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

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

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

(1)
Single Hypothesis

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

\[
R_m(H) = \mathbb{E}_{S \sim D^m, \sigma} \left[ \sup_{h \in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(z_i) \right] 
\]  \hspace{1cm} (1)

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

\[
\]  \hspace{1cm} (3)
Single Hypothesis

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

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

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

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

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

\[
\mathcal{R}_m(H) = \mathbb{E}_{S \sim D^m, \sigma} \left[ \sup_{h \in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i h(z_i) \right] \\
= \mathbb{E}_{S \sim D^m, \sigma} \left[ \frac{1}{m} \sum_{i=1}^{m} \sigma_i h_0(z_i) \right] \\
= \mathbb{E}_{S \sim D^m} \left[ \frac{1}{m} \sum_{i=1}^{m} \mathbb{E}_{\sigma} [\sigma_i] \sigma_i h_0(z_i) \right] \\
= \mathbb{E}_{S \sim D^m} \left[ \frac{1}{m} \sum_{i=1}^{m} 0 \cdot \sigma_i h_0(z_i) \right]
\]
Single Hypothesis

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

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

(1)

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

(2)

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

(3)

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

(4)

(5)
Rademacher Identity 1

Prove

\[ \mathcal{R}_m(\alpha H) = |\alpha| \mathcal{R}_m(H) \]

If \( \alpha \geq 0 \) \hspace{1cm} \text{If } \alpha < 0
Rademacher Identity 1

Prove

\[ R_m(\alpha H) = |\alpha|R_m(H) \]

If \( \alpha \geq 0 \)

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

\[ \sup_{h \in H} \sum_{i=1}^{m} \alpha \sigma_i h(x_i) = \quad (7) \quad \text{If } \alpha < 0 \]

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

Prove

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

If $\alpha \geq 0$

$$\sup_{h \in \alpha H} \sum_{i=1}^{m} \sigma_i h(x_i) = \sum_{i=1}^{m} \alpha \sigma_i h(x_i) = \alpha \sup_{h \in H} \sum_{i=1}^{m} \sigma_i h(x_i)$$ (6) (7) (8)

If $\alpha < 0$

$$\sup_{h \in \alpha H} \sum_{i=1}^{m} \sigma_i h(x_i) = \sum_{i=1}^{m} \alpha \sigma_i h(x_i) = \alpha \sup_{h \in H} \sum_{i=1}^{m} \sigma_i h(x_i)$$ (9) (10) (11)
Rademacher Identity 1

Prove

\[ \mathcal{R}_m(\alpha H) = |\alpha| \mathcal{R}_m(H) \]

If \( \alpha \geq 0 \)

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

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

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

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

If \( \alpha < 0 \)

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

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

\[
(-\alpha) \sup_{h \in H} \sum_{i=1}^{m} (-\sigma_i) h(x_i) \quad (11)
\]
Rademacher Identity 2

Prove

\[ \mathcal{R}_m(H + H') = \mathcal{R}_m(H) + \mathcal{R}_m(H') \]
Rademacher Identity 2

Prove

\[ \mathcal{R}_m(H + H') = \mathcal{R}_m(H) + \mathcal{R}_m(H') \]

(12)

\[ \mathcal{R}_m(H + H') \]

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

(13)

(14)
Rademacher Identity 2

Prove

\[ \mathcal{R}_m(H + H') = \mathcal{R}_m(H) + \mathcal{R}_m(H') \]

(12)

\[
\mathcal{R}_m(H + H') \\
= \frac{1}{m} \mathbb{E}_{\tilde{\sigma}, S} \left[ \sup_{h \in H, h' \in H'} \sum_{i=1}^{m} \sigma_i(h(x_i) + h'(x_i)) \right] \\
= \frac{1}{m} \mathbb{E}_{\tilde{\sigma}, S} \left[ \sup_{h \in H, h' \in H'} \sum_{i=1}^{m} \sigma_i h(x_i) + \sup_{h \in H, h' \in H'} \sum_{i=1}^{m} \sigma_i h'(x_i) \right]
\]

(13)

(14)

(15)
Rademacher Identity 2

Prove

\[ \mathcal{R}_m(H + H') = \mathcal{R}_m(H) + \mathcal{R}_m(H') \]

\[ \mathcal{R}_m(H + H') \]

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

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

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

![Diagram of hyperplanes in different dimensions and orientations.](image)
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

In general, the VC dimension of $d$-dimensional hyperplanes is $d + 1$
Show that the VC dimension of a finite hypothesis set $H$ is at most $\log |H|$. 
Show that the VC dimension of a finite hypothesis set $H$ is at most $\log |H|$.

- To shatter a set, it means that every point can take on a different binary label $h(x)$. 
Show that the VC dimension of a finite hypothesis set $H$ is at most $\log |H|$.

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

- 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