Obtaining Upper Bounds for the Approximation with Radial Basis Functions using Proof Techniques from Statistical Learning Theory

Dominik Köhler

Bayreuth, November 8, 2022

Introduction

Notation

$Z \subset \mathbb{R}^k$, $\mathcal{F} \subset \{f : Z \to [A, B] \text{ measurable}\}$, $\mathcal{M} \subset \{\phi : \mathbb{R}_0^+ \to [A, B] \text{ m.}\}$,

$z_1, \ldots, z_n \in Z$, $\lambda \in \mathcal{L}_1(Z) \geq 0, \|\lambda\|_1 = 1$, $\delta \in (0, 1)$

Introduction

Notation

$Z \subset \mathbb{R}^k$, $\quad \mathcal{F} \subset \{f : Z \to [A, B] \textit{ measurable}\}$, $\quad \mathcal{M} \subset \{\phi : \mathbb{R}_0^+ \to [A, B] \textit{ m.}\}$,

$z_1, \ldots, z_n \in Z$, $\quad \lambda \in \mathcal{L}_1(Z) \geq 0, \|\lambda\|_1 = 1$, $\quad \delta \in (0, 1)$

Goal: Find an upper bound for the uniform convergence of:

$$\sup_{f \in \mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq \alpha_{n,\mathcal{F}}, \quad \alpha_{n,\mathcal{F}} \overset{n\to\infty}{\longrightarrow} 0$$

Introduction

Notation

$Z \subset \mathbb{R}^k$, $\quad \mathcal{F} \subset \{f : Z \to [A,B] \textit{ measurable}\}$, $\quad \mathcal{M} \subset \{\phi : \mathbb{R}_0^+ \to [A,B] \textit{ m.}\}$,

$z_1, \dots, z_n \in Z$, $\quad \lambda \in \mathcal{L}_1(Z) \geq 0, \|\lambda\|_1 = 1$, $\quad \delta \in (0,1)$

Goal: Find an upper bound for the uniform convergence of:

$$\sup_{f \in \mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq \alpha_{n,\mathcal{F}}, \quad \alpha_{n,\mathcal{F}} \overset{n\to\infty}{\longrightarrow} 0$$

Method: Use bounds from statistical learning theory

$$P\left\{ z_1, \dots, z_n \in Z : \sup_{f \in \mathcal{F}} \left| E[f(z)] - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq \alpha_{n,\delta,\mathcal{F}} \right\} \geq 1 - \delta, \quad \alpha_{n,\delta,\mathcal{F}} \overset{n\to\infty}{\longrightarrow} 0$$

Introduction

Notation

$Z \subset \mathbb{R}^k$, $\mathcal{F} \subset \{f : Z \to [A, B] \text{ measurable}\}$, $\mathcal{M} \subset \{\phi : \mathbb{R}_0^+ \to [A, B] \text{ m.}\}$,

$z_1, \ldots, z_n \in Z$, $\lambda \in \mathcal{L}_1(Z) \geq 0, \|\lambda\|_1 = 1$, $\delta \in (0, 1)$

Goal: Find an upper bound for the uniform convergence of:

$$\sup_{f \in \mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq \alpha_{n,\mathcal{F}}, \quad \alpha_{n,\mathcal{F}} \xrightarrow{n\to\infty} 0$$

Method: Use bounds from statistical learning theory

$$P\left\{ z_1, \ldots, z_n \in Z : \sup_{f \in \mathcal{F}} \left| E[f(z)] - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq \alpha_{n,\delta,\mathcal{F}} \right\} \geq 1 - \delta, \quad \alpha_{n,\delta,\mathcal{F}} \xrightarrow{n\to\infty} 0$$

Use bounds for radial functions

$$\sup_{\phi \in \mathcal{M}, t \in Z} \left| \int_z \phi(\|z - t\|_2)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} \phi(\|z_i - t\|_2) \right| \leq C(B - A)\sqrt{\frac{c_{\mathcal{M}}(2k + 3)}{n}}$$

