Classification: Rademacher Complexity

Machine Learning: Jordan Boyd-Graber
University of Colorado Boulder
LECTURE 6B

Slides adapted from Rob Schapire

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Administrivia Questions

Administrivia Questions

Administrivia Questions

Administrivia Questions

Administrivia Questions

Single Hypothesis

What is the Rademacher complexity of a hypothesis set reduced to a single hypothesis?

(1)

Single Hypothesis

What is the Rademacher complexity of a hypothesis set reduced to a single hypothesis?

$$\mathscr{R}_m(H) = \mathbb{E}_{S\sim D^m,\sigma}\left[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(z_i)\right] \quad (1)$$

$$(2)$$

Single Hypothesis

What is the Rademacher complexity of a hypothesis set reduced to a single hypothesis?

$$\mathscr{R}_m(H) = \mathbb{E}_{S \sim D^m, \sigma}\left[\sup_{h \in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(z_i)\right] \tag{1}$$

$$= \mathbb{E}_{S \sim D^m, \sigma}\left[\frac{1}{m} \sum_{i=1}^{m} \sigma_i h_0(z_i)\right] \tag{2}$$

$$\tag{3}$$

Single Hypothesis

What is the Rademacher complexity of a hypothesis set reduced to a single hypothesis?

$$\mathscr{R}_m(H) = \mathbb{E}_{S\sim D^m,\sigma}\left[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(z_i)\right] \tag{1}$$

$$= \mathbb{E}_{S\sim D^m,\sigma}\left[\frac{1}{m}\sum_{i=1}^{m}\sigma_i h_0(z_i)\right] \tag{2}$$

$$= \mathbb{E}_{S\sim D^m}\left[\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_\sigma\left[\sigma_i\right]\sigma_i h_0(z_i)\right] \tag{3}$$

$$\tag{4}$$

Single Hypothesis

What is the Rademacher complexity of a hypothesis set reduced to a single hypothesis?

$$\mathscr{R}_m(H) = \mathbb{E}_{S \sim D^m, \sigma}\left[\sup_{h \in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(z_i)\right] \quad (1)$$

$$= \mathbb{E}_{S \sim D^m, \sigma}\left[\frac{1}{m} \sum_{i=1}^{m} \sigma_i h_0(z_i)\right] \quad (2)$$

$$= \mathbb{E}_{S \sim D^m}\left[\frac{1}{m} \sum_{i=1}^{m} \mathbb{E}_{\sigma}[\sigma_i] \sigma_i h_0(z_i)\right] \quad (3)$$

$$= \mathbb{E}_{S \sim D^m}\left[\frac{1}{m} \sum_{i=1}^{m} 0 \cdot \sigma_i h_0(z_i)\right] \quad (4)$$

Single Hypothesis

What is the Rademacher complexity of a hypothesis set reduced to a single hypothesis?

$$\mathscr{R}_m(H) = \mathbb{E}_{S\sim D^m,\sigma}\left[\sup_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i h(z_i)\right] \quad (1)$$

$$= \mathbb{E}_{S\sim D^m,\sigma}\left[\frac{1}{m}\sum_{i=1}^{m}\sigma_i h_0(z_i)\right] \quad (2)$$

$$= \mathbb{E}_{S\sim D^m}\left[\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}_\sigma\left[\sigma_i\right]\sigma_i h_0(z_i)\right] \quad (3)$$

$$= \mathbb{E}_{S\sim D^m}\left[\frac{1}{m}\sum_{i=1}^{m}0\cdot\sigma_i h_0(z_i)\right] = 0 \quad (4)$$

