Fundamentals of Machine Learning

Yassine Laguel

Mail : yassine.laguel@univ-cotedazur.fr

UNIVERSITÉ CÔTE D'AZUR

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \ldots, X_n$ be n centered random variables $\left(\mathbb{E}\left[X_i\right] = 0 \text{ for all } i\right)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \ldots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1,\ldots,n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \dots, X_n$ be n centered random variables $\left(\mathbb{E}\left[X_i\right] = 0 \text{ for all } i\right)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \dots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1,\dots,n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \ldots, X_n$ be n centered random variables $\left(\mathbb{E}\left[X_i\right] = 0 \text{ for all } i\right)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \ldots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1, \ldots, n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \ldots, X_n$ be n centered random variables $(\mathbb{E}[X_i] = 0$ for all $i)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \ldots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1,\ldots,n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.
- Note that we did not need any independence between the $X_i's$.

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \ldots, X_n$ be n centered random variables $(\mathbb{E}[X_i] = 0$ for all $i)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \ldots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1, \ldots, n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.
- Note that we did not need any independence between the $X_i's$.

Proposition : Maximal inequality for bounded losses

Assume the loss ℓ satisfies $0 \leq \ell \leq 1$.

Then,

$$\mathbb{E}\left[\max_{f \in \mathcal{H}} r(f) - R(f)\right] \leqslant \sqrt{\frac{2\log(|\mathcal{H}|)}{n}}.$$

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \ldots, X_n$ be n centered random variables $(\mathbb{E}[X_i] = 0$ for all $i)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \ldots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1,\ldots,n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.
- Note that we did not need any independence between the $X_i's$.

Proposition : Maximal inequality for bounded losses

Assume the loss ℓ satisfies $0 \leq \ell \leq 1$.

Then,

$$\mathbb{E}\left[\max_{f \in \mathcal{H}} r(f) - R(f)\right] \leqslant \sqrt{\frac{2\log(|\mathcal{H}|)}{n}}.$$

Proof : on the black board.

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \dots, X_n$ be n centered random variables $(\mathbb{E}[X_i] = 0$ for all $i)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \dots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1,\dots,n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.
- Note that we did not need any independence between the $X_i's$.

Proposition : Maximal inequality for bounded losses

Assume the loss ℓ satisfies $0 \leq \ell \leq 1$.

Then,

$$\mathbb{E}\left[\max_{f \in \mathcal{H}} r(f) - R(f)\right] \leqslant \sqrt{\frac{2\log(|\mathcal{H}|)}{n}}.$$

Proof : on the black board.

Remarks :

- Taking $\ell \leq c$ with $c > 0$ turns the final bound into $c\sqrt{2\log(|\mathcal{H}|)/n}$.
  $\longrightarrow$ Exercise

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \ldots, X_n$ be n centered random variables $(\mathbb{E}[X_i] = 0 \text{ for all } i)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \ldots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1,\ldots,n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.
- Note that we did not need any independence between the $X_i's$.

Proposition : Maximal inequality for bounded losses

Assume the loss ℓ satisfies $0 \leq \ell \leq 1$.

Then,

$$\mathbb{E}\left[\max_{f \in \mathcal{H}} r(f) - R(f)\right] \leqslant \sqrt{\frac{2\log(|\mathcal{H}|)}{n}}.$$

Proof : on the black board.

Remarks :

- Taking $\ell \leq c$ with $c > 0$ turns the final bound into $c\sqrt{2\log(|\mathcal{H}|)/n}$.
  $\longrightarrow$ Exercise
- Remember that our general goal was to establish a bound of the form

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} r(f) - R(f)\right] \leq \frac{\psi(\mathcal{H})}{n^{\alpha}}$$

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \ldots, X_n$ be n centered random variables $(\mathbb{E}[X_i] = 0$ for all $i)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \ldots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1,\ldots,n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.
- Note that we did not need any independence between the $X_i's$.

Proposition : Maximal inequality for bounded losses

Assume the loss ℓ satisfies $0 \leq \ell \leq 1$.

Then,

$$\mathbb{E}\left[\max_{f \in \mathcal{H}} r(f) - R(f)\right] \leqslant \sqrt{\frac{2\log(|\mathcal{H}|)}{n}}.$$

Proof : on the black board.

Remarks :

- Taking $\ell \leq c$ with $c > 0$ turns the final bound into $c\sqrt{2\log(|\mathcal{H}|)/n}$.
  $\longrightarrow$ Exercise
- Remember that our general goal was to establish a bound of the form

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} r(f) - R(f)\right] \leq \frac{\psi(\mathcal{H})}{n^{\alpha}}$$

- The complexity of the Hypothesis class, catured by $|\mathcal{H}|$ plays a **logarithmic** role in our final statistical bound.

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \ldots, X_n$ be n centered random variables $(\mathbb{E}[X_i] = 0$ for all $i)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \ldots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1, \ldots, n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.
- Note that we did not need any independence between the $X_i's$.

