

Rademacher Complexities

Peng Lingwei

August 2, 2019

Contents

| | |
|--|----------|
| 26 Rademacher Complexities | 2 |
| 26.1 THE RADEMACHER COMPLEXITY | 2 |
| 26.1.1 Rademacher Calculus | 4 |
| 26.2 RADEMACHER COMPLEXITY OF LINEAR CLASSES | 6 |
| 26.3 GENERALIZATION BOUNDS FOR SVM | 7 |

26 Rademacher Complexities

1. **Uniform convergence** is a sufficient condition for learnability.
2. **Rademacher complexities** measures the rate of uniform convergence.

26.1 THE RADEMACHER COMPLEXITY

Definition 1. (ϵ -Representative Sample). (w.r.t. domain $Z = (\mathcal{X}, \mathcal{Y}) \sim \mathcal{D}$, hypothesis class \mathcal{H} , loss function l). A training set S is called ϵ -representative if

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

We have $m_{\mathcal{H}}(\epsilon, \delta) \leq m_{\mathcal{H}}^{UC}(\epsilon/2, \delta)$.

Definition 2. (The representativeness of S with respect to \mathcal{F}).

$$Rep_{\mathcal{D}}(\mathcal{F}, S) := \sup_{f \in \mathcal{F}} (L_{\mathcal{D}}(f) - L_S(f)) \quad (1)$$

where,

$$\mathcal{F} := l \circ \mathcal{H} := \{z \mapsto l(h, z) : z \in Z, h \in \mathcal{H}\}$$

$$f \in \mathcal{F}, \quad L_{\mathcal{D}}(f) = \mathbb{E}_{z \sim \mathcal{D}}[f(z)], \quad L_S = \frac{1}{m} \sum_{i=1}^m f(z_i).$$

Analogizing the concept of validation set which used to estimate the representativeness of S , we define **Rademacher complexity**.

Definition 3. (The rademacher complexity of \mathcal{F} w.r.t. S).

$$R(\mathcal{F} \circ S) := \frac{1}{m} \mathbb{E}_{\sigma \sim \{\pm 1\}^m} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m \sigma_i f(z_i) \right] \quad (2)$$

where,

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

$$\sigma = \{\sigma_i : \mathbb{P}[\sigma_i = 1] = \mathbb{P}[\sigma_i = -1] = 0.5\}$$

More generally, given a set of vectors, $A \subset \mathbb{R}^m$, we define

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

Lemma 1.

$$\mathbb{E}_{S \sim \mathcal{D}^m} [Rep_{\mathcal{D}}(\mathcal{F}, S)] \leq 2 \mathbb{E}_{S \sim \mathcal{D}^m} R(\mathcal{F} \circ S) \quad (3)$$

Proof. Let $S' = \{z'_1, \dots, z'_m\}$ be another i.i.d. sample. Then,

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

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

$$\leq \mathbb{E}_{S'} \left[\sup_{f \in \mathcal{F}} (L_{S'}(f) - L_S(f)) \right]$$

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

In some techniques, we can get:

$$\begin{aligned} \mathbb{E}_{S, S'} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m (f(z'_i) - f(z_i)) \right] &= \mathbb{E}_{S, S', \sigma} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m \sigma_i (f(z'_i) - f(z_i)) \right] \\ &\leq \mathbb{E}_{S, S', \sigma} \left[\sup_{f \in \mathcal{F}} \sum_{i=1}^m \sigma_i (f(z'_i)) + \sup_{f \in \mathcal{F}} \sum_{i=1}^m (-\sigma_i) f(z_i) \right] \\ &= m \mathbb{E}_{S'} [R(\mathcal{F} \circ S')] + m \mathbb{E}_S [R(\mathcal{F} \circ S)] = 2m \mathbb{E}_S [R(\mathcal{F} \circ S)]. \end{aligned}$$

□

Theorem 1.

