Name: Vo Ngoc Bich Uyen

HOME ASSIGNMENT 4

Points for problems: 1, 1, 1, 1, 1
Total: 5

Problem 1: Let $V \subset \mathbb{R}^n$ be a finite set of vectors of cardinality N. Show that

$$\mathbb{E}_\varepsilon \max_{v \in V} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i v_i \right\} \leq \frac{c}{n} \sqrt{2 \log N} \quad \text{where} \quad c^2 = \max_{v \in V} \sum_{i=1}^{n} v_i^2$$

→ Proof:

+) Let $\mu = \mathbb{E}_\varepsilon \max_{v \in V} \left\{ \sum_{i=1}^{n} \varepsilon_i v_i \right\}$

+) For any $\lambda > 0$, we have $e^{\lambda \mu} \underset{\text{Jensen}}{\leq} \mathbb{E}_\varepsilon \left\{ e^{\lambda \max_{v \in V} \sum_{i=1}^{n} \varepsilon_i v_i} \right\} = \mathbb{E}_\varepsilon \max_{v \in V} \left\{ e^{\lambda \sum_{i=1}^{n} \varepsilon_i v_i} \right\} \leq \mathbb{E}_\varepsilon \left\{ \sum_{v \in V} e^{\lambda \sum_{i=1}^{n} \varepsilon_i v_i} \right\} =$

$$= \sum_{v \in V} \mathbb{E}_\varepsilon \left\{ e^{\lambda \sum_{i=1}^{n} \varepsilon_i v_i} \right\} = \sum_{v \in V} \prod_{i=1}^{n} \mathbb{E}_\varepsilon e^{\lambda \varepsilon_i v_i} = \sum_{v \in V} \prod_{i=1}^{n} \frac{e^{\lambda v_i} + e^{-\lambda v_i}}{2} \leq \left[\frac{e^u + e^{-u}}{2} \leq e^{\frac{u^2}{2}} \right] \leq \sum_{v \in V} \prod_{i=1}^{n} e^{\frac{\lambda^2 v_i^2}{2}} =$$

$$= \sum_{v \in V} e^{\frac{\lambda^2 \|v\|^2}{2}} \leq |N| e^{\frac{\lambda^2 c^2}{2}}$$

$$\Rightarrow \mu \leq \frac{\log |N|}{\lambda} + \frac{\lambda c^2}{2} \qquad (\lambda > 0)$$

+) Take $\lambda = \sqrt{\dfrac{2 \log |N|}{c^2}}$, we have $\mu \leq c \sqrt{2 \log N}$

$$\Rightarrow \mathbb{E}_\varepsilon \max_{v \in V} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i v_i \right\} \leq \frac{c}{n} \sqrt{2 \log N} \qquad \square$$

Problem 2: Let $V \subseteq \mathbb{R}^n$ and let $\mathrm{conv}(V)$ stand for its convex hull, that is

$$\mathrm{conv}(V) = \left\{ \sum_{j=1}^{N} \lambda_j v^{(j)} : N \in \mathbb{N}, \lambda_j \geqslant 0, \sum_{j=1}^{N} \lambda_j = 1, v^{(j)} \in V \right\}$$

Prove that,

$$\mathbb{E}_{\varepsilon} \max_{v \in V} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i v_i \right\} = \mathbb{E}_{\varepsilon} \max_{v \in \mathrm{conv}(V)} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i v_i \right\}$$

$\longrightarrow$ Proof:

+) Maximizing a linear function over the covex hull always yields a vertex, which is a point in original set V., i.e. if v is a vector, then $\max_j v_j = \sup_{\substack{\lambda \geqslant 0 \\ \sum_{j=1}^{N} \lambda_j = 1}} \sum_j \lambda_j v_j$

+) $$\mathbb{E}_{\varepsilon} \max_{v \in \mathrm{conv}(V)} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i v_i \right\} = \frac{1}{n} \cdot \mathbb{E}_{\varepsilon} \sup_{\lambda} \sup_{\lambda^{(1)}, \dots, \lambda^{(N)}} \sum_{i=1}^{n} \varepsilon_i \sum_{j=1}^{N} \lambda_j v_i^{(j)} =$$

$$= \frac{1}{n} \cdot \mathbb{E}_{\varepsilon} \sup_{\lambda} \sum_{j=1}^{N} \lambda_j \sup_{\lambda^{(j)}} \sum_{i=1}^{n} \varepsilon_i v_i^{(j)} = \frac{1}{n} \mathbb{E}_{\varepsilon} \sup_{\lambda} \sum_{i=1}^{n} \varepsilon_i v_i = \mathbb{E}_{\varepsilon} \max_{v \in V} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i v_i \right\} \quad \square$$

Problem 3: Let F be a class of functions taking their values in $\{0, 1\}$ and let $\ell(Y, Y') = \mathbb{1}(Y \neq Y$ be the binary loss function. Show that

$$\mathbb{E}_{\varepsilon} \max_{f \in F} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \ell(Y_i, f(X_i)) \right\} \leqslant \mathbb{E}_{\varepsilon} \max_{f \in F} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(X_i) \right\}$$

$\longrightarrow$ Proof:

$$\frac{1}{2}\tilde{f}(x_i) = f(x_i) - \frac{1}{2}$$

+) We have $f : \mathcal{X} \to \{0,1\}$, let define $\tilde{f} : \mathcal{X} \to \{-1,1\}$ by $\tilde{f}(x_i) = 2f(x_i) - 1$ ($\tilde{F}$)