Proposition : Maximal inequality for bounded losses

Assume the loss ℓ satisfies $0 \leq \ell \leq 1$.

Then,

$$\mathbb{E}\left[\max_{f \in \mathcal{H}} r(f) - R(f)\right] \leqslant \sqrt{\frac{2\log(|\mathcal{H}|)}{n}}.$$

Proof : on the black board.

Remarks :

- Taking $\ell \leq c$ with $c > 0$ turns the final bound into $c\sqrt{2\log(|\mathcal{H}|)/n}$.
  $\longrightarrow$ Exercise
- Remember that our general goal was to establish a bound of the form

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} r(f) - R(f)\right] \leq \frac{\psi(\mathcal{H})}{n^{\alpha}}$$

- The complexity of the Hypothesis class, catured by $|\mathcal{H}|$ plays a **logarithmic** role in our final statistical bound.
- The dependency on n scales as $\frac{1}{n^{\alpha}}$ with $\alpha = 0.5$.
  This will be a common characteristic of our upcoming results.

Maximal Inequality for bounded losses

Proposition : Upperbounding the max of bounded r.v.

Let $X_1, \dots, X_n$ be n centered random variables $\left(\mathbb{E}\left[X_i\right] = 0 \text{ for all } i\right)$, such that $X_i \in [a, b]$ almost surely for all $i \in \{1, \dots, n\}$.

Then,

$$\mathbb{E}\left[\max_{i \in \{1, \dots, n\}} X_i\right] \leq \frac{(b-a)}{2}\sqrt{2\log(n)}.$$

Proof : on the black board.

Remark :

- This bound is trivial for n large enough, since all variables are almost surely upperbounded by b.
- Note that we did not need any independence between the $X_i's$.

Proposition : Maximal inequality for bounded losses

Assume the loss ℓ satisfies $0 \leq \ell \leq 1$.

Then,

$$\mathbb{E}\left[\max_{f \in \mathcal{H}} r(f) - R(f)\right] \leqslant \sqrt{\frac{2\log(|\mathcal{H}|)}{n}}.$$

Proof : on the black board.

Remarks :

- Taking $\ell \leq c$ with $c > 0$ turns the final bound into $c\sqrt{2\log(|\mathcal{H}|)/n}$.
  $\longrightarrow$ Exercise
- Remember that our general goal was to establish a bound of the form

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} r(f) - R(f)\right] \leq \frac{\psi(\mathcal{H})}{n^{\alpha}}$$

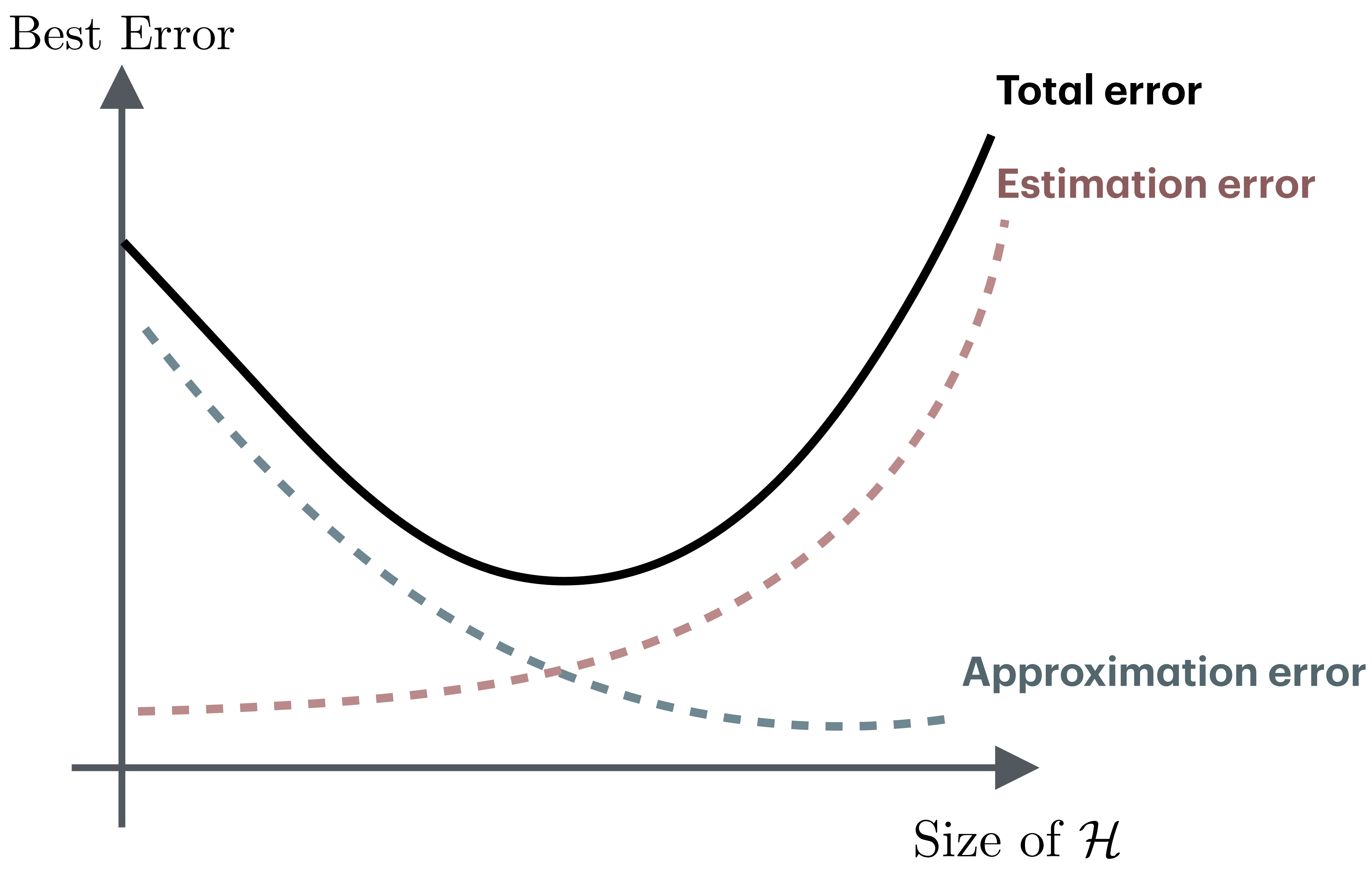
- The complexity of the Hypothesis class, catured by $|\mathcal{H}|$ plays a **logarithmic** role in our final statistical bound.
- The dependency on n scales as $\frac{1}{n^{\alpha}}$ with $\alpha = 0.5$.
  This will be a common characteristic of our upcoming results.
- When $|\mathcal{H}|$ is infinite, this bound is not exploitable.

