp\left\{ -\frac{n\sigma^2}{b^2} h\left(\frac{b\delta}{\sigma^2}\right) \right\},$$

where $h(t) = (1+t)\log(1+t) - t$ for $t \geq 0$.

(c) Show that Bennett's inequality is at least as good as Bernstein's inequality.

Problem 2: Eviation Inequalities in a Hilbert Space *[15 points]*

Let $\{X_i\}_{i=1}^n$ be a sequence of independent random variables taking values in a Hilbert space $\mathcal{H}$, and suppose that $\|X_i\|_{\mathcal{H}} \leq b_i$ almost surely. Consider the real-valued random variable

$$S_n = \Big\| \sum_{i=1}^{n} X_i \Big\|_{\mathcal{H}}.$$

(a) Show that, for all $\delta > 0$,

$$\mathbb{P}\Big[\big|S_n - \mathbb{E}[S_n]\big| \geq n\delta\Big] \leq 2\exp\Big(-\tfrac{n\delta^2}{8b^2}\Big),$$

where

$$b^2 = \frac{1}{n}\sum_{i=1}^{n} b_i^2.$$

(b) Show that

$$\mathbb{P}\Big[\tfrac{S_n}{n} \geq a + \delta\Big] \leq \exp\Big(-\tfrac{n\delta^2}{8b^2}\Big),$$

where

$$a = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n} \mathbb{E}\big[\|X_i\|_{\mathcal{H}}^2\big]}.$$

Problem 3: Shannon Entropy and Kullback–Leibler Divergence *[10 points]*

Given a discrete random variable $X \in \mathcal{X}$ with probability mass function p, its *Shannon entropy* is defined by

$$\mathbb{H}(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x).$$

(a) Consider the random variable $Z = p(U)$, where U is uniformly distributed over $\mathcal{X}$. Show that

$$\mathbb{H}(Z) = \frac{1}{|\mathcal{X}|}\Big[\log|\mathcal{X}| - \mathbb{H}(X)\Big].$$

(b) Use part (a) to show that the Shannon entropy for a discrete random variable is maximized by a uniform distribution.

(c) Given two probability mass functions p and q, specify a choice of random variable $Y = D\big(p\|q\big)$, corresponding to the *Kullback–Leibler divergence* between p and q.

Problem 4: Entropy and Constant Shifts *[10 points]*

(a) Show that for any random variable X and constant $c \in \mathbb{R}$,

$$\mathbb{H}\big(e^{\lambda(X+c)}\big) = e^{-\lambda c}\mathbb{H}\big(e^{\lambda X}\big).$$

(b) Use part (a) to show that if X satisfies

$$\mathbb{H}\big(e^{\lambda X}\big) \leq \frac{1}{2}\sigma^2\lambda^2\varphi_X(\lambda),$$

where $\varphi_X(.)$ denotes the MGF of X, then so does $X + c$ for any constant c.

Problem 5: Total Variation and Wasserstein *[10 points]*

Consider the Wasserstein distance based on the Hamming metric, namely

$$W_p(\mathbb{P}, \mathbb{Q}) = \inf_{\mathbb{M}} \mathbb{M}[X \neq Y],$$

where the infimum is taken over all couplings $\mathbb{M}$—that is, distributions on the product space $\mathcal{X} \times \mathcal{X}$ with marginals $\mathbb{P}$ and $\mathbb{Q}$, respectively. Show that

$$\inf_{\mathbb{M}} \mathbb{M}[X \neq Y] = \|\mathbb{P} - \mathbb{Q}\|_{\text{TV}} = \sup_A\big|\mathbb{P}(A) - \mathbb{Q}(A)\big|,$$

where the supremum ranges over all measurable subsets A of $\mathcal{X}$. Equivalently, show that

$$\inf_{\mathbb{M}} \mathbb{M}[X \neq Y] = \frac{1}{2} \int_{x \in \mathcal{X}} \big|p(x) - q(x)\big| dx.$$

Problem 6: Concentration on the Euclidean Ball *[15 points]*

Consider the uniform measure $\mathbb{P}$ over the Euclidean unit ball

$$\mathbb{B}_2^n = \big\{x \in \mathbb{R}^n : \|x\|_2 \leq 1\big\}.$$

(a) Given any subset $A \subseteq \mathbb{B}_2^n$, show that

$$\frac{1}{2}\|a + b\|_2 \leq 1 - \frac{\varepsilon^2}{8} \quad \text{for all } a \in A \text{ and } b \in (A^\epsilon)^c.$$

To be clear, we define $(A^\epsilon)^c = \mathbb{B}_2^n \setminus A^\epsilon$.

