Rademacher Complexities

Recall

① ε-representative Sample S (w.r.t $Z, \mathcal{H}, \ell, \mathcal{D}$)

$$\sup_{h \in \mathcal{H}} |L_{\mathcal{D}}(h) - L_S(h)| < \varepsilon$$

② S $\frac{\varepsilon}{2}$-representative

$\Rightarrow$ ERM-rule is consistent

$$L_{\mathcal{D}}(ERM_{\mathcal{H}}(S)) \le \min_{h \in \mathcal{H}} L_{\mathcal{D}}(h) + \varepsilon$$

Notation

$\mathcal{F} := \ell \circ \mathcal{H} := \{ z \longmapsto \ell(h,z) : h \in \mathcal{H} \}$

given $f \in \mathcal{F}$

$L_{\mathcal{D}}(f) = \mathbb{E}_{z \sim \mathcal{D}}[f(z)]$

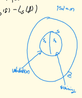
$L_S(f) = \frac{1}{m} \sum_{i=1}^m f(z_i)$

Definition

ⓐ representativeness of S with respect to $\mathcal{F}$

$$Rep_{\mathcal{D}}(\mathcal{F}, S) := \sup_{f \in \mathcal{F}} (L_{\mathcal{D}}(f) - L_S(f))$$

$|S| = m$



$\sup_{f \in \mathcal{F}} (L_{S_1}(f) - L_{S_2}(f))$

$\sigma = (\sigma_1, \dots, \sigma_m) \in \{\pm 1\}^m$

$S_1 : \{ z_i : \sigma_i = 1 \}$

$S_2 : \{ z_i : \sigma_i = -1 \}$

$\mathbb{P}[\sigma_i = 1] = \mathbb{P}[\sigma_i = -1] = \frac{1}{2}$

$\mathcal{F} \circ S = \{ f(z_1), \dots f(z_m) : f \in \mathcal{F} \}$

ⓑ Rademacher Complexity

$$R(\mathcal{F} \circ S) := \frac{1}{m} \mathbb{E}_{\sigma \sim \{\pm 1\}^m} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m \sigma_i f(z_i) \right]$$

Lemma 26.2

$$\mathbb{E}_{S \sim \mathcal{D}^m} [Rep_{\mathcal{D}}(\mathcal{F}, S)] \le 2 \mathbb{E}_{S \sim \mathcal{D}^m} R(\mathcal{F} \circ S)$$

$\sup_f (L_{\mathcal{D}}(f) - L_S(f))$ $\quad$ $\mathbb{E}_\sigma \sup_f (L_{S_1}(f) - L_{S_2}(f))$

Proof -

Let $S' = \{ z_1', \dots z_m' \}$ i.i.d sample

$\forall f \in \mathcal{F} : L_{\mathcal{D}}(f) = \mathbb{E}_{S'}[L_{S'}(f)]$

$$\sup_{f \in \mathcal{F}} L_{\mathcal{D}}(f) - L_S(f) = \sup_{f \in \mathcal{F}} \mathbb{E}_{S'} [L_{S'}(f) - L_S(f)]$$
$$\le \mathbb{E}_{S'} \sup_{f \in \mathcal{F}} [L_{S'}(f) - L_S(f)]$$

$$\mathbb{E}_S \left[\sup_{f \in \mathcal{F}} L_{\mathcal{D}}(f) - L_S(f) \right]$$
$$\le \mathbb{E}_{S,S'} \left(\sup_{f \in \mathcal{F}} (L_{S'}(f) - L_S(f)) \right]$$
$$= \frac{1}{m} \mathbb{E}_{S,S'} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m (f(z_i') - f(z_i)) \right]$$

z_i, z_i' iid

$$= \frac{1}{m} \mathbb{E}_{S,S',\sigma} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m \sigma_i (f(z_i') - f(z_i)) \right]$$
$$\le \frac{1}{m} \mathbb{E}_{S,S',\sigma} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m \sigma_i f(z_i') + \sup_{f \in \mathcal{F}} \sum_{i=1}^m -\sigma_i f(z_i) \right]$$
$$= \mathbb{E}_{S'} [R(\mathcal{F} \circ S')] + \mathbb{E}_S [R(\mathcal{F} \circ S)]$$

$$\left(\mathbb{E}_{S \sim \mathcal{D}^m} [Rep_{\mathcal{D}}(\mathcal{F}, S)] \le 2 \mathbb{E}_{S \sim \mathcal{D}^m} R(\mathcal{F} \circ S) \right.$$