$$(5)$$

Rademacher Identity 1

Prove

$$\mathscr{R}_m(\alpha H) = |\alpha|\mathscr{R}_m(H)$$

If $\alpha \geq 0$

If $\alpha < 0$

Rademacher Identity 1

Prove

$$\mathscr{R}_m(\alpha H) = |\alpha| \mathscr{R}_m(H)$$

If $\alpha \geq 0$

$$\sup_{h \in \alpha H} \sum_{i=1}^{m} \sigma_i h(x_i) = \quad (6)$$

$$\sup_{h \in H} \sum_{i=1}^{m} \alpha \sigma_i h(x_i) = \quad (7)$$

$$\alpha \sup_{h \in H} \sum_{i=1}^{m} \sigma_i h(x_i) \quad (8)$$

If $\alpha < 0$

Rademacher Identity 1

Prove

$$\mathscr{R}_m(\alpha H) = |\alpha| \mathscr{R}_m(H)$$

If $\alpha \geq 0$

$$\sup_{h \in \alpha H} \sum_{i=1}^{m} \sigma_i h(x_i) = \quad (6)$$

$$\sup_{h \in H} \sum_{i=1}^{m} \alpha \sigma_i h(x_i) = \quad (7)$$

$$\alpha \sup_{h \in H} \sum_{i=1}^{m} \sigma_i h(x_i) \quad (8)$$

If $\alpha < 0$

$$\sup_{h \in \alpha H} \sum_{i=1}^{m} \sigma_i h(x_i) = \quad (9)$$

$$\sup_{h \in H} \sum_{i=1}^{m} \alpha \sigma_i h(x_i) = \quad (10)$$

$$(-\alpha) \sup_{h \in H} \sum_{i=1}^{m} (-\sigma_i) h(x_i) \quad (11)$$

Rademacher Identity 1

Prove

$$\mathscr{R}_m(\alpha H) = |\alpha|\mathscr{R}_m(H)$$

If $\alpha \geq 0$

$$\sup_{h \in \alpha H} \sum_{i=1}^{m} \sigma_i h(x_i) = \tag{6}$$

$$\sup_{h \in H} \sum_{i=1}^{m} \alpha \sigma_i h(x_i) = \tag{7}$$

$$\alpha \sup_{h \in H} \sum_{i=1}^{m} \sigma_i h(x_i) \tag{8}$$

If $\alpha < 0$

$$\sup_{h \in \alpha H} \sum_{i=1}^{m} \sigma_i h(x_i) = \tag{9}$$

$$\sup_{h \in H} \sum_{i=1}^{m} \alpha \sigma_i h(x_i) = \tag{10}$$

$$(-\alpha) \sup_{h \in H} \sum_{i=1}^{m} (-\sigma_i) h(x_i) \tag{11}$$

Since σ_i and $-\sigma$ have the same distribution, $\mathscr{R}_m(\alpha H) = |\alpha|\mathscr{R}_m(H)$

Rademacher Identity 2

Prove

$$\mathscr{R}_m(H+H') = \mathscr{R}_m(H) + \mathscr{R}_m(H') \tag{12}$$

Rademacher Identity 2

Prove

$$\mathscr{R}_m(H+H') = \mathscr{R}_m(H) + \mathscr{R}_m(H')$$

$$\mathscr{R}_m(H+H') \tag{12}$$

$$= \frac{1}{m}\mathbb{E}_{\vec{\sigma},S}\left[\sup_{h \in H, h' \in H'} \sum_{i=1}^{m} \sigma_i \left(h(x_i) + h'(x_i)\right)\right] \tag{13}$$

$$\tag{14}$$

Rademacher Identity 2

Prove

$$\mathscr{R}_m(H+H') = \mathscr{R}_m(H) + \mathscr{R}_m(H')$$

$$\mathscr{R}_m(H+H') \tag{12}$$

$$= \frac{1}{m}\mathbb{E}_{\vec{\sigma},S}\left[\sup_{h\in H, h'\in H'} \sum_{i=1}^{m} \sigma_i (h(x_i) + h'(x_i))\right] \tag{13}$$

$$= \frac{1}{m}\mathbb{E}_{\vec{\sigma},S}\left[\sup_{h\in H, h'\in H'} \sum_{i=1}^{m} \sigma_i h(x_i) + \sup_{h\in H, h'\in H'} \sum_{i=1}^{m} \sigma_i h'(x_i))\right] \tag{14}$$

(15)