$$\mathbb{E}_{S \sim \mathcal{D}^m} [L_{\mathcal{D}}(\text{ERM}_{\mathcal{H}}(S)) - L_S(\text{ERM}_{\mathcal{H}}(S))] \leq 2 \mathbb{E}_{S \sim \mathcal{D}^m} R(l \circ \mathcal{H} \circ S)$$

$$\mathbb{E}_{S \sim \mathcal{D}^m} [L_{\mathcal{D}}(\text{ERM}_{\mathcal{H}}(S)) - L_S(h^*)] \leq 2 \mathbb{E}_{S \sim \mathcal{D}^m} R(l \circ \mathcal{H} \circ S), \text{ where } h^* = \arg \min_h L_{\mathcal{D}}(h)$$

Because $L_{\mathcal{D}}(\text{ERM}_{\mathcal{H}}(S)) - L_{\mathcal{D}}(h^*) \geq 0$, then

$$\mathbb{P} \left\{ L_{\mathcal{D}}(\text{ERM}_{\mathcal{H}}(S)) - L_{\mathcal{D}}(h^*) \geq \frac{2 \mathbb{E}_{S' \sim \mathcal{D}^m} R(l \circ \mathcal{H} \circ S')}{\delta} \right\} \leq \delta$$

Lemma 2. (McDiarmid's Inequality).

If

$$|f(x_1, \dots, x_m) - f(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_m)| \leq c_i.$$

then,

$$\mathbb{P} \{f - \mathbb{E}f \geq \epsilon\} \leq \exp \left(\frac{-2\epsilon^2}{\sum_{i=1}^m c_i} \right)$$

$$\mathbb{P} \{f - \mathbb{E}f \leq -\epsilon\} \leq \exp \left(\frac{-2\epsilon^2}{\sum_{i=1}^m c_i} \right)$$

which also means

$$\mathbb{P} \left\{ |f - \mathbb{E}f| \geq c \sqrt{\ln(2/\delta)m/2} \right\} \leq \delta$$

Theorem 2. (Data-dependent bound). Assume that for all z and $h \in \mathcal{H}$, we have that $|l(h, z)| \leq c$. Then,

1.

$$\mathbb{P} \left\{ \forall h \in \mathcal{H}, L_{\mathcal{D}}(h) - L_S(h) \leq 2 \mathbb{E}_{S' \sim \mathcal{D}^m} R(l \circ \mathcal{H} \circ S') + c \sqrt{2 \ln(1/\delta)/m} \right\} \geq 1 - \delta$$

Proof. $\text{Rep}_{\mathcal{D}}(\mathcal{F}, S)$ satisfies the preceding condition with a constant $2c/m$,

□

2.

$$\mathbb{P} \left\{ \forall h \in \mathcal{H}, L_{\mathcal{D}}(h) - L_S(h) \leq 2R(l \circ \mathcal{H} \circ S) + 3c \sqrt{2 \ln(2/\delta)/m} \right\} \geq 1 - \delta$$

Proof.

$$\mathbb{P} \left\{ \text{Rep}_{\mathcal{D}}(F, S) \leq \mathbb{E}_{S'} \text{Rep}_{\mathcal{D}}(l \circ \mathcal{H} \circ S') + c\sqrt{2\ln(2/\delta)/m} \right\} \geq 1 - \delta/2$$

$$\mathbb{P} \left\{ \mathbb{E} \text{Rep}_{\mathcal{D}}(F, S) \leq 2\mathbb{E} R(l \circ \mathcal{H} \circ S') \right\} = 1$$

$$\mathbb{P} \left\{ \mathbb{E}_{S'} R(l \circ \mathcal{H} \circ S') \leq R(l \circ \mathcal{H} \circ S) + c\sqrt{2\ln(2/\delta)/m} \right\} \geq 1 - \delta/2$$

□

3.

$$\mathbb{P} \left\{ L_{\mathcal{D}}(\text{ERM}_{\mathcal{H}}(S)) - L_S(h^*) \leq 2R(l \circ \mathcal{H} \circ S) + 4c\sqrt{2\ln(3/\delta)/m} \right\} \geq 1 - \delta$$

Proof.