1. Supervised Learning Setting

2. Estimation vs Approximation

3. Maximal inequalities

4. Rademacher Complexity



Rademacher Complexity of a set

- Rademacher variables and complexity of a set

Rademacher Complexity of a set

- Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Rademacher Complexity of a set

- Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Remark :

-

Rademacher Complexity of a set

Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Remark :

- If $U \sim \mathcal{B}(1/2)$, then $X \triangleq 2U - 1$ is a Rademacher variable.

Definition : Rademacher complexity of a set

Let $\Omega_1, \ldots, \Omega_n$ be n independent Rademacher variables.
The Rademacher complexity of a set $T \subset \mathbb{R}^n$, denoted $\operatorname{Rad}(T)$, is defined as:

$$\operatorname{Rad}(T) = \mathbb{E}\left[\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i\right].$$

Rademacher Complexity of a set

Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Remark :

- If $U \sim \mathcal{B}(1/2)$, then $X \triangleq 2U - 1$ is a Rademacher variable.

Definition : Rademacher complexity of a set

Let $\Omega_1, \dots, \Omega_n$ be n independent Rademacher variables.
The Rademacher complexity of a set $T \subset \mathbb{R}^n$, denoted $\mathrm{Rad}(T)$, is defined as:

$$\mathrm{Rad}(T) = \mathbb{E}\left[\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i\right].$$

Remarks :

- The quantity $\mathrm{Rad}(T)$ measures the complexity of T, as $\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i$

Rademacher Complexity of a set

Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Remark :

- If $U \sim \mathcal{B}(1/2)$, then $X \triangleq 2U - 1$ is a Rademacher variable.

Definition : Rademacher complexity of a set

Let $\Omega_1, \ldots, \Omega_n$ be n independent Rademacher variables.
The Rademacher complexity of a set $T \subset \mathbb{R}^n$, denoted $\mathrm{Rad}(T)$, is defined as:

$$\mathrm{Rad}(T) = \mathbb{E}\left[\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i\right].$$

Remarks :

- The quantity $\mathrm{Rad}(T)$ measures the complexity of T, as $\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i$
- describes how well elements in T can replicate the sign pattern of a random signal $(\Omega_1, \ldots, \Omega_n)$.

Rademacher Complexity of a set

Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Remark :

- If $U \sim \mathcal{B}(1/2)$, then $X \triangleq 2U - 1$ is a Rademacher variable.

Definition : Rademacher complexity of a set

Let $\Omega_1, \ldots, \Omega_n$ be n independent Rademacher variables.
The Rademacher complexity of a set $T \subset \mathbb{R}^n$, denoted $\mathrm{Rad}(T)$, is defined as:

$$\mathrm{Rad}(T) = \mathbb{E}\left[\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i\right].$$

Remarks :

- The quantity $\mathrm{Rad}(T)$ measures the complexity of T, as $\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i$
- describes how well elements in T can replicate the sign pattern of a random signal $(\Omega_1, \ldots, \Omega_n)$.