Rademacher Identity 2

Prove

$$\mathscr{R}_m(H+H') = \mathscr{R}_m(H) + \mathscr{R}_m(H')$$

$$\mathscr{R}_m(H+H') \tag{12}$$

$$= \frac{1}{m}\mathbb{E}_{\vec{\sigma},S}\left[\sup_{h\in H, h'\in H'} \sum_{i=1}^{m} \sigma_i (h(x_i) + h'(x_i))\right] \tag{13}$$

$$= \frac{1}{m}\mathbb{E}_{\vec{\sigma},S}\left[\sup_{h\in H, h'\in H'} \sum_{i=1}^{m} \sigma_i h(x_i) + \sup_{h\in H, h'\in H'} \sum_{i=1}^{m} \sigma_i h'(x_i))\right] \tag{14}$$

$$= \frac{1}{m}\mathbb{E}_{\vec{\sigma},S}\left[\sup_{h\in H} \sum_{i=1}^{m} \sigma_i h(x_i)\right] + \frac{1}{m}\mathbb{E}_{\vec{\sigma},S}\left[\sup_{h'\in H'} \sum_{i=1}^{m} \sigma_i h(x_i)\right] \tag{15}$$

VC Dimension

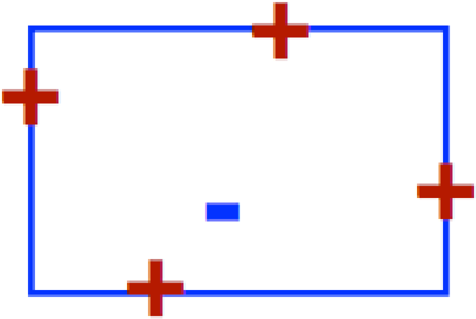
To show VC dimension of a set of points

- Show that **a** set of d *can* be shattered
- Show that **no** set of $d+1$ can be shattered

Axis Aligned Rectangles

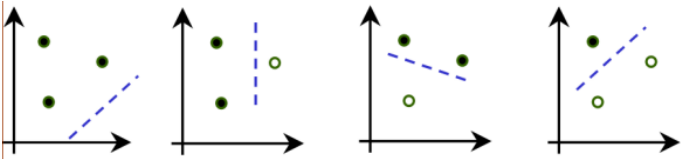
Axis Aligned Rectangles

Axis Aligned Rectangles



Hyperplanes

Hyperplanes



Hyperplanes

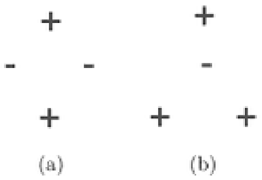


Figure 3.2 Unrealizable dichotomies for four points using hyperplanes in $\mathbb{R}^2$. (a) All four points lie on the convex hull. (b) Three points lie on the convex hull while the remaining point is interior.

Hyperplanes

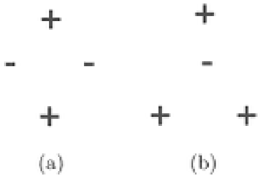


Figure 3.2 Unrealizable dichotomies for four points using hyperplanes in $\mathbb{R}^2$. (a) All four points lie on the convex hull. (b) Three points lie on the convex hull while the remaining point is interior.

In general, the VC dimension of d-dimensional hyperplanes is $d+1$

Finite Subsets

Show that the VC dimension of a finite hypothesis set H is at most $\lg|H|$.

Finite Subsets

Show that the VC dimension of a finite hypothesis set H is at most $\lg|H|$.

- To shatter a set, it means that every point can take on a different binary label $h(x)$

Finite Subsets

Show that the VC dimension of a finite hypothesis set H is at most $\lg|H|$.

- To shatter a set, it means that every point can take on a different binary label $h(x)$
- If a set has d points, there are 2^d ways to do that

Finite Subsets

Show that the VC dimension of a finite hypothesis set H is at most $\lg|H|$.

- To shatter a set, it means that every point can take on a different binary label $h(x)$
- If a set has d points, there are 2^d ways to do that
- Each configuration requires a different hypothesis

Finite Subsets

Show that the VC dimension of a finite hypothesis set H is at most $\lg|H|$.

- To shatter a set, it means that every point can take on a different binary label $h(x)$
- If a set has d points, there are 2^d ways to do that
- Each configuration requires a different hypothesis
- Solving for the number of hypotheses gives $\lg|H|$

Next time

- Getting more practical
- SVMs
- Excellent theoretical properties