$$\begin{aligned} L_{\mathcal{D}}(h_S) - L_{\mathcal{D}}(h^*) &= L_{\mathcal{D}}(h_S) - L_S(h_S) + L_S(h_S) - L_S(h^*) + L_S(h^*) - L_{\mathcal{D}}(h^*) \\ &\leq (L_{\mathcal{D}}(h_S) - L_S(h_S)) + (L_S(h^*) - L_{\mathcal{D}}(h^*)) \end{aligned}$$

$$\mathbb{P} \left\{ L_{\mathcal{D}}(h_S) - L_S(h_S) \leq 2R(l \circ \mathcal{H} \circ S) + 3c\sqrt{2\ln(3/\delta)/m} \right\} \geq 1 - 2\delta/3$$

Because $L_{\mathcal{D}}(h^*)$ does not depend on S , so we can use hoeffding's inequality to get

$$\mathbb{P} \left\{ L_S(h^*) - L_{\mathcal{D}}(h^*) \leq c\sqrt{2\ln(3/\delta)/m} \right\} \geq 1 - \delta/3$$

□

26.1.1 Rademacher Calculus

Lemma 3. $\forall A \subset \mathbb{R}^m, c \in \mathbb{R}, \mathbf{a}_0 \in \mathbb{R}^m$, we have

$$R(\{c\mathbf{a} + \mathbf{a}_0 : \mathbf{a} \in A\}) = |c|R(A) \quad (4)$$

Lemma 4. $\forall A \subset \mathbb{R}^m$, if $A' = \left\{ \sum_{j=1}^N \alpha_j \mathbf{a}^{(j)} : N \in \mathbb{N}, \forall j, \mathbf{a}^{(j)} \in A, \alpha_j \geq 0, \|\vec{\alpha}\|_1 = 1 \right\}$, then $R(A') = R(A)$.

Proof.

$$\begin{aligned} mR(A') &= \mathbb{E}_{\sigma} \sup_{\vec{\alpha} \succeq \vec{0}: \|\vec{\alpha}\|_1=1} \sup_{\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(N)}} \sum_{i=1}^m \sigma_i \sum_{j=1}^N \alpha_j a_i^{(j)} \\ &= \mathbb{E}_{\sigma} \sup_{\vec{\alpha} \succeq \vec{0}: \|\vec{\alpha}\|_1=1} \sum_{j=1}^N \alpha_j \sup_{\mathbf{a}^{(j)}} \sum_{i=1}^m \sigma_i a_i^{(j)} \\ &= \mathbb{E}_{\sigma} \sup_{\mathbf{a} \in A} \sum_{i=1}^m \sigma_i a_i \end{aligned}$$

□

Lemma 5. (Massart Lemma). Let $A = \{\mathbf{a}_1, \dots, \mathbf{a}_N\}$ be a finite set of vectors in \mathbf{R}^m . Define $\bar{\mathbf{a}} = \frac{1}{N} \sum_{i=1}^N \mathbf{a}_i$. Then,

$$R(A) \leq \max_{\mathbf{a} \in A} \|\mathbf{a} - \bar{\mathbf{a}}\|_2 \frac{\sqrt{2 \log(N)}}{m} \quad (5)$$

Proof.

$$\begin{aligned} \forall A, \quad mR(A) &= \mathbb{E}_{\vec{\sigma}} \left[\max_{\mathbf{a} \in A} \langle \vec{\sigma}, \mathbf{a} \rangle \right] = \mathbb{E}_{\vec{\sigma}} \left[\log \left(\max_{\mathbf{a} \in A} e^{\langle \vec{\sigma}, \mathbf{a} \rangle} \right) \right] \\ &= \mathbb{E}_{\vec{\sigma}} \left[\log \left(\sum_{\mathbf{a} \in A} e^{\langle \vec{\sigma}, \mathbf{a} \rangle} \right) \right] \leq \log \left[\mathbb{E}_{\vec{\sigma}} \left(\sum_{\mathbf{a} \in A} e^{\langle \vec{\sigma}, \mathbf{a} \rangle} \right) \right] \\ &\leq \log \left(\sum_{\mathbf{a} \in A} \prod_{i=1}^m \mathbb{E}_{\sigma_i} [e^{\sigma_i a_i}] \right) \leq \log \left(\sum_{\mathbf{a} \in A} \prod_{i=1}^m [e^{a_i} + e^{-a_i}] / 2 \right) \\ &\leq \log \left(\sum_{\mathbf{a} \in A} \prod_{i=1}^m e^{a_i^2 / 2} \right) = \log \left(\sum_{\mathbf{a} \in A} \exp(\|\mathbf{a}\|_2^2 / 2) \right) \\ &\leq \log \left(|A| \max_{\mathbf{a} \in A} \exp(\|\mathbf{a}\|_2^2 / 2) \right) = \log(|A|) + \max_{\mathbf{a} \in A} (\|\mathbf{a}\|_2^2 / 2) \end{aligned}$$