Thm 26.3

(i) $$\mathbb{E}_{S \sim \mathcal{D}^m} [L_{\mathcal{D}}(ERM_{\mathcal{H}}(S)) - L_S(ERM_{\mathcal{H}}(S))]$$
$$\le 2 \mathbb{E}_{S \sim \mathcal{D}^m} R(\ell \circ \mathcal{H} \circ S)$$

(ii) $\forall h^*$ $$\mathbb{E}_{S \sim \mathcal{D}^m} [L_{\mathcal{D}}(ERM_{\mathcal{H}}(S)) - L_{\mathcal{D}}(h^*)]$$
$$\le 2 \mathbb{E}_{S \sim \mathcal{D}^m} R(\ell \circ \mathcal{H} \circ S)$$

(iii) $h^* = \arg\min L_{\mathcal{D}}(h)$, $\delta \in (0,1)$

with prob at least $1-\delta$

$$L_{\mathcal{D}}(ERM_{\mathcal{H}}(S)) - L_{\mathcal{D}}(h^*) \le \frac{2 \mathbb{E}_{S \sim \mathcal{D}^m} R(\ell \circ \mathcal{H} \circ S)}{\delta}$$

Thm 26.5

$\forall z, h \in \mathcal{H}$ $\quad |\ell(h,z)| \le c$

$\Rightarrow$

1. prob at least $1-\delta$, $\forall h \in \mathcal{H}$

$$L_{\mathcal{D}}(h) - L_S(h) \le 2 \mathbb{E}_{S' \sim \mathcal{D}^m} R(\ell \circ \mathcal{H} \circ S') + c \sqrt{\frac{2 \ln(2/\delta)}{m}}$$

$Rep_{\mathcal{D}}(\mathcal{F}, S) = \sup_h L_{\mathcal{D}}(h) - L_S(h)$

(constant $\frac{2c}{m}$)

2. prob at least $1-\delta$, $\forall h \in \mathcal{H}$

$$L_{\mathcal{D}}(h) - L_S(h) \le 2 R(\ell \circ \mathcal{H} \circ S) + 4c \sqrt{\frac{2 \ln(4/\delta)}{m}}$$

3. $\forall h^*$, with prob at least $1-\delta$

$$L_{\mathcal{D}}(ERM_{\mathcal{H}}(S)) - L_{\mathcal{D}}(h^*) \le 2 R(\ell \circ \mathcal{H} \circ S) + 5c \sqrt{\frac{2 \ln(8/\delta)}{m}}$$

($h_S := ERM_{\mathcal{H}}(S)$)

$L_{\mathcal{D}}(h_S) - L_{\mathcal{D}}(h^*)$

$= \underbrace{L_{\mathcal{D}}(h_S) - L_S(h_S)}_{(ii)} + \underbrace{L_S(h_S) - L_S(h^*)}_{\le 0} + \underbrace{L_S(h^*) - L_{\mathcal{D}}(h^*)}_{c\sqrt{\dots}}$

Rademacher Calculus

$A \subset \mathbb{R}^m$

$$R(A) := \frac{1}{m} \mathbb{E}_\sigma \left[\sup_{a \in A} \sum_{i=1}^m \sigma_i a_i \right]$$

① $A \subset \mathbb{R}^m$, $c \in \mathbb{R}$, $a_0 \in \mathbb{R}^m$

$R(\{ca + a_0 : a \in A\}) \le |c| R(A)$

② $A' = \{ \sum_{i=1}^N \alpha_i a^{(i)} : N \in \mathbb{N}, a^{(i)} \in A, \alpha_i \ge 0, \|\alpha\|_1 = 1 \}$