First properties

Rademacher Complexity of a set

Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Remark :

- If $U \sim \mathcal{B}(1/2)$, then $X \triangleq 2U - 1$ is a Rademacher variable.

Definition : Rademacher complexity of a set

Let $\Omega_1, \ldots, \Omega_n$ be n independent Rademacher variables.
The Rademacher complexity of a set $T \subset \mathbb{R}^n$, denoted $\mathrm{Rad}(T)$, is defined as:

$$\mathrm{Rad}(T) = \mathbb{E}\left[\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i\right].$$

Remarks :

- The quantity $\mathrm{Rad}(T)$ measures the complexity of T, as $\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i$
- describes how well elements in T can replicate the sign pattern of a random signal $(\Omega_1, \ldots, \Omega_n)$.

First properties

Proposition : Homotetie and translation

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.
Then,

$$\mathrm{Rad}(cT + v) = |c| \, \mathrm{Rad}(T).$$

Rademacher Complexity of a set

Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Remark :

- If $U \sim \mathcal{B}(1/2)$, then $X \triangleq 2U - 1$ is a Rademacher variable.

Definition : Rademacher complexity of a set

Let $\Omega_1, \ldots, \Omega_n$ be n independent Rademacher variables.
The Rademacher complexity of a set $T \subset \mathbb{R}^n$, denoted $\mathrm{Rad}(T)$, is defined as:

$$\mathrm{Rad}(T) = \mathbb{E}\left[\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i\right].$$

Remarks :

- The quantity $\mathrm{Rad}(T)$ measures the complexity of T, as $\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i$
- describes how well elements in T can replicate the sign pattern of a random signal $(\Omega_1, \ldots, \Omega_n)$.

First properties

Proposition : Homotetie and translation

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$. Then,

$$\mathrm{Rad}(cT + v) = |c| \, \mathrm{Rad}(T).$$

Proof : on the black board.

Rademacher Complexity of a set

Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}[X = 1] = \mathbb{P}[X = -1] = 1/2.$$

Remark :

- If $U \sim \mathcal{B}(1/2)$, then $X \triangleq 2U - 1$ is a Rademacher variable.

Definition : Rademacher complexity of a set

Let $\Omega_1, \ldots, \Omega_n$ be n independent Rademacher variables.
The Rademacher complexity of a set $T \subset \mathbb{R}^n$, denoted $\mathrm{Rad}(T)$, is defined as:

$$\mathrm{Rad}(T) = \mathbb{E}\left[\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i\right].$$

Remarks :

- The quantity $\mathrm{Rad}(T)$ measures the complexity of T, as $\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i$ describes how well elements in T can replicate the sign pattern of a random signal $(\Omega_1, \ldots, \Omega_n)$.

First properties

Proposition : Homotetie and translation

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$. Then,

$$\mathrm{Rad}(cT + v) = |c| \, \mathrm{Rad}(T).$$

Proof : on the black board.

Proposition : Minkowski sum

For any sets $T, T' \subset \mathbb{R}^d$, we have

$$\mathrm{Rad}\left(T + T'\right) = \mathrm{Rad}(T) + \mathrm{Rad}\left(T'\right).$$

where $T + T' = \{t + t', t \in T, t' \in T'\}$.

Rademacher Complexity of a set

Rademacher variables and complexity of a set

Definition : Rademacher random variables

A Rademacher random variable is a random variable $X : \Omega \to \{-1, 1\}$ such that

$$\mathbb{P}\left[X = 1\right] = \mathbb{P}\left[X = -1\right] = 1/2.$$

Remark :

- If $U \sim \mathcal{B}(1/2)$, then $X \triangleq 2U - 1$ is a Rademacher variable.

Definition : Rademacher complexity of a set

Let $\Omega_1, \ldots, \Omega_n$ be n independent Rademacher variables.
The Rademacher complexity of a set $T \subset \mathbb{R}^n$, denoted $\mathrm{Rad}(T)$, is defined as:

$$\mathrm{Rad}(T) = \mathbb{E}\left[\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i\right].$$

Remarks :

- The quantity $\mathrm{Rad}(T)$ measures the complexity of T, as $\sup_{t \in T} \frac{1}{n} \sum_{i=1}^{n} \Omega_i t_i$
- describes how well elements in T can replicate the sign pattern of a random signal $(\Omega_1, \ldots, \Omega_n)$.

First properties

Proposition : Homotetie and translation

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.
Then,

$$\mathrm{Rad}(cT + v) = |c| \, \mathrm{Rad}(T).$$

Proof : on the black board.

Proposition : Minkowski sum

For any sets $T, T' \subset \mathbb{R}^d$, we have

$$\mathrm{Rad}\left(T + T'\right) = \mathrm{Rad}(T) + \mathrm{Rad}\left(T'\right).$$

where $T + T' = \{t + t', t \in T, t' \in T'\}$.

Proof : on the black board.

