Leanability with VC-dimensions

Daniel Khashabi

Fall 2013
Last Update: September 25, 2016

1 VC-dimension

Let $\mathcal{F}$ be class of functions defined from $\mathcal{X}$ to $\{-1, 1\}$ [*]. Let $X = (X_1, \ldots, X_n)$ be a set of samples. Given $\mathcal{F}$, define the following:

$$S_{\mathcal{F}}(X) \triangleq \{h(X_1), \ldots, h(X_n)\}$$

Definition 1.1 (Growth Function)**.**

$$\Pi_{\mathcal{F}}(n) \triangleq \max_{\mathcal{S}:|\mathcal{S}|=n} |S_{\mathcal{F}}(X)|$$

By the above definitions we know,

$$\Pi_{\mathcal{F}}(n) \leq |\mathcal{F}|$$
$$\Pi_{\mathcal{F}}(n) \leq |2^n|$$

Definition 1.2 (Shattering)**.** *A hypothesis class $\mathcal{F}$ shatters a finite set $S \subset \mathcal{X}$, iff $|S_{\mathcal{F}}(S)| = 2^{|S|}$*

In an informal language, shattering means that you have to be able to separate all +/- labelings of the same set of points, given the function class.

We want to bound Rademacher average using a function of VC-dimension.

Lemma 1.1. *Let $\mathcal{F}$ be a class of functions defined on X to $\{+1, -1\}$, then,*

$$R_n(\mathcal{F}) \leq \sqrt{\frac{2 \log 2\Pi_{\mathcal{F}}(n)}{n}}$$

If the function space is symmetric, i.e. given $f \in \mathcal{F}$ then $-f \in \mathcal{F}$:

$$R_n(\mathcal{F}) \leq \sqrt{\frac{2 \log \Pi_{\mathcal{F}}(n)}{n}}$$

Proof. Proof with finite class lemma. ∎

Definition 1.3 (VC-dimension)**.** *The **Vapnik-Chervonenkis dimension** of a class $\mathcal{F}$ on a set X, is the cardinality of the largest set shattered by $\mathcal{F}$, that is, the largest n such that there exists a set $S \subset X$, and $|S| = n$ that $\mathcal{F}$ shatters the set S. We denote VC-dimension with $d_{VC}(\mathcal{F})$.*

[*] The output being mapped to ± 1 is just for simplicity and holds for any binary functions.

Example 1.2. Here we provide a couple of example hypothesis families with their corresponding VC-dimension.

- $\mathcal{H}$ is axis parallel rectangles, $\mathcal{X}$ is $\mathbb{R}^2$.
  VC Dimension 4
- $\mathcal{H}$ is axis-parallel rectangles, $\mathcal{X}$ is $\mathbb{R}^3$.
  VC Dimension 6
  Hint: first show that it can shatter the 6 points (1,0,0),(0,1,0),(0,0,1),(1,0,0),(0,1,0),(0,0,1). If we draw a bounding box for these points, then by excluding/including each point by moving a face of the box, we can get any labeling for the points. For 7 points, consider the bounding box. If the bounding rectangle has at least one point in its interior, then we cannot accomplish the labeling where the interior point is labeled - and the rest are labeled $+$.
- $\mathcal{H}$ is axis-parallel rectangles, $\mathcal{X}$ is $\mathbb{R}^d$.
  VC Dimension $2d$
- $\mathcal{H}$ is the union of 2 intervals, $\mathcal{X}$ is $\mathbb{R}$.
  VC Dimension 4.
- $\mathcal{H}$ is $1\{a\sin(x) > 0\}$, $\mathcal{X}$ is $\mathbb{R}$.
  VC Dimension 1.
- $\mathcal{H}$ is $1\{\sin(x+a) > 0\}$, $\mathcal{X}$ is $\mathbb{R}$.
  VC Dimension 2.
- $\mathcal{H}$ is half-spaces in $\mathbb{R}^n$.
  VC dimension $n+1$.

Proposition 1.1. *A finite concept class C has VC dimension at most $\log|C|$. The different number of ways d points can be labeled with $\pm$ is 2^d.*

We will use the following lemma to find another bound:

Lemma 1.3. *For any $d \leq n$, we have*

$$\sum_{i=1}^{k} \binom{n}{i} \leq \left(\frac{en}{k}\right)^k$$

Proof.

$$\begin{aligned}
\left(\frac{k}{n}\right)^k \sum_{i=1}^{k} \binom{n}{i} &\leq \sum_{i=1}^{k} \left(\frac{k}{n}\right)^i \binom{n}{i} \\
&\leq \sum_{i=1}^{n} \binom{n}{i} \left(\frac{k}{n}\right)^i \times 1^{n-i} \\
&\leq \left(1+\frac{k}{n}\right)^n \leq e^k
\end{aligned}$$

$$\Rightarrow \sum_{i=1}^{k} \binom{n}{i} \leq \left(\frac{en}{k}\right)^k$$

■

Now using the above lemma, we find a bound on the growth function.