Since $R(A) = R(A')/\lambda$ we obtain that

$$R(A) \leq \frac{\log(|A|) + \lambda^2 \max_{\mathbf{a} \in A} (\|\mathbf{a}\|_2^2 / 2)}{\lambda m}$$

□

Lemma 6. (Contraction Lemma). $\forall i \in [m]$, let $\phi_i : \mathbb{R} \rightarrow \mathbb{R}$ be a ρ -Lipschitz function. For $\mathbf{a} \in \mathbb{R}^m$ let $\phi(\mathbf{a}) = (\phi_1(a_1), \dots, \phi_m(a_m))$. Let $\phi \circ A = \{\phi(\vec{a}) : \mathbf{a} \in A\}$. Then,

$$R(\phi \circ A) \leq \rho R(A).$$

Proof. First, $\rho = 1$. Let $A_i = \{(a_1, \dots, a_{i-1}, \phi_i(a_i), a_{i+1}, \dots, a_m) : \mathbf{a} \in A\}$.

$$\begin{aligned} mR(A_1) &= \mathbb{E}_{\sigma} \left[\sup_{\mathbf{a} \in A_1} \sum_{i=1}^m \sigma_i a_i \right] \\ &= \mathbb{E}_{\sigma} \left[\sup_{\mathbf{a} \in A} \sigma_1 \phi(a_1) + \sum_{i=2}^m \sigma_i a_i \right] \\ &= \frac{1}{2} \mathbb{E}_{\sigma_2, \dots, \sigma_m} \left[\sup_{\mathbf{a}, \mathbf{a}' \in A} \left(\phi(a_1) - \phi(a'_1) + \sum_{i=2}^m \sigma_i a_i + \sum_{i=2}^m \sigma_i a'_i \right) \right] \\ &\leq \frac{1}{2} \mathbb{E}_{\sigma_2, \dots, \sigma_m} \left[\sup_{\mathbf{a}, \mathbf{a}' \in A} \left(|a_1 - a'_1| + \sum_{i=2}^m \sigma_i a_i + \sum_{i=2}^m \sigma_i a'_i \right) \right] \\ &= \frac{1}{2} \mathbb{E}_{\sigma_2, \dots, \sigma_m} \left[\sup_{\mathbf{a}, \mathbf{a}' \in A} \left(a_1 - a'_1 + \sum_{i=2}^m \sigma_i a_i + \sum_{i=2}^m \sigma_i a'_i \right) \right] \\ mR(A_1) &\leq mR(A) \end{aligned}$$

□

26.2 RADEMACHER COMPLEXITY OF LINEAR CLASSES

1. $\mathcal{H}_1 = \{\mathbf{x} \mapsto \langle \mathbf{w}, \mathbf{x} \rangle : \|\mathbf{w}\|_1 \leq 1\}$
2. $\mathcal{H}_2 = \{\mathbf{x} \mapsto \langle \mathbf{w}, \mathbf{x} \rangle : \|\mathbf{w}\|_2 \leq 1\}$

Lemma 7.

$$R(\mathcal{H}_2 \circ S) \leq \frac{\max_i \|\mathbf{x}_i\|_2}{\sqrt{m}} \quad (6)$$

Proof.