More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\mathrm{Rad}(\mathrm{conv}(T)) = \mathrm{Rad}(T).$$

More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\mathrm{Rad}(\mathrm{conv}(T)) = \mathrm{Rad}(T).$$

Proof : on the black board.

More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\mathrm{Rad}(\mathrm{conv}(T)) = \mathrm{Rad}(T).$$

Proof : on the black board.

Proposition : Massart's Lemma - Complexity for finite Sets

Let $T \subset \mathbb{R}^n$ be finite.

Then,

$$\mathrm{Rad}(T) \leq \max_{t \in T} \|t\|_2 \frac{\sqrt{2 \log(|T|)}}{n}.$$

More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\mathrm{Rad}(\mathrm{conv}(T)) = \mathrm{Rad}(T).$$

Proof : on the black board.

Proposition : Massart's Lemma - Complexity for finite Sets

Let $T \subset \mathbb{R}^n$ be finite.

Then,

$$\mathrm{Rad}(T) \leq \max_{t \in T} \|t\|_2 \frac{\sqrt{2\log(|T|)}}{n}.$$

Proof : on the black board.

More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\mathrm{Rad}(\mathrm{conv}(T)) = \mathrm{Rad}(T).$$

Proof : on the black board.

Proposition : Massart's Lemma - Complexity for finite Sets

Let $T \subset \mathbb{R}^n$ be finite.

Then,

$$\mathrm{Rad}(T) \leq \max_{t \in T} \|t\|_2 \frac{\sqrt{2\log(|T|)}}{n}.$$

Proof : on the black board.

Remark :

- The two previous propositions allow to upper bound the Rademacher complexity of the containing polyhedra.

More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\mathrm{Rad}(\mathrm{conv}(T)) = \mathrm{Rad}(T).$$

Proof : on the black board.

Proposition : Massart's Lemma - Complexity for finite Sets

Let $T \subset \mathbb{R}^n$ be finite.

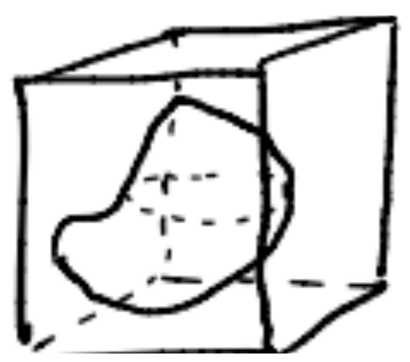
Then,

$$\mathrm{Rad}(T) \leq \max_{t \in T} \|t\|_2 \frac{\sqrt{2 \log(|T|)}}{n}.$$

Proof : on the black board.

Remark :

- The two previous propositions allow to upper bound the Rademacher complexity of the containing polyhedra.



More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\mathrm{Rad}(\mathrm{conv}(T)) = \mathrm{Rad}(T).$$

Proof : on the black board.

Proposition : Massart's Lemma - Complexity for finite Sets

Let $T \subset \mathbb{R}^n$ be finite.

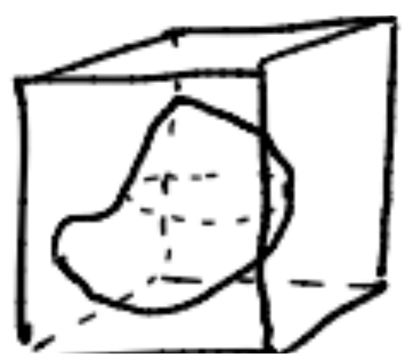
Then,

$$\mathrm{Rad}(T) \leq \max_{t \in T} \|t\|_2 \frac{\sqrt{2\log(|T|)}}{n}.$$

Proof : on the black board.

Remark :

- The two previous propositions allow to upper bound the Rademacher complexity of the containing polyhedra.



Proposition : Talagrand's lemma

Let $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$ be n G-lipchitz functions.

Let $T \subset R^n$ and define $(f_1, \ldots, f_n) \circ T$ as

$$(f_1, \ldots, f_n) \circ T = \{(f_1(t_1), \ldots, f_n(t_n)), t \in T\}.$$

Then,

$$\mathrm{Rad}((f, \ldots, f_n) \circ T) \leqslant G\,\mathrm{Rad}(T).$$

More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\text{Rad}(\text{conv}(T)) = \text{Rad}(T).$$

Proof : on the black board.

Proposition : Massart's Lemma - Complexity for finite Sets

Let $T \subset \mathbb{R}^n$ be finite.

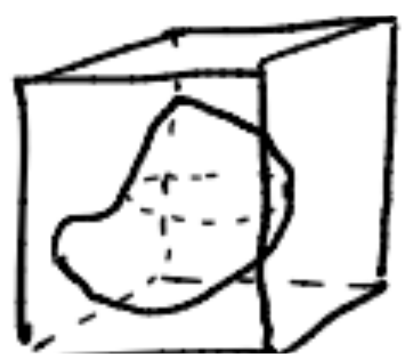
Then,

$$\text{Rad}(T) \leq \max_{t \in T} \|t\|_2 \frac{\sqrt{2\log(|T|)}}{n}.$$

Proof : on the black board.