Lemma 1.4 (Sauer's lemma)**.** *Let F be a class of functions, mapping from X to a binary space, with $d_{VC}(\mathcal{F}) = d$. Then,*

$$\Pi_{\mathcal{F}}(n) \leq \sum_{i=0}^{d} \binom{n}{i}$$

in addition, for any $n \geq d$

$$\Pi_{\mathcal{F}}(n) \leq \left(\frac{en}{d}\right)^d$$

Proof. The second part of the lemma is trivial based Lemma 1.3. The proof of the first part is by induction on $d + n$. ■

Example 1.5. For a class of real-valued functions $\mathcal{F}$ on $\mathbb{R}$, we define

$$R_n(\mathcal{F}) \triangleq \sup_{z^n \in Z^n} R_n(\mathcal{F}(z^n))$$

For each of the following function classes, prove the Rademacher averages, without relying on the VC-theory,

- $\mathcal{F}_1$ the collection of indicators of all semi-infinite intervals of the form $(-\infty, t], t \in \mathbb{R}$.

$$R_n(\mathcal{F}_1) \leq 2\sqrt{\frac{\log(n+1)}{n}}, \quad \forall n$$

- $\mathcal{F}_2$ is the collection of indicators of all closed intervals of the form $[s, t]$ for $-\infty < s < t < +\infty$.

$$R_n(\mathcal{F}_2) \leq 2\sqrt{\frac{2\log(n) + \log 2}{n}}, \quad \forall n$$

- $\mathcal{F}_3$ is the collection of indicators of all subsets of R that can be represented as unions of no more than k disjoint intervals from $\mathcal{F}_2$.

$$R_n(\mathcal{F}_3) \leq 2\sqrt{\frac{k\log(ne/k)}{n}}, \quad \forall n \geq k$$

Example 1.6. The *sup-norm* for any space of functions is defined as

$$||f|| \triangleq \sup_{z \in Z} |f(z)|$$

Given $\mathcal{F}$ the class of positive-valued functions on Z and $\epsilon > 0$, the $\epsilon - net$ of $\mathcal{F}$ with respect to the *sup-norm* is any $f \in \mathcal{F}$, such that,

$$||f - f_j||_\infty = \sup_{z \in Z} |f(z) - f_j(z)| \leq \epsilon,$$

for at least one of the functions in $\mathcal{F}' = \{f_1, \ldots, f_k\}$, which are not necessarily in $\mathcal{F}$. The ϵ*-covering number* of $\mathcal{F}$ w.r.t. to the *sup-norm*, or the cardinality of a minimal $\epsilon - net$ of $\mathcal{F}$, is denoted by $N_\infty(\mathcal{F}, \epsilon)$. If $\mathcal{F}$ does not accept ϵ-net $N_\infty(\mathcal{F}, \epsilon) = \infty$. The logarithm of the ϵ-covering number, is usually called ϵ-number of $\mathcal{F}$ and denoted by $H_\infty(\mathcal{F}, \epsilon)$.

1. For $\mathcal{F}$ the family of uniformly-bounded functions (i.e. $\exists L > 0 \quad s.t. \quad \forall f \in \mathcal{F} \Rightarrow ||f||_\infty \leq L$). Show that,
$$R_n(\mathcal{F}) \leq \inf_{\epsilon > 0} \left(\epsilon + 2L \sqrt{\frac{\log N_\infty(\mathcal{F}, \epsilon)}{n}} \right)$$
2. Let
$$Z = \left\{ \left(z^{(1)}, \ldots, z^{(d)} \right) \in \mathbb{R}^d : ||z||_1 = \sum_{j=1}^{d} z^{(j)} \leq 1 \right\},$$
and $\mathcal{F}$ consisting of functions of the form $f(z) = f_w(z) = \langle w, z \rangle$, for all $w \in \mathbb{R}^d$ with $||w||_\infty = \max_{1 \leq j \leq d} \left| w^{(j)} \right| \leq 1$.
Show that
$$N_\infty(\mathcal{F}, \epsilon) \leq \left(\frac{2}{\epsilon} \right)^d$$
and prove that,
$$R_n(\mathcal{F}) = O\left(\sqrt{\frac{d \log n}{n}} \right).$$
3. Suppose $\mathcal{F}$ is such that $H_\infty(\mathcal{F}, \epsilon) \leq C\epsilon^{-\frac{1}{\alpha}}$ for some constant $C > 0$ and $\alpha > 0$. For example,
   - the class of functions $\mathcal{F}$ all differentiable $f : [0, 1] \to [0, 1]$ with $|f'| \leq 1$, then the above bound holds with $\alpha = 1$.