Outline

Definitions

Notation

$$X \subset \mathbb{R}^k, \qquad Y \subset \mathbb{R}, \qquad P \textit{ unknown probability measure on } X \times Y,$$
$$(x_1, y_1), \ldots, (x_n, y_n), (x, y) \in (X \times Y) \textit{ i.i.d. following } P$$

- Goal: Find a function f, which explains the distribution of the points:
$$(x, f(x)) \approx (x, y).$$
- To measure the error, we use a *loss function*:
$$L(x, y, f(x)) : X \times Y \times \mathbb{R} \to \mathbb{R}_0^+, \qquad L \text{ measurable},$$
and get the *empirical risk* on the data $(x_1, y_1), \ldots, (x_n, y_n)$:
$$\widehat{R}(f) := \frac{1}{n} \sum_{i=1}^{n} L(x_i, y_i, f(x_i))$$
and for the error on $X \times Y$ we get the *risk*:
$$R(f) := \int_{X \times Y} L(x, y, f(x)) dP(x, y)$$

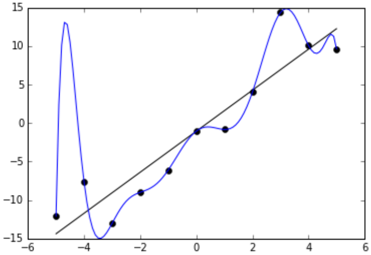
Bounding the risk

Question

Is it possible to generalize from the empirical risk to the risk for a set of functions $\mathcal{F} \subset \{f : X \to Y \text{ measurable}\}$?

$$P\left\{(x_1, y_1), \ldots, (x_n, y_n) \in X \times Y : \sup_{f \in \mathcal{F}} \left| R(f) - \widehat{R}(f) \right| \leq \alpha_{n,\delta,\mathcal{F}} \right\} \geq 1 - \delta?$$

Not in general:



To find suitable sets $\mathcal{F}$, we use *complexity measures*.

Relaxed Notation

We substitute:

$$
\begin{aligned}
X \times Y &\mapsto Z \subset \mathbb{R}^k \\
L(x, y, f(x)) &\mapsto f(z), && f : Z \to \mathbb{R} \\
P &\mapsto \lambda(z), && \lambda \in \mathcal{L}_1(Z) \geq 0, \|\lambda\|_1 = 1 \\
R(f) = \int_{X \times Y} L(x, y, f) dP(x, y) &\mapsto \int_Z f(z)\lambda(z) dz \\
\widehat{R}(f) = \frac{1}{n} \sum_{i=1}^{n} L(x_i, y_i, f) &\mapsto \frac{1}{n} \sum_{i=1}^{n} f(z_i)
\end{aligned}
$$

Example for Classification

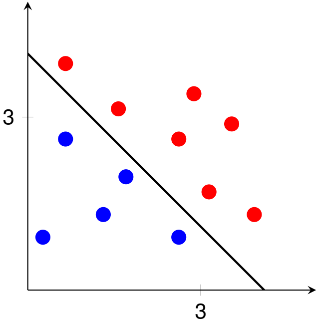
Notation

$S = \{z_1, \ldots, z_n\} \subset Z \subset \mathbb{R}^k, \qquad \mathcal{F}_{0,1} \subset \{f : Z \to \{0, 1\} \textit{ measurable}\}$

Example for Classification

Notation

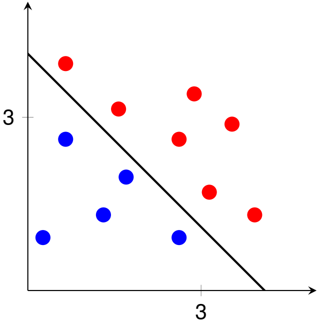
$S = \{z_1, \dots, z_n\} \subset Z \subset \mathbb{R}^k, \qquad \mathcal{F}_{0,1} \subset \{f : Z \to \{0,1\} \textit{ measurable}\}$