Remark :

- The two previous propositions allow to upper bound the Rademacher complexity of the containing polyhedra.



Proposition : Talagrand's lemma

Let $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$ be n G-lipchitz functions.

Let $T \subset R^n$ and define $(f_1, \ldots, f_n) \circ T$ as

$$(f_1, \ldots, f_n) \circ T = \{(f_1(t_1), \ldots, f_n(t_n)), t \in T\}.$$

Then,

$$\text{Rad}((f, \ldots, f_n) \circ T) \leqslant G\,\text{Rad}(T).$$

Proof : on the black board.

More properties

Proposition : Convex hull

Let $T \subset \mathbb{R}^n, v \in \mathbb{R}^n, c \in \mathbb{R}$, and define $cT + v = \{ct + v, t \in T\}$.

Then,

$$\operatorname{Rad}(\operatorname{conv}(T)) = \operatorname{Rad}(T).$$

Proof : on the black board.

Proposition : Massart's Lemma - Complexity for finite Sets

Let $T \subset \mathbb{R}^n$ be finite.

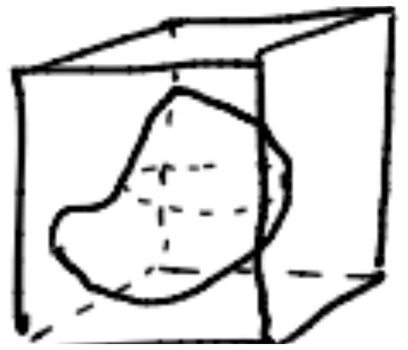
Then,

$$\operatorname{Rad}(T) \leq \max_{t \in T} \|t\|_2 \frac{\sqrt{2\log(|T|)}}{n}.$$

Proof : on the black board.

Remark :

- The two previous propositions allow to upper bound the Rademacher complexity of the containing polyhedra.



Proposition : Talagrand's lemma

Let $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$ be n G-lipchitz functions.

Let $T \subset R^n$ and define $(f_1, \ldots, f_n) \circ T$ as

$$(f_1, \ldots, f_n) \circ T = \{(f_1(t_1), \ldots, f_n(t_n)), t \in T\}.$$

Then,

$$\operatorname{Rad}((f, \ldots, f_n) \circ T) \leqslant G \operatorname{Rad}(T).$$

Proof : on the black board.

Remark :

- This lemma will later be used to establish statistical bounds when the loss ℓ satisfies a Lispchitz condition.

Rademacher complexity of a hypothesis class

- Formal definition & Symmetrization Lemma

Definition : Rademacher complexity of a class of functions

Let $(\Omega_1, \ldots, \Omega_n)$ be n i.i.d. Rademacher variables.

We define the empirical Rademacher complexity of the hypothesis class $\mathcal{H}$, denoted $\widetilde{\mathrm{Rad}}(\mathcal{H})$, as

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) = \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \ell\left(f\left(X_i\right), Y_i\right) \mid \left(X_1, Y_1\right), \ldots, \left(X_n, Y_n\right)\right],$$

and the Rademacher complexity of the class $\mathcal{H}$, denoted $\mathrm{Rad}(\mathcal{H})$, as:

$$\mathrm{Rad}(\mathcal{H}) = \mathbb{E}[\widetilde{\mathrm{Rad}}(\mathcal{H})] = \mathbb{E}\left[\sup_{f_{\in} \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \, \ell\left(f\left(X_i\right), Y_i\right)\right].$$

Rademacher complexity of a hypothesis class

Formal definition & Symmetrization Lemma

Definition : Rademacher complexity of a class of functions

Let $(\Omega_1, \ldots, \Omega_n)$ be n i.i.d. Rademacher variables.

We define the empirical Rademacher complexity of the hypothesis class $\mathcal{H}$, denoted $\widetilde{\mathrm{Rad}}(\mathcal{H})$, as

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) = \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \ell\left(f\left(X_i\right), Y_i\right) \mid \left(X_1, Y_1\right), \ldots, \left(X_n, Y_n\right)\right],$$

and the Rademacher complexity of the class $\mathcal{H}$, denoted $\mathrm{Rad}(\mathcal{H})$, as:

$$\mathrm{Rad}(\mathcal{H}) = \mathbb{E}[\widetilde{\mathrm{Rad}}(\mathcal{H})] = \mathbb{E}\left[\sup_{f_\in \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \, \ell\left(f\left(X_i\right), Y_i\right)\right].$$

Remarks :

- We will use from now on the Rademacher complexity to establish as a measure of complexity of the hypothesis class $\mathcal{H}$.

Rademacher complexity of a hypothesis class

Formal definition & Symmetrization Lemma

Definition : Rademacher complexity of a class of functions

Let $(\Omega_1, \ldots, \Omega_n)$ be n i.i.d. Rademacher variables.

