CS 777 : SALT If $\ell \in [0,B]$, L-Lipschitz, then for $S \sim D^n$ 31/01/18

Lecture-10.

$$\mathbb{P}_S\left(\sup_{f\in\mathcal{F}} \left|er_D^{\ell}[f] - er_S^{\ell}[f]\right| > 2\cdot L\cdot R_n(\mathcal{F}) + \epsilon\right) \le 2\exp\left(\frac{-n\epsilon^2}{2B^2}\right)$$

$$R_n(\mathcal{F}) \triangleq \mathbb{E}_{S,\varepsilon_i} \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(x_i)\right|$$

Radamacher Complexity of $\mathcal{F}$.

$\mathcal{F} = \{w : \|w\|_2 \le R\}$, $\|x\|_2 \le r$ then $R_n(\mathcal{F}) \le \frac{R\cdot r}{\sqrt{n}}$

We able to handle,

hinge loss, logistic, exp, least square, $\varepsilon$-insensitive

Not handled:

1. Regularisation.
2. classification (not Lipschitz).

(Comment).

Finding exact VC-dimension of a class is NP-hard, Same holds for $R_n(\mathcal{F})$, but we can upper-bound it.

Emperical R.A.

$S = \{x_1, x_2 \ldots x_n\}$

$$\hat{R}_{S,\hat{\varepsilon}_i}(\mathcal{F}) = \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum \hat{\varepsilon}_i f(x_i)\right|$$

Sample $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots \hat{\varepsilon}_n$.

Example: $w \in \mathcal{W}$, (not necessarily $\mathcal{B}(0,R)$)

$$\sup_{w\in\mathcal{W}} \left|\left\langle \frac{1}{n}\sum \hat{\varepsilon}_i x_i, w\right\rangle\right|$$

Exercise:- $\mathbb{P}\left(\left|\hat{R}_{S,\hat{\varepsilon}_i}(\mathcal{F}) - R_n(\mathcal{F})\right| > \varepsilon\right) \le 2\exp\left(-\frac{n\varepsilon^2}{2B^2}\right)$

holds whenever $|f(x)| \le B$.

$g : (S, \hat{\varepsilon}_i) \mapsto \hat{R}_{S,\hat{\varepsilon}_i}(\mathcal{F})$. $\rightarrow$ Show is stable.

---

Why people are interested in Gaussian Avg? (rather than Radamacher avg) ~~although symmetrisation does~~

Ans: $R_n(\mathcal{F}) \le G_n(\mathcal{F}) \le \ln n \; R_n(\mathcal{F})$.

$G_n$: Gaussian average.

## Massart's Finite class lemma:-

Bounds $R_n(\mathcal{F})$ if $|\mathcal{F}| < \infty$

Fix $|S| = n$ $\quad S \in (\mathcal{X} \times \mathcal{Y})^n$

$f \mapsto (f(x_1), \ldots, f(x_n)) \in \mathbb{R}^n$

$A = \{ a \in \mathbb{R}^n : a = (f(x_1), f(x_2) \ldots, f(x_n)) \text{ for some } f \in \mathcal{F} \}$

We want to upper bound this.

$R_S(\mathcal{F}) = \boxed{\mathbb{E}_{\varepsilon_i} \sup_{a \in A} \frac{1}{n} \sum \varepsilon_i a_i}$ , $|A| \leq |\mathcal{F}|$

$\boxed{R_n(\mathcal{F}) = \mathbb{E}_S R_S(\mathcal{F}) \leq \sup_S R_S(\mathcal{F}).}$

We will use cramer-chernoff:

$$\exp\left( s \, \mathbb{E}_{\varepsilon_i} \sup_{a \in A} \frac{1}{n} \sum \varepsilon_i a_i \right)$$

$$\leq \mathbb{E}_{\varepsilon_i} \exp\left( s \sup_a \frac{1}{n} \sum \varepsilon_i a_i \right) \quad \text{[Jensen's Ine]}$$

$$= \mathbb{E}_{\varepsilon_i} \sup_a \exp\left( \frac{s}{n} \sum \varepsilon_i a_i \right) \quad \text{(exp - is monotonous.)}$$

$$= \mathbb{E}_{\varepsilon_i} \sup_a \prod \exp\left( \frac{s \, \varepsilon_i a_i}{n} \right)$$

$$\leq \mathbb{E}_{\varepsilon_i} \sum_a \prod_i \exp\left( \frac{s \, \varepsilon_i a_i}{n} \right)$$

$$= \sum_a \prod_i \mathbb{E}_{\varepsilon_i} \exp\left( \frac{s \, \varepsilon_i a_i}{n} \right).$$

$$\leq \sum_a \prod_i \exp\left( \frac{s^2 a_i^2}{2n^2} \right) \quad \text{(Hoeffding's Lemma)}$$

$$= \sum_a \exp\left( \frac{s^2}{2n^2} \|a\|_2^2 \right)$$

$$= |A| \exp\left( \frac{s^2 c^2}{2n^2} \right) \qquad \Big| \sup_{a \in A} \|a\|_2 = c.$$

$$R_S(\mathcal{F}) \leq \frac{1}{s} \lg |A| + \frac{s c^2}{2n^2} \leq \frac{c \sqrt{2 \lg |A|}}{n}$$