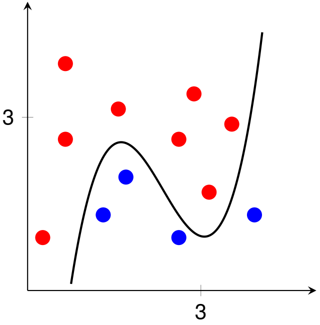


Example for Classification

Notation

$S = \{z_1, \ldots, z_n\} \subset Z \subset \mathbb{R}^k, \qquad \mathcal{F}_{0,1} \subset \{f : Z \to \{0,1\} \text{ measurable}\}$





Using Realvalued Functions for Classification

Notation

$$\mathcal{F} \subset \{f : Z \to \mathbb{R}\}, \quad \mathcal{H} : \mathbb{R} \to \{0, 1\}, f(z) \mapsto \begin{cases} 1, & f(z) \geq 0, \\ 0, & f(z) < 0. \end{cases}$$

Definition

- *To induce: $S^+ \subset S$ is induced by $\mathcal{F}$, if it exists $f \in \mathcal{F}$ and $\beta \in \mathbb{R}$, such that:*

$$\begin{cases} z \in S^+ & \implies \mathcal{H}(f(z) - \beta) = 1 \\ z \in S \setminus S^+ & \implies \mathcal{H}(f(z) - \beta) = 0 \end{cases}$$

- *To shatter: S is shattered by $\mathcal{F}$, if all 2^n subsets of S are induced by $\mathcal{F}$.*

The Vapnik-Chervonenkis-Dimension

Definition

$$\mathscr{G}_{\mathcal{F},S} := \textit{ the number of subsets of } S \textit{ induced by } \mathcal{F}, \quad \mathscr{G}_{\mathcal{F},S} \leq 2^n$$

The Vapnik-Chervonenkis-Dimension

Definition

$$\mathscr{G}_{\mathcal{F},S} := \textit{ the number of subsets of } S \textit{ induced by } \mathcal{F}, \quad \mathscr{G}_{\mathcal{F},S} \leq 2^n$$

The growth function of $\mathcal{F}$ is defined as:

$$\mathscr{G}_{\mathcal{F}}(n) := \max_{S \subset Z, \#S = n} \mathscr{G}_{\mathcal{F},S}$$

The Vapnik-Chervonenkis-Dimension

Definition

$$\mathscr{G}_{\mathcal{F},S} := \textit{ the number of subsets of } S \textit{ induced by } \mathcal{F}, \quad \mathscr{G}_{\mathcal{F},S} \leq 2^n$$

The growth function of $\mathcal{F}$ is defined as:

$$\mathscr{G}_{\mathcal{F}}(n) := \max_{S \subset Z, \#S = n} \mathscr{G}_{\mathcal{F},S}$$

Definition

$$\mathrm{VCD}(\mathcal{F}) := \begin{cases} \max\{n \in \mathbb{N}_0 \mid \mathscr{G}_{\mathcal{F}}(n) = 2^n\}, & \textit{if existent,} \\ \infty, & \textit{elsewise.} \end{cases}$$

$\mathrm{VCD}(\mathcal{F}) \geq n \Leftrightarrow \exists S \subset Z, \#S = n$: all 2^n subsets of S are induced by $\mathcal{F}$.

The Upper Bound

Notation

$\mathcal{F} \subset \{f : Z \to [A, B]\}$, *C constant independent of* $\mathcal{F}$, $\lambda \in \mathcal{L}_1(Z) \geq 0, \|\lambda\|_1 = 1$.

It holds:

- For all $\delta \in (0, 1)$ it holds with a probability of at least $1 - \delta$:

$$\sup_{f \in \mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq C(B-A)\frac{1}{\sqrt{n}} \max\left\{ \sqrt{\mathrm{VCD}(\mathcal{F})}, -\frac{\ln(\delta)}{\sqrt{n}}, \sqrt{-\ln(\delta)} \right\}.$$

The Upper Bound

Notation

$\mathcal{F} \subset \{f : Z \to [A, B]\}$, *C constant independent of* $\mathcal{F}$, $\lambda \in \mathcal{L}_1(Z) \geq 0, \|\lambda\|_1 = 1$.

