CS 839 Probability and Learning in High Dimension	Lecture 3 - 02/02/2022

Lecture 3: Title of Lecture

Lecturer: Yudong Chen — *Scribe: Student A, Student B*

In this lecture,[1] we will introduce ... We will also ...

1 Notation

A quick summary of the notation.

1. **Random variables:** X, Y, U, V
2. **Ranges/alphabets:** $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{V}$
3. **Specific values:** x, y, u, v

For a vector $u \in \mathbb{R}^d$, we use $\|u\|_2$ to denote its ℓ_2 norm and $\|u\|_\infty$ its ℓ_∞ norm. For a matrix $A \in \mathbb{R}^{d_1 \times d_2}$, we use $\|A\|_{\mathrm{F}}$ to denote its Frobenius norm and $\|A\|_{\mathrm{op}}$ its operator/spectral norm (i.e., the largest singular value of A). For two matrices A, B of the same dimension, $\langle A, B \rangle := \mathsf{tr}(A^\top B)$ denotes their trace inner product. The trace inner product reduces to the usual inner product between vectors for when $A, B \in \mathbb{R}^{d \times 1}$.

2 Preliminaries

Lemma 1 (Jensen's Inequality)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ denote a* convex *function, and X denote any random variable. We have*

$$\mathbb{E}[g(X)] \ge g(\mathbb{E}[X]).$$

Further, if g is strictly convex, equality holds iff X is deterministic.

A variable X is σ^2-sub-Gaussian if

$$\mathbb{E}\, e^{\lambda(X - \mathbb{E}[X])} \le e^{\lambda^2 \sigma^2/2}, \qquad \text{for all } \lambda \in \mathbb{R}. \tag{1}$$

and is called (σ^2, b)-sub-exponential if

$$\mathbb{E}\, e^{\lambda(X - \mathbb{E}[X])} \le e^{\lambda^2 \sigma^2/2}, \qquad \text{for all } \lambda \in \mathbb{R} \text{ with } |\lambda| \le \frac{1}{b}. \tag{2}$$

3 Hoeffding

Theorem 1 (Hoeffding's Inequality)**.** *If X_i's are independent σ_i^2-sub-Gaussian RV variables, then for each $t \ge 0$,*

$$\mathbb{P}\left[\Big| \sum_i (X_i - \mathbb{E}\, X_i) \Big| \ge t \right] \le 2 \exp\left[-\frac{t^2}{2 \sum_i \sigma_i^2} \right].$$

Proof This is the proof. □

[1] *Reading:* Sec 3.1 of Wainwright book. Also relevant: Duchi notes Sec 3.3 and Vershynin HDP book Sec 5.

4 Optimization

Consider the problem

$$\begin{array}{ll} \text{maximize} & x^\top A x \\ \text{subject to} & x^\top B x \leq 1, \\ & \|x\|_2 \leq 1. \end{array}$$

Let $y^* := \arg\min_y f(y)$.