Applications:

① ~~Convex~~ Sparse Models:

Let $\mathcal{F}$ be linear class of sparse models.

$$\mathcal{F} = \{ x \mapsto \langle w, x \rangle, \quad \|w\|_0 \le s, \quad \|w\|_\infty \le R \}$$
$$w, x \in \mathbb{R}^d, \quad \|x\|_\infty \le r$$

$\Rightarrow \|w\|_1 \le s \cdot R$

$$R_S(\mathcal{F}) = \mathbb{E}_{\varepsilon_i} \sup_{w \in W} \frac{1}{n} \sum \varepsilon_i \langle w, x_i \rangle$$

$$\le \mathbb{E}_{\varepsilon_i} \sup_{w \in W} \|w\|_1 \left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i x_i \right\|_\infty$$

[Hölder Inequality]

$$\boxed{\begin{array}{l} \frac{1}{p} + \frac{1}{q} = 1, \quad \|x\|_p \|y\|_q \ge |\langle x, y \rangle| \\ \text{for } p = 1 \text{ we have} \\ \left|\sum a_i b_i\right| \le \sum |a_i b_i| \quad \text{let } \|a\|_\infty = A \\ \qquad \le A \sum |b_i| = A \cdot \|b\|_1 \end{array}}$$

$$\le s \cdot R \cdot \mathbb{E}_{\varepsilon_i} \left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i x_i \right\|_\infty$$

$$= s \cdot R \cdot \mathbb{E}_{\varepsilon_i} \sup_j \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i x_i \right|_j$$

(add the supremum)

$$\boxed{\begin{array}{l} \cancel{\text{To apply}} \quad A = \{ x_1^i, x_2^i, \ldots x_n^i \} \\ \text{observe,} \quad |A| = d \end{array}}$$

$$\frac{r\sqrt{n}\sqrt{2\log d} \cdot sR}{n}$$

$$\le \frac{s\,r\,R\sqrt{2\log d}}{\sqrt{n}}$$

② Covering Number $\rightarrow$ R.A.

$\mathcal{F} \rightarrow \varepsilon$ - cover $\mathcal{C}$ $\qquad |f(x)| \le B$

$\forall f \in \mathcal{F}, \; \exists g \in \mathcal{C}, \; \forall x \in \mathcal{X}, \; |f(x) - g(x)| \le \varepsilon,$

$$R_n(\mathcal{F}) \le \varepsilon + R_n(\mathcal{C}) \le \varepsilon + B\sqrt{\frac{\lg |\mathcal{C}|}{n}} =$$

using MFCL

$$= \varepsilon + B\sqrt{\frac{d \lg(1+\frac{2B}{\varepsilon})}{n}}$$ for linear models of $L_2$ – norm at most $B$.

$$R_n(\mathcal{F}) \le \inf_{\alpha} \left\{ \alpha + B\sqrt{\frac{N_\infty(\mathcal{F},\alpha)}{n}} \right\}$$

Classification: $\xi$

$\mathcal{F} \subseteq \{-1, 1\}^{\mathcal{X}}$, $\ell = \ell^{0-1}$ (non-liptchitz).

$$\mathbb{P}_S \left( \sup_{f \in \mathcal{F}} | er_D^{0-1}[f] - er_S^{0-1}[f] | > 2R_n(\ell^{0-1} \circ \mathcal{F}) + \varepsilon \right)$$
$$\le 2\exp\left(\frac{-n\varepsilon^2}{2}\right)$$

can be proved using Mc'Diaramid; and boundedness of loss function.

TLCE:- $R_n(\ell \circ \mathcal{F}) \le L \cdot R_n(\mathcal{F})$ if loss is L-Lipchitz

$$R_n(\ell^{0-1} \circ \mathcal{F}) \le 2 R_n(\mathcal{F}).$$
if $\mathcal{F} \subseteq \{-1, 1\}^{\mathcal{X}}$.

$A = \{a \in \mathbb{R}^n,\ a = \{f(x_1), \ldots f(x_n)\},\ f \in \mathcal{F}\}$

$\Pi_S(\mathcal{F}) = |A|$

$\uparrow$ Growth function. We have $\Pi_S(\mathcal{F}) \le 2^n$.

$$R_n(\mathcal{F}) \le \sqrt{\frac{\lg \Pi_S(\mathcal{F})}{n}}$$ ( $|c| \le \sqrt{n}$ )

(will show) $\simeq \sqrt{\frac{d \log n}{n}}$.

Scanned by CamScanner