Bounded differences

We want to show that

$$\hat{\theta} \in \operatorname*{argmin}_{\theta \in \Theta} \frac{1}{n}\sum \ell(\theta; z_i)$$

achieves near-optimal population loss.

To show such results, we will use uniform concentration guarantees

$$\sup_{\theta \in \Theta} \left| \frac{1}{n}\sum \ell(\theta; z_i) - \mathbb{E}\,\ell(\theta; z) \right| \to 0 \quad \text{fast enough}.$$

We begin with concentration results for light-tailed RVs.

Def A RV X is σ^2-subGaussian if $\mathbb{E} e^{\lambda(X - \mathbb{E}X)} \le \underline{\exp\left(\frac{\sigma^2}{2}\lambda^2\right)}$ $\forall \lambda \in \mathbb{R}$

↳ Light-tailed RV ↳ MGF of $N(0, \sigma^2)$

From Markov's inequality, $\forall \lambda \ge 0$

$$\mathbb{P}(X - \mathbb{E}X \ge t) = \mathbb{P}(\lambda(X - \mathbb{E}X) \ge \lambda t) = \mathbb{P}(e^{\lambda(X-\mathbb{E}X)} \ge e^{\lambda t}) \le e^{-\lambda t}\,\mathbb{E} e^{\lambda(X - \mathbb{E}X)}$$
$$\le \exp\left(\frac{\sigma^2}{2}\lambda^2 - \lambda t\right)$$

Taking min over $\lambda \ge 0$, we get $\mathbb{P}(X - \mathbb{E}X \ge t) \le \exp\left(-\frac{t^2}{2\sigma^2}\right)$.

Similarly, we have $\mathbb{P}(X - \mathbb{E}X \le -t) \le \exp\left(-\frac{t^2}{2\sigma^2}\right)$.

Example ε : random signs (Rademacher) is 1-sub-Gaussian.

$$\mathbb{E} e^{\lambda\varepsilon} = \frac{1}{2}(e^{-\lambda} + e^{\lambda}) = \frac{1}{2}\left(\sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) = \sum_{k \ge 0} \frac{\lambda^{2k}}{(2k)!} \le 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{(2k)!}$$
$$\le 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{2^k k!} = e^{\frac{\lambda^2}{2}}$$

Example $X \in [a, b]$, $\mathbb{E}X = 0$. Then, X is $(b-a)^2$-sub-Gaussian.

Pf) $$\mathbb{E}_X e^{\lambda X} = \mathbb{E}_X e^{\lambda(X - \mathbb{E}_{X'} X')} \underset{\text{Jensen}}{\le} \mathbb{E} e^{\lambda(X - X')}$$ where X' indep copy of X

Let ε be random signs indep of everything so that $X - X' \overset{D}{=} \varepsilon(X - X')$

$$\mathbb{E} e^{\lambda(X - X')} = \mathbb{E}_{X,X'} \mathbb{E}_{\varepsilon}\, e^{\lambda\varepsilon(X - X')} \le \mathbb{E}_{X,X'}\, e^{\frac{\lambda^2}{2}(X - X')^2} \quad \text{by Ex 1}$$
$$\le e^{\frac{\lambda^2}{2}(b-a)^2}$$

Actually, we can show X is $\frac{(b-a)^2}{4}$ - sub-G.

cf. $X \in [a,b]$ is $\frac{(b-a)^2}{4}$ - sub-Gaussian.

By convexity of $x \mapsto e^{\lambda x}$, $e^{\lambda x} \le \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}$

Take expectations on both sides. For $h = \lambda(b-a)$, $p = \frac{-a}{b-a}$. $L(h) = -hp + \log(1-p+pe^h)$,

$\mathbb{E} e^{\lambda X} \le e^{L(h)}$. $L(0) = L'(0) = 0$, $L''(h) \le \frac{1}{4}$ $\forall h$ so $L(h) \le \frac{1}{8} h^2$ by Taylor ⊠.

Bounded differences will play a key role in showing this result.

Thm Let g be a function satisfying

one coordinate doesn't change ftn too much

$$|g(z_1, \cdots z_i, \cdots, z_n) - g(z_1, \cdots, z_i', \cdots, z_n)| \le c_i \quad \forall\, 1 \le i \le n$$

For independent RVs Z_i's,

$$\mathbb{P}\left(g(Z_1^n) - \mathbb{E} g(Z_1^n) \ge t \right) \le \exp\left(-\frac{2t^2}{\sum_1^n c_i^2}\right).$$

Generalization of Hoeffding bound

Def $\{M_i\}_{i=0}^n$ is a martingale seq w.r.t. RVs $Z_1, \cdots, Z_n$

if M_i is $(Z_1, \cdots, Z_i)$-measurable, $\mathbb{E}|M_i| < \infty$, and

$\mathbb{E}[M_i \mid Z_{i-1}] = M_{i-1}$.

We call $\{D_i := M_i - M_{i-1}\}_{i=1}^n$ a martingale difference sequence w.r.t. Z_1^n.

$(\mathbb{E}[D_i \mid X_1^{i-1}] = 0)$

Lemma Let D_i be a martingale difference sequence w.r.t. Z_1^n s.t. $\exists\, \sigma_i^2$

$$\mathbb{E}[e^{\lambda D_i} \mid Z_1^{i-1}] \le \exp\left(\frac{\sigma_i^2 t^2}{2}\right) \quad \forall i \quad \cdots (\star)$$

Then, $M_n - M_0 = \sum_{i=1}^n D_i$ is $(\sum \sigma_i^2)$ - sub-Gaussian.

Pf

$$\mathbb{E} e^{\lambda \sum D_i} = \mathbb{E}\, e^{\lambda D_n} \cdot e^{\lambda \sum_1^{n-1} D_i} = \mathbb{E}\left[\mathbb{E}\left[e^{\lambda D_n} \cdot \underline{e^{\lambda \sum_1^{n-1} D_i}} \mid Z_1^{n-1}\right]\right]$$

$$\le \exp\left(\frac{\sigma_n^2 t^2}{2}\right) \cdot \mathbb{E}\left[e^{\lambda \sum_1^{n-1} D_i}\right]$$

↳ Z_1^{n-1} - measurable

By induction, we get the result. ⊠.

Proof of bdd differences

Define the Doob martingale $M_i = \mathbb{E}[g(Z_1^n) \mid Z_1^i]$ $\quad \left(\begin{array}{l} M_0 = \mathbb{E}\, g(Z_1^n) \\ M_n = g(Z_1^n) \end{array} \right)$

So we'd like to bound $\mathbb{P}(M_n - M_0 \ge t)$.

Note that $|D_i| = \left| \mathbb{E}[g(Z_1^n) \mid Z_1^i] - \mathbb{E}[g(Z_1^n) \mid Z_1^{i-1}] \right|$

$$\le \sup_{z,z'} \left| \mathbb{E}_{Z_{i+1}^n}\, g(Z_1^{i-1}, z, Z_{i+1}^n) - \mathbb{E}_{Z_{i+1}^n}\, g(Z_1^{i-1}, z', Z_{i+1}^n) \right| \le c_i \quad . \quad \cdots (*)$$

So $\mathbb{E}[e^{\lambda D_i} \mid Z_1^{i-1}] = \mathbb{E}[e^{\lambda(D_i - \mathbb{E}[D_i \mid Z_1^{i-1}])} \mid Z_1^{i-1}] \le \exp\left(\frac{\lambda^2}{2} \cdot \frac{c_i^2}{4} \right)$

From previous lemma, and tail inequality for sub-G RVs, we have the result $\boxtimes$.