We define the empirical Rademacher complexity of the hypothesis class $\mathcal{H}$, denoted $\widetilde{\mathrm{Rad}}(\mathcal{H})$, as

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) = \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \ell\left(f\left(X_i\right), Y_i\right) \mid \left(X_1, Y_1\right), \ldots, \left(X_n, Y_n\right)\right],$$

and the Rademacher complexity of the class $\mathcal{H}$, denoted $\mathrm{Rad}(\mathcal{H})$, as:

$$\mathrm{Rad}(\mathcal{H}) = \mathbb{E}[\widetilde{\mathrm{Rad}}(\mathcal{H})] = \mathbb{E}\left[\sup_{f_\in \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \, \ell\left(f\left(X_i\right), Y_i\right)\right].$$

Remarks :

- We will use from now on the Rademacher complexity to establish as a measure of complexity of the hypothesis class $\mathcal{H}$.
- A fundamental result in this direction is the symmetrisation lemma.

Rademacher complexity of a hypothesis class

Formal definition & Symmetrization Lemma

Definition : Rademacher complexity of a class of functions

Let $(\Omega_1, \ldots, \Omega_n)$ be n i.i.d. Rademacher variables.

We define the empirical Rademacher complexity of the hypothesis class $\mathcal{H}$, denoted $\widetilde{\mathrm{Rad}}(\mathcal{H})$, as

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) = \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \ell\left(f\left(X_i\right), Y_i\right) \mid \left(X_1, Y_1\right), \ldots, \left(X_n, Y_n\right)\right],$$

and the Rademacher complexity of the class $\mathcal{H}$, denoted $\mathrm{Rad}(\mathcal{H})$, as:

$$\mathrm{Rad}(\mathcal{H}) = \mathbb{E}[\widetilde{\mathrm{Rad}}(\mathcal{H})] = \mathbb{E}\left[\sup_{f \in \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \, \ell\left(f\left(X_i\right), Y_i\right)\right].$$

Remarks :

- We will use from now on the Rademacher complexity to establish as a measure of complexity of the hypothesis class $\mathcal{H}$.
- A fundamental result in this direction is the symmetrisation lemma.

Proposition : Symmetrization lemma

Under the Empirical Risk Minimization Framework, we have :

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} \{r(f) - R(f)\}\right] \leqslant 2 \, \mathrm{Rad}(\mathcal{H}).$$

Rademacher complexity of a hypothesis class

Formal definition & Symmetrization Lemma

Definition : Rademacher complexity of a class of functions

Let $(\Omega_1, \dots, \Omega_n)$ be n i.i.d. Rademacher variables.

We define the empirical Rademacher complexity of the hypothesis class $\mathcal{H}$, denoted $\widetilde{\mathrm{Rad}}(\mathcal{H})$, as

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) = \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \ell\left(f\left(X_i\right), Y_i\right) \mid \left(X_1, Y_1\right), \dots, \left(X_n, Y_n\right)\right],$$

and the Rademacher complexity of the class $\mathcal{H}$, denoted $\mathrm{Rad}(\mathcal{H})$, as:

$$\mathrm{Rad}(\mathcal{H}) = \mathbb{E}[\widetilde{\mathrm{Rad}}(\mathcal{H})] = \mathbb{E}\left[\sup_{f_\in \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \, \ell\left(f\left(X_i\right), Y_i\right)\right].$$

Remarks :

- We will use from now on the Rademacher complexity to establish as a measure of complexity of the hypothesis class $\mathcal{H}$.
- A fundamental result in this direction is the symmetrisation lemma.

Proposition : Symmetrization lemma

Under the Empirical Risk Minimization Framework, we have :

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} \{r(f) - R(f)\}\right] \leqslant 2 \, \mathrm{Rad}(\mathcal{H}).$$

Proof : on the black board.

Rademacher complexity of a hypothesis class

Formal definition & Symmetrization Lemma

Definition : Rademacher complexity of a class of functions

Let $(\Omega_1, \ldots, \Omega_n)$ be n i.i.d. Rademacher variables.

We define the empirical Rademacher complexity of the hypothesis class $\mathcal{H}$, denoted $\widetilde{\mathrm{Rad}}(\mathcal{H})$, as

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) = \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \ell\left(f\left(X_i\right), Y_i\right) \mid \left(X_1, Y_1\right), \ldots, \left(X_n, Y_n\right)\right],$$

and the Rademacher complexity of the class $\mathcal{H}$, denoted $\mathrm{Rad}(\mathcal{H})$, as:

$$\mathrm{Rad}(\mathcal{H}) = \mathbb{E}[\widetilde{\mathrm{Rad}}(\mathcal{H})] = \mathbb{E}\left[\sup_{f_{\in} \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \, \ell\left(f\left(X_i\right), Y_i\right)\right].$$

Remarks :

- We will use from now on the Rademacher complexity to establish as a measure of complexity of the hypothesis class $\mathcal{H}$.
- A fundamental result in this direction is the symmetrisation lemma.

