清華大學
TSINGHUA UNIVERSITY

Machine Learning
10/23/2020

1. $\sup_{z_1 \cdots z_n, z_i'} \left| \hat{R}_{S_n}(\mathcal{G}) - \hat{R}_{S_n'}(\mathcal{G}) \right|$

$= \sup_{z_1 \cdots z_n, z_i'} \left| E_\sigma \sup_{g\in\mathcal{G}} \frac{1}{n}\sum_{j=1}^{n} \sigma_j g(z_j) - E_\sigma \sup_{g\in\mathcal{G}} \frac{1}{n}\left(\sum_{j=1}^{n}\sigma_j g(z_j) + \sigma_i\left(g(z_i') - g(z_i)\right)\right) \right|$

$= \sup_{z_1 \cdots z_n, z_i'} \left| E_\sigma \left(\underbrace{\sup_{g\in\mathcal{G}} \frac{1}{n}\sum_{j=1}^{n}\sigma_j g(z_j)}_{A} - \sup_{g\in\mathcal{G}} \underbrace{\frac{1}{n}\left(\sum_{j=1}^{n}\sigma_j g(z_j) + \sigma_i\left(g(z_i') - g(z_i)\right)\right)}_{A \quad + \quad B} \right) \right|$

$\sup(A+B) \le \sup A + \sup B$

$\le \sup_{z_1 \cdots z_n, z_i'} \left| E_\sigma \sup_{g\in\mathcal{G}} \frac{1}{n} \sup_{z_i, z_i' \in S} \left| g(z_i') - g(z_i) \right| \right|$

$\le \frac{1}{n} \sup_{\substack{\sigma\in\{1,-1\} \\ g\in\mathcal{G} \\ z\in D}} \left| \sigma' g'(z') - \sigma g(z) \right|$ $\quad \because \sigma \in \{-1, 1\}$, $\forall g\in\mathcal{G},\ 0 \le g(z) \le 1$,

$\therefore$ 原式 $\le \frac{1}{n}$

代入 mcDiarmid 不等式,有

$$P\left(\left|\hat{R}_{S_n}(\mathcal{G}) - E_{S_n \sim D^n} \hat{R}_{S_n}(\mathcal{G})\right| \ge \varepsilon\right)$$

$$= P\left(\left|\hat{R}_{S_n}^{n}(\mathcal{G}) - R_n(\mathcal{G})\right| \ge \varepsilon\right) \le 2e^{\frac{-2\varepsilon^2}{\sum_{i=1}^{n}\frac{1}{n^2}}} = 2e^{-2n\varepsilon^2}$$

即 $\delta \le 2e^{-2n\varepsilon^2}$, $\varepsilon \le \sqrt{\frac{\log\frac{2}{\delta}}{2n}}$

$R_n(\mathcal{G}) \le \hat{R}_{S_n}(\mathcal{G}) + O\left(\sqrt{\frac{\log\frac{1}{\delta}}{n}}\right)$,得证

(2) $R_n(H) = \underset{S\sim D^n}{E} E_\sigma \sup_{h\in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)$

(a) 右边: $|\alpha| R_n(H) = |\alpha| \underset{S\sim D^n}{E} E_\sigma \sup_{h\in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)$

左边: $R_n(\alpha H) = \underset{S\sim D^n}{E} E_\sigma \sup_{h\in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i \alpha h(z_i)$

若 $\alpha \geq 0$:

$\sup_{h\in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i \alpha h(z_i)$ 仍在原先的 h 取最大值,

$= \alpha \sup \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)$, 根据期望性质, 有 $R_n(\alpha H)$

$= \alpha \underset{S\sim D^n}{E} E_\sigma \sup \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i) =$ 右边

若 $\alpha < 0$:

$E_\sigma \sup_{h\in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i \alpha h(z_i)$

$= \frac{1}{2^n}\sum_{j=1}^{2^n} \sup \frac{1}{n}\sum_{i=1}^{n} \sigma_{ji} \cdot -|\alpha| h(z_i)$

设对 $\forall j \in [1, 2^n]$,

$\exists\, \sigma_{nj}$ 与 σ_j 所有值都完全相反,

则根据 σ 对称性,

有 原式 $= \frac{1}{2^n}\sum_{j=1}^{2^n} \sup \frac{1}{n}\sum_{i=1}^{n} \sigma_{nji} |\alpha| h(z_i)$

带 sup 的式子, 在 σ 为 σ_{nj} 时, 与原先式子 σ 为 σ_j 时取最大值的 h 一致,

有原式 $= |\alpha| \frac{1}{2^n}\sum_{j=1}^{2^n} \frac{1}{n}\sum_{i=1}^{n} \sigma_{nji} h(z_i)$

根据期望性质, 有 $R_n(\alpha H)$

$= |\alpha| \underset{S\sim D^n}{E} E_\sigma \sup \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)$ = 右边

得证

(b)

左边 $R_n(H+H')$

$= \underset{S\sim D^n}{E} E_\sigma \sup_{h\in H, h'\in H'} \frac{1}{n}\sum_{i=1}^{n} \sigma_i (h(z_i)+h'(z_i))$

$\because h$ 与 h' 独立

$\therefore = \underset{S\sim D^n}{E} E_\sigma \left(\sup_{h\in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i) + \sup_{h'\in H'} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h'(z_i)\right)$

根据期望性质

$= \underset{S\sim D^n}{E} E_\sigma \sup_{h\in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h(z_i)$

$+ \underset{S\sim D^n}{E} E_\sigma \sup_{h'\in H} \frac{1}{n}\sum_{i=1}^{n} \sigma_i h'(z_i)$

$= R_n(H) + R_n(H')$

得证

3. ~~根据 Hoeffding 不等式~~.

$$\sum_{i=1}^{n}(X_i - EX_i) = n\left\{\left[\frac{1}{n}\sum_{i=1}^{n} l(h(x_i), y_i)\right] - E_{(x,y)\sim p}\, l(h(x), y)\right\}$$

$$= n\left(\hat{\varepsilon}_{D_n}(h) - \varepsilon(h)\right)$$

根据 Hoffeding 不等式，有

$$P\left(n\left|\hat{\varepsilon}_{D_n}(h) - \varepsilon(h)\right| \geq \varepsilon\right) \leq 2e^{-\frac{2\varepsilon^2}{n}}$$

将 ε 替为 $n\varepsilon$，有

$$P\left(\left|\hat{\varepsilon}_{D_n}(h) - \varepsilon(h)\right| \geq \varepsilon\right) \leq 2e^{-2n\varepsilon^2},$$

即 $\delta p(h) \leq 2e^{-2n\varepsilon^2}$,

有 $\varepsilon \leq \sqrt{\dfrac{\log 2 + \log\frac{1}{p(h)} + \log\frac{1}{\delta}}{2n}}$,

即

$$\varepsilon(h) \leq \hat{\varepsilon}_{D_n}(h) + O\left(\sqrt{\frac{\log\frac{1}{p(h)} + \log\frac{1}{\delta}}{n}}\right)$$

得证