$$\begin{aligned} mR(\mathcal{H}_2 \circ S) &= \mathbb{E}_\sigma \left[\sup_{\mathbf{a} \in \mathcal{H}_2 \circ S} \sum_{i=1}^m \sigma_i a_i \right] = \mathbb{E}_\sigma \left[\sup_{\mathbf{w}: \|\mathbf{w}\|_2 \leq 1} \sum_{i=1}^m \sigma_i \langle \mathbf{w}, \mathbf{x}_i \rangle \right] \\ &= \mathbb{E}_\sigma \left[\sup_{\mathbf{w}: \|\mathbf{w}\|_2 \leq 1} \langle \mathbf{w}, \sum_{i=1}^m \sigma_i \mathbf{x}_i \rangle \right] \leq \mathbb{E}_\sigma \left[\left\| \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\|_2 \right] \\ &\leq \left(\mathbb{E}_\sigma \left[\left\| \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\|_2^2 \right] \right)^{1/2} \\ \mathbb{E}_\sigma \left[\left\| \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\|_2^2 \right] &= \sum_{i \neq j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \mathbb{E}_\sigma [\sigma_i \sigma_j] + \sum_{i=1}^m \langle \mathbf{x}_i, \mathbf{x}_i \rangle \mathbb{E}_\sigma [\sigma_i^2] \\ &= \sum_{i=1}^m \|\mathbf{x}_i\|_2^2 \leq m \max_i \|\mathbf{x}_i\|_2^2 \end{aligned}$$

□

Lemma 8. Let $S = (\mathbf{x}_1, \dots, \mathbf{x}_m)$ be the vectors in \mathbb{R}^n , then,

$$R(\mathcal{H}_1 \circ S) \leq \max_i \|\mathbf{x}_i\|_\infty \sqrt{\frac{2 \log(2n)}{m}} \quad (7)$$

Proof. Using Holder's inequality, we have $\langle \mathbf{w}, \mathbf{v} \rangle \leq \|\mathbf{w}\|_1 \|\mathbf{v}\|_\infty$. Therefore,

$$mR(\mathcal{H}_1 \circ S) = \mathbb{E}_\sigma \left[\sup_{\mathbf{w}: \|\mathbf{w}\|_1 \leq 1} \langle \mathbf{w}, \sum_{i=1}^m \sigma_i \mathbf{x}_i \rangle \right] \leq \mathbb{E}_\sigma \left[\left\| \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\|_\infty \right].$$

Let $j \in [n]$ and $\mathbf{v}_j = (x_{1,j}, \dots, x_{m,j}) \in \mathbb{R}^m$, and $V = \{\mathbf{v}_1, \dots, \mathbf{v}_n, -\mathbf{v}_1, \dots, -\mathbf{v}_n\}$. Note that $\|\mathbf{v}_j\|_2 \leq \sqrt{m} \max_i \|\mathbf{x}_i\|_\infty$.

$$\begin{aligned} \mathbb{E}_\sigma \left[\left\| \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\|_\infty \right] &= \mathbb{E}_\sigma \left[\max_j |\langle \mathbf{v}_j, \sigma \rangle| \right] = mR(V) \\ &\leq \max_j \|\mathbf{v}_j\|_2 \frac{\sqrt{2 \log(2n)}}{m} \\ &\leq \max_i \|\mathbf{x}_i\|_\infty \sqrt{2 \log(2n)/m} \end{aligned}$$

□

In chapter 6, we defined “Restriction of \mathcal{H} to S ”:

$$\mathcal{H}_S = \{(h(s_1), h(s_2), \dots, h(s_m)) : h \in \mathcal{H}, s_i \in S\},$$

Here is the growth function:

$$\forall m \in \mathbb{N}, \Pi_{\mathcal{H}}(m) = \max_{S \sim \mathcal{D}^m} |\mathcal{H}_S|$$

Then, the VC-dimension is:

$$VCdim(\mathcal{H}) = \max \{m : \Pi_{\mathcal{H}}(m) = 2^m\}$$

Using Massart Lemma:

$$R_m(\mathcal{H} \circ S) = \mathbb{E}_S \left[\mathbb{E}_{\vec{\sigma}} \left[\sup_{h_S \in \mathcal{H}_S} \frac{1}{m} \langle \vec{\sigma}, h_S \rangle \right] \right] \leq \mathbb{E}_S \left[\frac{\sqrt{m} \sqrt{2 \log(\Pi_{\mathcal{H}})}}{m} \right]$$

26.3 GENERALIZATION BOUNDS FOR SVM