10/16/23

Lecture 11: Halfspaces w/ Massart Noise

$$\text{LeakyReLU}_\lambda(z) = \begin{cases} (1-\lambda) z & \text{if } z \geq 0 \\ \lambda z & \text{if } z < 0 \end{cases}$$

Note: $\text{LeakyReLU}_\lambda(z) = \frac{1}{2}\left[z + (1-2\lambda)|z| \right]$

LeakyReLU loss:

$$\ell(w; x, y) \triangleq \text{LeakyReLU}_\lambda(-y\langle w, x\rangle)$$

$$\ell(w) \triangleq \mathbb{E}_{x,y}\, \ell(w, x, y)$$

True test loss:

$$\text{err}(w; x, y) \triangleq \mathbb{1}\left[y \neq \text{sgn}(w, x) \mid x \right]$$

$$\text{err}(w) \triangleq \mathbb{E}_{x,y}\left[\text{err}(w; x, y)\right] = \Pr_{x,y}\left[y \neq \text{sgn}(\langle w, x\rangle)\right]$$

Under Massart noise,

$$\mathbb{E}_{y|x}\left[\text{err}(w^*; x, y) | x\right] = \eta(x) \leq \eta \quad \forall x \qquad (*)$$

Lemma 1 (w^* achieves small loss): If $\lambda \geq \eta + \varepsilon$

$$\ell(w^*) \leq -\tau(\lambda - \text{err}(w^*))$$

margin

Pf: $\ell(w^*) = \mathbb{E}_{x,y}\Big[\overbrace{\mathbb{1}[y = \text{sgn}(\langle w^*, x\rangle)]}^{1-\text{err}(w^*; x, y)} \cdot \lambda \cdot (-|\langle w^*, x\rangle|)$

$$\underbrace{\mathbb{1}[y \neq \text{sgn}(\langle w^*, x\rangle)]}_{\text{err}(w^*; x, y)} \cdot (1-\lambda) \cdot |\langle w^*, x\rangle| \Big]$$

$$= \mathbb{E}_{x}\left[\left(\underbrace{\mathbb{E}_{y|x}\left[err(w^*;x,y)\,|\,x\right]}_{\le \eta\ \forall x \text{ by } (*)}-\lambda\right)\cdot \underbrace{|\langle w^*,x\rangle|}_{\ge \tau}\right]$$

(b/c $\lambda \ge \eta$)

$$\le -\tau\left(\lambda - err(w^*)\right) \le -\tau\varepsilon . \quad \square$$

In fact, Lemma 1 also holds under arbitrary conditionings of D, e.g for any event $\mathcal{E}$,

$$\mathbb{E}_{x|\mathcal{E}}\left[\left(\underbrace{\mathbb{E}_{y|x}\left[err(w^*;x,y)\,|\,x\right]}_{\le \eta\ \forall x \text{ by } (*)}-\lambda\right)\cdot \underbrace{|\langle w^*,x\rangle|}_{\ge \tau}\,\Big|\,\mathcal{E}\right]$$

$$\le -\tau\left(\lambda - \mathbb{E}_{x,y|\mathcal{E}}\left[err(w;x,y)\,|\,\mathcal{E}\right]\right)$$

$$\le -\tau(\lambda-\eta) \le -\tau\varepsilon.$$

(this is key distinction between Massart and agnostic)

Lemma 2: If w satisfies $err(w) \ge \lambda$, then there is some slab $\{x : |\langle w,x\rangle| \le \gamma\}$ s.t.

$$\underbrace{L^{\gamma}(w)}_{\text{recall, } = \mathbb{E}_{x,y}\left[\ell(w;x,y)\,|\,|\langle w,x\rangle| \le \gamma\right]} \ge 0$$

$\square$

Pf: Suppose to the contrary there is

no such slab. So $\forall\, \gamma > 0$,

$$0 > \mathop{\mathbb{E}}_{x,y}\left[\ell(w; x, y)\, \mathbb{1}\left[|\langle w, x\rangle| \le \gamma \right] \right]$$

$$= \mathop{\mathbb{E}}_{x,y}\left[\left\{ \left(1 - \mathrm{err}(w; x, y)\right) \cdot \lambda \cdot \left(-|\langle w, x\rangle|\right) + \mathrm{err}(w; x, y) \cdot (1-\lambda) \cdot |\langle w, x\rangle| \right\} \mathbb{1}\left[|\langle w, x\rangle| \le \gamma \right] \right]$$

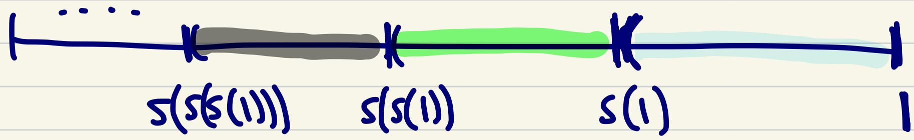
$$= \mathop{\mathbb{E}}_{x}\left[\Big(\underbrace{\mathop{\mathbb{E}}_{y|x}\left[\mathrm{err}(w; x, y) \,|\, x \right]}_{\triangleq\, \mathrm{err}(w; x)} - \lambda \Big) \cdot |\langle w, x\rangle|\, \mathbb{1}\left(|\langle w, x\rangle| \le \gamma \right) \right]$$

$$|\langle w, x\rangle| = \int_0^\infty \mathbb{1}\left[s < |\langle w, x\rangle| \right] ds$$

$$= \int_0^\infty \mathop{\mathbb{E}}_{x}\left[\left(\mathrm{err}(w; x) - \lambda \right) \mathbb{1}\left[s < |\langle w, x\rangle| \le \gamma \right] \right] ds$$

i.e. by averaging argument, implies that $\forall\, \gamma$, exists $s(\gamma) < \gamma$ s.t.

$$0 > \mathop{\mathbb{E}}_{x}\left[\left(\mathrm{err}(w; x) - \lambda \right) \mathbb{1}\left[s(\gamma) < |\langle w, x\rangle| \le \gamma \right] \right] \qquad (\ast\ast)$$



By adding up (*) over partition of $[0,1]$ $[s(\gamma),\gamma]$ intervals, get

$$0 > \mathbb{E}_x\left[\mathrm{err}(w;x) - \lambda\right]$$
$$= \mathrm{err}(w) - \lambda,$$

so $\mathrm{err}(w) < \lambda$, contradiction! $\square$

So:

1). w^* achieves small $L^\gamma(\cdot)$ $\forall\ \gamma$
2) any w with large $\mathrm{err}(w)$ achieves large $L^\gamma(\cdot)$ for some γ,

so approx equilibrium for 2-player game

$$\min_w \max_\gamma L^\gamma(w)$$

has $\mathrm{err}(\cdot) \le \eta + o(1)$. $\square$