It holds:

- For all $\delta \in (0, 1)$ it holds with a probability of at least $1 - \delta$:

$$\sup_{f \in \mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq C(B-A)\frac{1}{\sqrt{n}} \max\left\{ \sqrt{\mathrm{VCD}(\mathcal{F})}, -\frac{\ln(\delta)}{\sqrt{n}}, \sqrt{-\ln(\delta)} \right\}.$$

- There exist points $\tilde{z}_1, \ldots, \tilde{z}_n \in Z$, such that:

$$\sup_{f \in \mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(\tilde{z}_i) \right| \leq C(B-A)\sqrt{\frac{\mathrm{VCD}(\mathcal{F})}{n}}.$$

VC-Dimension for Radial Functions

A set of radial functions

$$\mathcal{R} := \left\{ \Phi : Z \to \mathbb{R}, (\cdot, t) \mapsto \phi(\|\cdot - t\|_2), \quad \phi : \mathbb{R}_0^+ \to \mathbb{R} \text{ measurable}, t \in Z \subset \mathbb{R}^k \right\}$$

can be seen as the composition of two sets $\mathcal{R} = \mathcal{M} \circ \mathcal{F}_{Ball}$:

$$\mathcal{F}_{Ball} = \{g : Z \to \mathbb{R}_0^+, z \mapsto \|z - t\|_2, \qquad t \in Z\},$$
$$\mathcal{M} = \{\phi : \mathbb{R}_0^+ \to \mathbb{R}, x \mapsto \phi(x), \qquad \phi \text{ measurable}\}.$$

Lemma

If all functions of $\mathcal{M}$ are monotone, it holds:

$$\mathrm{VCD}(\mathcal{M} \circ \mathcal{F}_{Ball}) \leq 2\,\mathrm{VCD}(\mathcal{F}_{Ball}) + 1$$

The Monotonic Requirement

With this idea we decompose ϕ:

$$\phi(z) = \sum_{i=1}^{p} \phi|_{\mathcal{I}_i}(z), \qquad \begin{array}{l} \mathcal{I}_1 \cup \ldots \cup \mathcal{I}_p = \mathbb{R}_0^+, \quad \mathcal{I}_i \cap \mathcal{I}_j = \emptyset, \\ \phi \text{ monotone on } \mathcal{I}_i, \quad 1 \leq i, j \leq p. \end{array} \tag{1}$$

We set

$$\mathcal{M}_p := \left\{\phi : \mathbb{R}_0^+ \to [A, B], \phi \text{ satisfies (1) for } p \in \mathbb{N}\right\}$$

and get:

Theorem

$$\mathrm{VCD}(\mathcal{M}_p \circ \mathcal{F}_{Ball}) \leq c_p \left(2\,\mathrm{VCD}(\mathcal{F}_{Ball}) + 1\right), \quad c_p \in \mathcal{O}(p \log(p))$$

With $\mathrm{VCD}(\mathcal{F}_{Ball}) = k + 1$, it follows:

$$\mathrm{VCD}(\mathcal{M}_p \circ \mathcal{F}_{Ball}) \leq c_p(2k + 3).$$

Result for Radial Functions

Notation

$$\mathcal{F} = \mathcal{M}_p \circ \mathcal{F}_{Ball} = \{\phi(\|\cdot - t\|_2) : Z \to [A, B], \quad \phi \in \mathcal{M}_p, t \in Z\}$$

Theorem

With the probability of at least $1 - \delta$ we find points $z_1, \ldots z_n \in Z$, such that:

$$\sup_{f \in \mathcal{F}} \left| R(f) - \widehat{R}(f) \right| \leq C(B - A) \frac{1}{\sqrt{n}} \max \left\{ \sqrt{c_p(2k+3)}, -\frac{\ln(\delta)}{\sqrt{n}}, \sqrt{-\ln(\delta)} \right\}.$$

Summary

We used bounds from statistical learning theory:

$$P\left\{z_1,\dots,z_n \in Z : \sup_{f\in\mathcal{F}} \left| E[f(z)] - \frac{1}{n}\sum_{i=1}^{n} f(z_i)\right| \leq \alpha_{n,\delta,\mathcal{F}}\right\} \geq 1-\delta, \quad \alpha_{n,\delta,\mathcal{F}} \xrightarrow{n\to\infty} 0$$