$\tilde{Y}_i = 2Y_i - 1$

+) $$\mathbb{E}_\varepsilon \max_{\tilde{f}\in\tilde{F}} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i \mathbb{1}\{\tilde{f}(x_i) \neq \tilde{Y}_i\} \right\} = \mathbb{E}_\varepsilon \max_{\tilde{f}\in\tilde{F}} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i \cdot \frac{1-\tilde{Y}_i\tilde{f}(x_i)}{2} \right\} =$$

$$= \mathbb{E}_\varepsilon \left\{ \frac{1}{2n}\sum_{i=1}^n \varepsilon_i + \frac{1}{2}\max_{\tilde{f}\in\tilde{F}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i \cdot (-\tilde{Y}_i)\tilde{f}(x_i) \right\} = \left[\begin{array}{l} \mathbb{E}_\varepsilon[\varepsilon_i] = 0 \\ \varepsilon_i \text{ and } \varepsilon_i(-\tilde{Y}_i) \\ \text{have the same} \\ \text{distribution} \end{array} \right] = \frac{1}{2}\mathbb{E}_\varepsilon \max \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i \tilde{f}(x_i) \right\}$$

$$\Rightarrow \mathbb{E}_\varepsilon \max_{f\in F} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i \ell(Y_i, f(x_i)) \right\} = \mathbb{E}_\varepsilon \max_{\tilde{f}\in\tilde{F}} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i \ell(\tilde{Y}_i, \tilde{f}(x_i)) \right\} = \frac{1}{2}\mathbb{E}_\varepsilon \max_{\tilde{f}\in\tilde{F}} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i \tilde{f}(x_i) \right\}$$

$$= \mathbb{E}_\varepsilon \max_{f\in F} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(x_i) \right\} - \frac{1}{2}\mathbb{E}_\varepsilon \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i \right\} \leq \mathbb{E}_\varepsilon \max_{f\in F} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(x_i) \right\} \qquad \square$$

Problem 4: Let $F = \{ f(x) = w^T x : w \in \mathbb{R}^d, \|w\| \leq R \}$. Prove that

$$\mathbb{E}_\varepsilon \max_{f\in F} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(x_i) \right\} \leq \frac{R}{n}\sqrt{\sum_{i=1}^n \|X_i\|^2}$$

$\longrightarrow$ Proof:

$$\mathbb{E}_\varepsilon \max_{f\in F} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i) \right\} = \mathbb{E}_\varepsilon \max_{f\in F} \left\{ \frac{1}{n}\sum_{i=1}^n \varepsilon_i w^T X_i \right\} = \mathbb{E}_\varepsilon \max_{\|w\|\leq R} \left\{ w^T \cdot \frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i \right\} \leq$$

$$\leq \underset{\text{Cauchy-Schwartz}}{\left[\max_{\|w\|\leq R} (w^T u) \leq \|w\|.\|u\| \right]} \leq R \cdot \mathbb{E}_\varepsilon \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i \right\| \right\} \overset{\text{Jensen}}{\leq} R \cdot \mathbb{E}_\varepsilon \left\{ \left\| \frac{1}{n}\sum_{i=1}^n \varepsilon_i X_i \right\|^2 \right\}^{\frac{1}{2}} =$$

$$= \frac{R}{n} \mathbb{E}_\varepsilon \left\{ \sum_{i=1}^n \varepsilon_i^2 \|X_i\|^2 + \sum_{\substack{i,j=1 \\ i\neq j}}^n \varepsilon_i \varepsilon_j X_i^T X_j \right\}^{\frac{1}{2}} = \frac{R}{n}\sqrt{\sum_{i=1}^n \|X_i\|^2} \qquad \square$$

Được quét bằng CamScanner

Problem 5: Let $c > 0$. Prove that

$$\mathbb{E}_{\varepsilon} \max_{w \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i w^T X_i - c\|w\|^2 \right\} \leq \frac{1}{4cn^2} \sum_{i=1}^{n} \|X_i\|^2$$

→ Proof:

+) $$\mathbb{E}_{\varepsilon} \max_{w \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i w^T X_i \right\} = \mathbb{E}_{\varepsilon} \max_{w \in \mathbb{R}^d} \left\{ w^T \cdot \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i X_i \right\} \leq \left[\begin{array}{l} \max u^T v \leq \|u\| . \|v\| \\ u = \sqrt{2c} . w \\ v = \frac{1}{\sqrt{2c}} \cdot \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i X_i \end{array} \right] \leq$$

$$\leq \sqrt{2c} \|w\| . \mathbb{E}_{\varepsilon} \left\| \frac{1}{n\sqrt{2c}} \cdot \sum_{i=1}^{n} \varepsilon_i X_i \right\| \leq \left[u.v \leq \frac{u^2 + v^2}{2} \right] \leq \frac{2c\|w\|^2}{2} + \frac{1}{2} \mathbb{E}_{\varepsilon} \left\| \frac{1}{n\sqrt{2c}} \cdot \sum_{i=1}^{n} \varepsilon_i X_i \right\|^2$$

$$= c\|w\|^2 + \frac{1}{4cn^2} \cdot \mathbb{E}_{\varepsilon} \left\{ \sum_{i=1}^{n} \varepsilon_i^2 \|X_i\|^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \varepsilon_i \varepsilon_j X_i^T X_j \right\}^2 = c\|w\|^2 + \frac{1}{4cn^2} \cdot \sum_{i=1}^{n} \|X_i\|^2$$

$$\Rightarrow \mathbb{E}_{\varepsilon} \max_{w \in \mathbb{R}^d} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i w^T X_i - c\|w\|^2 \right\} \leq c\|w\|^2 + \frac{1}{4cn^2} \sum_{i=1}^{n} \|X_i\|^2 - c\|w\|^2 = \frac{1}{4cn^2} \sum_{i=1}^{n} \|X_i\|^2 \quad \square$$