(b) Use the Brunn–Minkowski inequality (3.45) to show that

$$\mathbb{P}[A]\big(1 - \mathbb{P}[A^\epsilon]\big) \leq \Big(1 - \tfrac{\varepsilon^2}{8}\Big)^{2n}.$$

(c) Conclude that

$$\alpha_{\mathbb{P},(\mathcal{X},\rho)}(\varepsilon) \leq 2e^{-\frac{n\varepsilon^2}{4}} \quad \text{for } \mathcal{X} = \mathbb{B}_2^n \text{ with } \rho(\cdot) = \|\cdot\|_2.$$

Reading Material

Read the paper *The Curse of Concentration in Robust Learning: Evasion and Poisoning Attacks from Concentration of Measure* and answer the following problems.

Note: For simplicity and brevity, we have omitted the full definitions and notations from the paper.

Problem 7: Paper Summarization *(Mandatory)* *[15 points]*

In this problem, you should write a clear and concise overview of the paper in **no more than two pages**. Please ensure that your summary:

- **Highlights the paper's main goals and contributions.** What are the authors aiming to achieve, and why is it important?
- **Describes the key technical ideas and methods.** How do the authors approach the problem? What techniques or frameworks do they employ?
- **Discusses the significance and implications of the results.** Why are these findings impactful, and how might they influence future research or applications?

Above all, your summary should be well-structured, straightforward, and demonstrate a thorough understanding of the paper's content.

Problem 8: Concentration and Adversarial Risk *[15 points]*

(a) Prove the following lemma:

Lemma 1 *Let $\mu \equiv \mu_1 \times \cdots \times \mu_n$ be a product probability measure of dimension n, and let $f : Supp(\mu) \to \mathbb{R}$ be a measurable function such that $|f(x) - f(y)| \leq 1$ whenever x and y differ in only one coordinate. If $a = \mathbb{E}_{x\sim\mu}[f(x)]$, then*

$$\Pr_{x\sim\mu}\left[f(x) \leq a - b\right] \leq e^{-2b^2/n}.$$

(b) How do the authors prove that the adversarial risk can be large for any learning problem over concentrated spaces?

(c) Prove the following lemma and explain its usages in the paper.

Lemma 2 *For a nice metric probability space (X, d, μ), let $E \subseteq X$ be a Borel set. If $\rho = Risk_\ell(E)$, then we have*

$$Rob_\rho(E) = \rho \cdot \ell - \int_0^\ell Risk_z(E)\,dz.$$

Problem 9: Analyzing a New Concentrated Space *[10 points]*

Let $\{(X_n, d_n, \mu_n)\}_{n\in\mathbb{N}}$ be a family of metric probability spaces such that:

1. The diameter $\mathrm{Diam}(X_n)$ is $\Theta(1)$.
2. There are universal constants $k_1, k_2 > 0$ and a real $p < 2$ for which

$$\alpha_n(b) = 1 - \inf_{\mu_n(S)\geq 1/2} \mu_n\big(S^b\big) \leq k_1 \exp\Big(-k_2 b^p n\Big),$$

   for all $b \geq 0$, where S^b is the b-enlargement of S under d_n.

Consider a classification problem $\big(X_n, Y_n, \mu_n, C_n, H_n, d_n\big)$ where each $h_n \in H_n$ and $c_n \in C_n$ is a hypothesis and ground-truth concept. Suppose

$$\varepsilon_n = \Pr_{x\leftarrow\mu_n}\big[h_n(x) \neq c_n(x)\big].$$

Show or conjecture how the usual adversarial-risk theorems must be modified if we now have $\exp\big(-k_2 b^p n\big)$-type concentration. Derive an upper bound on $\mathrm{Risk}_b(h_n, c_n)$ and on the target-error robustness $\mathrm{Rob}_\rho(h_n, c_n)$ (or argue that they are unbounded).

Problem 10: Poisoning Attack with Mixed Constraints *[10 points]*

Let (X, Y, μ, H, C) be a classification problem with a (deterministic) learner L. We define a poisoning adversary A that receives the entire training set $T \in (X \times Y)^m$. We want A to satisfy two following properties:

1. Average budget:

$$\mathbb{E}_{T \leftarrow (\mu, c(\mu))^m}\Big[\mathrm{HD}\big(A(T), T\big)\Big] \le B_{\mathrm{avg}} \quad \text{for any } c \in C.$$

2. High-probability worst-case budget:

$$\Pr_{T \leftarrow (\mu, c(\mu))^m}\Big[\mathrm{HD}\big(A(T), T\big) \le B_{\mathrm{wc}}\Big] \ge 1 - \beta \quad \text{for any } c \in C.$$

Use this adversary to degrade $\mathrm{Conf}_A(m, c, \varepsilon)$ or $\mathrm{Err}_A(m, c, x)$, as in Theorems 4.5 or 4.8. In other words, show an adversary that forces high misclassification rates on the final hypothesis (or drastically lowers the confidence) while obeying both the average and the high-probability tampering constraints.