We transferred them into approximation theory:

$$\sup_{f\in\mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(\tilde{z}_i)\right| \leq C(B-A)\frac{1}{\sqrt{n}}\sqrt{\mathrm{VCD}(\mathcal{F})}$$

We applied them to Radial Basis Functions $\mathcal{F} = \{\phi(\cdot - t) : Z \to \mathbb{R},\ \phi \in \mathcal{M}_p, t \in Z\}$:

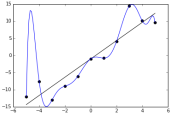
$$\mathrm{VCD}(\mathcal{F}) \leq c_p(2k+3)$$

Key Takeaways

$$\sup_{f \in \mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq \alpha_{\mathrm{VCD}(\mathcal{F}),[A,B]} \tag{1}$$

1. (1) holds on the fixed points $z_1, \ldots, z_n$ uniformly for all $f \in \mathcal{F}$
2. the RHS of (1) only depends on the VCD of $\mathcal{F}$, which we calculated for RBFs
3. Connection between statistical learning theory and approximation theory

Table of Figures



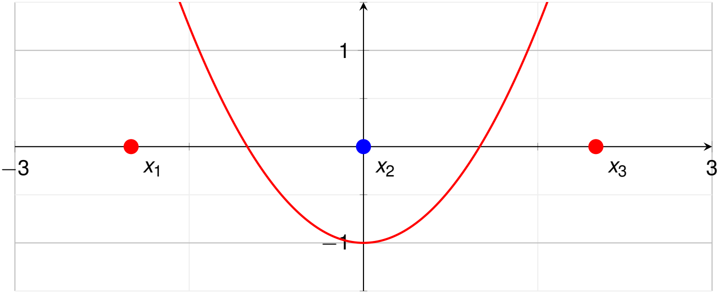
Sources

Federico Girosi
Approximation Error Bounds That Use Vc-Bounds
Proc. Internat. Conf. Artificial Neural Networks 1995.

Giorgio Gnecco und Marcello Sanguineti
Approximation error bounds via Rademacher's complexity
Applied Mathematical Sciences, 2008.

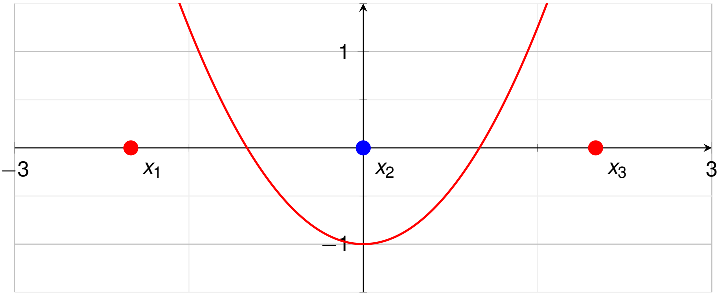
Example: Parabulas pointing "up"

$$\mathcal{F} := \{f : \mathbb{R} \to \mathbb{R}, x \mapsto (x-a)(x-b), \quad a, b \in \mathbb{R}\}$$



Example: Parabulas pointing "up"

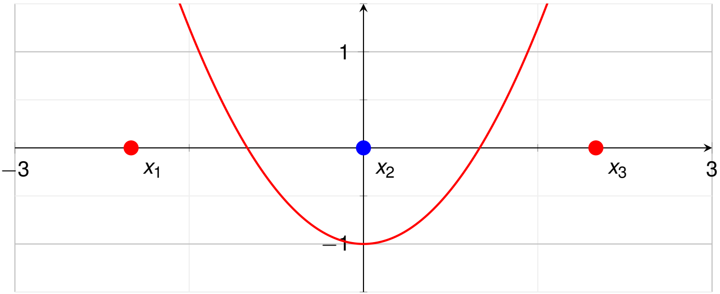
$$\mathcal{F} := \{f : \mathbb{R} \to \mathbb{R}, x \mapsto (x-a)(x-b), \quad a, b \in \mathbb{R}\}$$



- $\{x_2\} \subset \{x_1, x_2, x_3\}$ is not induced by $\mathcal{F}$ $\implies$ $\mathrm{VCD}(\mathcal{F}) < 3$