Proposition : Symmetrization lemma

Under the Empirical Risk Minimization Framework, we have :

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} \{r(f) - R(f)\}\right] \leqslant 2\,\mathrm{Rad}(\mathcal{H}).$$

Proof : on the black board.

Applications in the regression setting

Proposition : Complexity for Lipschitz Losses

Under the Empirical Risk Minimization Framework, if $\hat{y} \mapsto \ell(\hat{y}, y)$ is G-lipschitz, then

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) \leq G\, \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \Omega_i \; f\left(X_i\right) \mid \; X_1, \ldots, X_n\right].$$

Rademacher complexity of a hypothesis class

Formal definition & Symmetrization Lemma

Definition : Rademacher complexity of a class of functions

Let $(\Omega_1, \ldots, \Omega_n)$ be n i.i.d. Rademacher variables.

We define the empirical Rademacher complexity of the hypothesis class $\mathcal{H}$, denoted $\widetilde{\mathrm{Rad}}(\mathcal{H})$, as

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) = \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \ell\left(f\left(X_i\right), Y_i\right) \mid (X_1, Y_1), \ldots, (X_n, Y_n)\right],$$

and the Rademacher complexity of the class $\mathcal{H}$, denoted $\mathrm{Rad}(\mathcal{H})$, as:

$$\mathrm{Rad}(\mathcal{H}) = \mathbb{E}[\widetilde{\mathrm{Rad}}(\mathcal{H})] = \mathbb{E}\left[\sup_{f_\in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{N} \Omega_i \, \ell\left(f\left(X_i\right), Y_i\right)\right].$$

Remarks :

- We will use from now on the Rademacher complexity to establish as a measure of complexity of the hypothesis class $\mathcal{H}$.
- A fundamental result in this direction is the symmetrisation lemma.

Proposition : Symmetrization lemma

Under the Empirical Risk Minimization Framework, we have :

$$\mathbb{E}\left[\sup_{f \in \mathcal{H}} \{r(f) - R(f)\}\right] \leqslant 2\,\mathrm{Rad}(\mathcal{H}).$$

Proof : on the black board.

Applications in the regression setting

Proposition : Complexity for Lipschitz Losses

Under the Empirical Risk Minimization Framework, if $\hat{y} \mapsto \ell(\hat{y}, y)$ is G-lipschitz, then

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) \leq G\, \mathbb{E}\left[\sup_{f \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^{n} \Omega_i \, f\left(X_i\right) \mid X_1, \ldots, X_n\right].$$

Proof : on the black board.

Rademacher complexity of a hypothesis class

Formal definition & Symmetrization Lemma

Definition : Rademacher complexity of a class of functions

Let $(\Omega_1, \ldots, \Omega_n)$ be n i.i.d. Rademacher variables.

We define the empirical Rademacher complexity of the hypothesis class $\mathcal{H}$, denoted $\widetilde{\mathrm{Rad}}(\mathcal{H})$, as

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) = \mathbb{E}\left[\sup_{f\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{N} \Omega_i \ell\left(f\left(X_i\right), Y_i\right) \mid (X_1, Y_1), \ldots, (X_n, Y_n)\right],$$

and the Rademacher complexity of the class $\mathcal{H}$, denoted $\mathrm{Rad}(\mathcal{H})$, as:

$$\mathrm{Rad}(\mathcal{H}) = \mathbb{E}[\widetilde{\mathrm{Rad}}(\mathcal{H})] = \mathbb{E}\left[\sup_{f\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{N} \Omega_i\, \ell\left(f\left(X_i\right), Y_i\right)\right].$$

Remarks :

- We will use from now on the Rademacher complexity to establish as a measure of complexity of the hypothesis class $\mathcal{H}$.
- A fundamental result in this direction is the symmetrisation lemma.

Proposition : Symmetrization lemma

Under the Empirical Risk Minimization Framework, we have :

$$\mathbb{E}\left[\sup_{f\in\mathcal{H}}\{r(f) - R(f)\}\right] \leqslant 2\,\mathrm{Rad}(\mathcal{H}).$$

Proof : on the black board.

Applications in the regression setting

Proposition : Complexity for Lipschitz Losses

Under the Empirical Risk Minimization Framework, if $\hat{y} \mapsto \ell(\hat{y}, y)$ is G-lipschitz, then

$$\widetilde{\mathrm{Rad}}(\mathcal{H}) \leq G\,\mathbb{E}\left[\sup_{f\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \Omega_i\, f\left(X_i\right) \mid X_1, \ldots, X_n\right].$$

Proof : on the black board.

Remarks :

- Hence, assuming assuming a Lipschitz error function allows us to focus on the complexity spanned by the set of estimators rather than the composition $\ell(f(X), Y)$.