$R(A') = R(A)$

③ $A = \{a_1, \dots, a_N\}$ finite, $\subset \mathbb{R}^m$

$\bar{a} = \frac{1}{N} \sum_{i=1}^N a_i$

$$R(A) \le \max_{a \in A} \|a - \bar{a}\| \frac{\sqrt{2 \log(N)}}{m}$$

④ $\forall i \in [m] : \phi_i : \mathbb{R} \to \mathbb{R}$ ρ-Lipschitz-Function

$(\forall \alpha, \beta \in \mathbb{R} : |\phi_i(\alpha) - \phi_i(\beta)| \le \rho |\alpha - \beta|)$

for $a \in \mathbb{R}^m : \phi(a) = (\phi_1(a_1), \dots \phi_m(a_m))$

$\phi \circ A = \{ \phi(a) : a \in A \}$

$$R(\phi \circ A) \le \rho R(A)$$

Rademacher Complexity of Linear Classes

$\mathcal{H}_1 = \{ x \longmapsto \langle w, x \rangle : \|w\|_1 \le 1 \}$

$\mathcal{H}_2 = \{ x \longmapsto \langle w, x \rangle : \|w\|_2 \le 1 \}$

Lemma ($\mathcal{H}_2$)

$S = (x_1, \dots x_m)$ vectors in Hilbert Space

$\mathcal{H}_2 \circ S = \{ (\langle w, x_1 \rangle, \dots, \langle w, x_m \rangle) : \|w\|_2 \le 1 \}$

$$R(\mathcal{H}_2 \circ S) \le \frac{\max_i \|x_i\|_2}{\sqrt{m}}$$

Lemma ($\mathcal{H}_1$)

$S = (x_1, \dots x_m) \subset \mathbb{R}^n$

$$R(\mathcal{H}_1 \circ S) \le \max \|x_i\|_\infty \sqrt{\frac{2 \log(2n)}{m}}$$

$\mathcal{H}_1 \rightarrow \langle w, x \rangle \le \|w\|_1 \cdot \|x\|_2$

Apply to SVM $\quad \mathcal{H}_2$

$\mathcal{H} = \{ w : \|w\|_2 \le B \}$

$Z = X \times Y$

loss function

$\ell(w,(x,y)) = \phi(\langle w, x \rangle, y)$

ϕ ρ-Lipschitz

Thm 26.12

$\|x\|_2 \le R$

prob $1-\delta$

$$\forall w \in \mathcal{H}. \ L_{\mathcal{D}}(w) \le L_S(w) + \frac{2\rho BR}{\sqrt{m}} + \sqrt{\frac{2 \ln(2/\delta)}{m}}$$

Generalization Bound for predictors with low ℓ_1-norm

$\mathcal{H} = \{ w : \|w\|_1 \le B \}$

Thm 26.15 ($\mathcal{H}_1$)

$\|x\|_\infty \le R$

$$\forall w \in \mathcal{H} : L_{\mathcal{D}}(w) \le L_S(w) + 2\rho BR \sqrt{\frac{2 \log(2d)}{m}} + c \sqrt{\frac{2 \ln(2/\delta)}{m}}$$

Hoeffding's Inequality

Let $\Theta_1, \dots, \Theta_m$ be a sequence of i.i.d. random variables and assume that for all i,

$\mathbb{E}[\Theta_i] = \mu$ and $\mathbb{P}[a \le \Theta_i \le b] = 1$. $\quad$ $b = c, a = 0$

Then, for any $\varepsilon < 0$ $\quad$ $\delta = e^{-2m\varepsilon^2}$

$$\mathbb{P}\left[\left| \frac{1}{m} \sum_{i=1}^m \Theta_i - \mu \right| > \varepsilon \right] \le 2 \exp(-2m\varepsilon^2 / (b-a)^2)$$

$c \sqrt{\frac{2 \ln \frac{2}{\delta}}{m}} = \varepsilon$