Example: Parabulas pointing "up"

$$\mathcal{F} := \{f : \mathbb{R} \to \mathbb{R}, x \mapsto (x-a)(x-b), \quad a, b \in \mathbb{R}\}$$



- $\{x_2\} \subset \{x_1, x_2, x_3\}$ is not induced by $\mathcal{F}$ $\implies$ $\mathrm{VCD}(\mathcal{F}) < 3$
- $\{x_1, x_3\}$ is shattered by $\mathcal{F}$ $\implies$ $\mathrm{VCD}(\mathcal{F}) \geq 2$

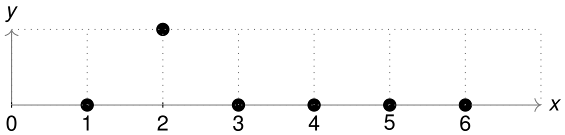
Hence:

$$\mathrm{VCD}(\mathcal{F} = 2)$$

Example: Indicator functions on $\mathbb{N}$

$$\mathcal{F} := \{f_i(j) : \mathbb{N} \to \{0,1\}, f_i(j) = \delta_{i,j}, i \in \mathbb{N}\}.$$

Looking at $f_2 \in \mathcal{F}$:



Calculating the Growth function:

$$\mathscr{G}_{\mathcal{F}}(n) = n + 1.$$

Hence:

$$\mathrm{VCD}(\mathcal{F}) = \max\{n \in \mathbb{N}_0 | \mathscr{G}_{\mathcal{F}}(n) = 2^n\} = 1.$$

Application for the Span of Functions

$$\mathcal{S} := \left\{ \sum_{i=1}^{m} a_i f_i(z) : a_1, \ldots, a_m \in \mathbb{R} \right\}, \ \mathcal{S}_M := \left\{ \sum_{i=1}^{m} a_i f_i(z) : \sum_{i=1}^{m} |a_i| \leq M, a_1, \ldots, a_m \in \mathbb{R} \right\}$$

Lemma

$$\sup_{f \in \mathcal{S}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(\tilde{z}_i) \right| \leq (B-A)C\sqrt{\frac{m+1}{n}},$$

$$\sup_{f \in \mathcal{S}_M} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(z) \right| \leq (B-A)CM\max\{|A|, |B|\}\sqrt{\frac{2\ln(2m)}{n}}.$$

2 Bounds: through $\mathscr{G}$ und R.C.

With the probability of at least $(1-\delta)$ it holds:

$$\sup_{f \in \mathcal{F}} \left| \int_Z f(z)\lambda(z)dz - \frac{1}{n}\sum_{i=1}^{n} f(z_i) \right| \leq \alpha.$$

- Bound using the Growth function:

$$\alpha = (B-A)\sqrt{8\frac{\ln\left(\mathscr{G}_{\mathcal{F}}(n)\right) - \ln\left(\frac{\delta}{4}\right)}{n}}.$$

- Bound using the Rademacher Comlexity

$$\alpha = (B-A)C\frac{1}{\sqrt{n}}\max\left\{\frac{\mathcal{R}_n(\mathcal{F}-A)}{B-A}, -\frac{\ln(\delta)}{\sqrt{n}}, \sqrt{-\ln(\delta)}\right\}.$$

It holds:

$$\mathscr{G}_{\mathcal{F}}(n) \leq \left(\frac{en}{\mathrm{VCD}(\mathcal{F})}\right)^{\mathrm{VCD}(\mathcal{F})}$$

$$\mathcal{R}_n(\mathcal{F}-A) \leq C(B-A)\sqrt{\mathrm{VCD}(\mathcal{F})}.$$

Proofing the Growth-Function-Bound - Idea

Idea:

1. Use *Hoeffding*-inequality with $E[Z_i] = \int_Z Z_i(z)dP(z)$:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i(z) - E[Z_i]\right| \geq \epsilon\right) \leq 2\exp\left(-2n\epsilon^2(B-A)^{-2}\right).$$

2. Use the union-Bound $P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$ for finitly many functions:

$$P(A_1 > \epsilon \cup A_2 > \epsilon) = P\left(\sup_{A \in \{A_1, A_2\}} A > \epsilon\right).$$

Proofing the Growth-Function-Bound

Notation

$\mathcal{F}_{\mathcal{H}} = \{\mathcal{H}(f(\cdot - \beta)), f \in \mathcal{F}, \beta \in \mathbb{R}\}$

1. It is sufficient to only use finitly many functions in $\mathcal{F}$:
   - Restrict to functions with image in $\{0, 1\}$
   - Only calculate the functions on $2n$ points

   Hence, we look at at most 2^{2n} functions:

$$P\left(\left\{\sup_{f\in\mathcal{F}}|R(f) - \widehat{R}(f)| > \epsilon\right\}\right) \leq P\left(\left\{\sup_{f\in\mathcal{F}_{\mathcal{H}}}\left|\int_Z f(z)dP(z) - \frac{1}{n}\sum_{i=1}^{n} f(z_i)\right| > \frac{\varepsilon}{B-A}\right\}\right)$$
$$\leq 2P\left(\sup_{f\in\mathcal{F}_{\mathcal{H}}}\left|\frac{1}{n}\sum_{i=1}^{n} f(z_i) - \frac{1}{n}\sum_{i=1}^{n} f(z_{i+n})\right| > \frac{\epsilon}{2(B-A)}\right)$$

2. Hoeffding-Inequality with $E\left[\frac{1}{n}\sum_{i=1}^{n} f(z_i) - f(z_{i+n})\right] = 0$ and $[A, B] = [0, 1]$

Proof of the RC-Bound

Notation

$$D := \sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} (f(x_i) - E_P[f(x)]) \right|, \quad \mathcal{R}_n(\mathcal{F}) := E_P E_\epsilon \left[\sup_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \epsilon_i f(x_i) \right| \right]$$

1. With the probability of at least $(1-\delta)$ it holds:

$$nD \leq 2E_P[nD] + C\left(\sqrt{-n\ln(\delta)} + \ln(\delta)\right)$$

2. Symmetrization:

$$E_P[D] \leq 2\frac{\mathcal{R}_n(\mathcal{F})}{\sqrt{n}}.$$

3. Hence, with the probability of at least $(1-\delta)$ it holds:

$$\sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(z_i) - E_P[f(z)] \right| \leq \frac{4\mathcal{R}_n(\mathcal{F})}{\sqrt{n}} + \frac{\tilde{C}\sqrt{-\ln(\delta)}}{\sqrt{n}} - \frac{\tilde{C}}{n}\ln(\delta)$$

$$\leq (B-A)C\frac{1}{\sqrt{n}} \max\left\{\mathcal{R}_n(\mathcal{F}), -\frac{\ln(\delta)}{\sqrt{n}}, \sqrt{-\ln(\delta)}\right\}.$$

Bounding the RC by the VCD

Notation

$\hat{\mathcal{R}}_S(\mathcal{F}) := E_\epsilon \left[\sup_{f \in \mathcal{F}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \epsilon_i f(x_i) \right| \right], \quad \mathcal{R}_n(\mathcal{F}) := E_P \left[\hat{\mathcal{R}}_S(\mathcal{F}) \right],$
$\mathcal{F} \subset \{f : Z \to [A, B]\}$

To prove:

$$\mathcal{R}_n(\mathcal{F}) \leq (B - A) C \sqrt{\mathrm{VCD}(\mathcal{F})}.$$

1. Use Dudleys Integral:

$$\hat{\mathcal{R}}(\mathcal{F}) \leq \hat{C} \int_0^\infty (\ln \mathcal{N}(\varepsilon, \mathcal{F}, \mathcal{L}_2(\mu_n)))^{\frac{1}{2}} \, d\varepsilon$$

2. Furthermore:

$$\mathcal{N}(\mathcal{F}, \mathcal{L}_2(\mu), \epsilon) \leq \left(\frac{2}{\varepsilon} \right)^{C_1 \left(\frac{2}{\varepsilon} - 1 \right) \mathrm{VCD}(\mathcal{F})}$$