   Prove that
$$R_n(\mathcal{F}) \leq Cn^{-\frac{\alpha}{2\alpha+1}}, \quad C > 0$$

Lemma 1.7 (Holder's inequality)**.** *Consider f and g are two measurable real-valued functions defined on a measurable space. Let $p, q \in [1, +\infty]$, and $1/p + 1/q = 1$. Then,*
$$||fg||_1 \leq ||f||_p ||g||_q$$
The theorem also holds in the extremal cases, when $p = \infty$, $q = 1$.

Lemma 1.8. *For non-negative random variable Z, if we know,*
$$\mathbb{P}(Z \geq t) \leq Ce^{-2nt^2},$$
for some universal constant $C > 0$ and $C < +\infty$, one can show that,
$$\mathbb{E}[Z] \leq \sqrt{\frac{\ln(Ce)}{2n}}$$

Proof. Proof in the Section 4. ∎

Example 1.9. Let X be real-valued random variable with CDF $F(x) = \mathbb{P}(X \leq x)$. If $X_1, \ldots, X_n$ are i.i.d. copies of X, the empirical CDF is,

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}.$$

1. Using the Rademacher complexity techniques prove that

$$\mathbb{E}\left[\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F_n(x)\right|\right] \leq \frac{C}{\sqrt{n}}, \quad C > 0$$

2. If $C = 1$, prove *Massart*'s inequality,

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F_n(x)\right| > t\right) \leq 2e^{-2nt^2}, \quad \forall t > 0.$$

Note: it can be shown $C = 1$ is optimal.

Lemma 1.10. *For non-negative random variable Z, if we know,*

$$\mathbb{P}(Z \geq t) \leq Ce^{-2nt^2},$$

for some universal constant $C > 0$ and $C < +\infty$, one can show that,

$$\mathbb{E}[Z] \leq \sqrt{\frac{\ln(Ce)}{2n}}$$

Proof. Proof in the Section 4. ∎

Example 1.11. Let X be real-valued random variable with CDF $F(x) = \mathbb{P}(X \leq x)$. If $X_1, \ldots, X_n$ are i.i.d. copies of X, the empirical CDF is,

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}.$$

1. Using the Rademacher complexity techniques prove that

$$\mathbb{E}\left[\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F_n(x)\right|\right] \leq \frac{C}{\sqrt{n}}, \quad C > 0$$

2. If $C = 1$, prove *Massart*'s inequality,

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F_n(x)\right| > t\right) \leq 2e^{-2nt^2}, \quad \forall t > 0.$$

Note: it can be shown $C = 1$ is optimal.

Theorem 1.12. *Let F be class of function defined on space X mapped to Y, from an unknown distribution $\mathbb{P}_{XY}$. Then given i.i.d. samples $\{(X_i, Y_i)\}_{i=1}^n$, then with probability at least $1-\delta$:*

$$\begin{aligned}
\mathbb{P}(h(X) \neq Y) &\leq \hat{\mathbb{P}}(h(x_i) \neq y_i) + 4R_n(\mathcal{F}) + \sqrt{\frac{1/\delta}{2n}} \\
&\leq \hat{\mathbb{P}}(h(x_i) \neq y_i) + 4\sqrt{\frac{2d\log(en) - 2d\log d}{n}} + \sqrt{\frac{1/\delta}{2n}}
\end{aligned}$$

This could be genelized to all of the samples,

$$\mathbb{P}(h(X) \neq Y) \leq \hat{\mathbb{P}}(h(X) \neq Y) + O\left(\sqrt{\frac{d\log n + \log 1/\delta}{n}}\right)$$

2 Exercise Problems

1. Which of the following procedures is sufficient and necessary and most efficient for proving that the VC dimension of a learner is N?
   (a) Show that the classifier can shatter all possible dichotomies with N points.
   (b) Show that the classifier can shatter a subset of all possible dichotomies with N points.
   (c) Show that the classifier can shatter all possible dichotomies with N points and that it cannot shatter any of the dichotomies with N+1 points.
   (d) Show that the classifier can shatter all possible dichotomies with N points and that it cannot shatter one of the dichotomies with N+1 points.
   (e) Show that the classifier can shatter a subset of all possible dichotomies with N points and that it cannot shatter one of the dichotomies with N+1 points.
2. Find the VC-dimension of the following: $f(w^\top w + \theta)$, where f is an arbitrary increasing non-linear function ($w, x \in \mathbb{R}^d$ and $\theta \in \mathbb{R}$)
3. For the concept class in the previous question, find the minimum number of training instances (sample complexity) necessary to learn a hypothesis with error at most ϵ with probability at least $1-\delta$.

3 Bibliographical notes

References

[1] Maxim Raginsky. Lecture notes: Ece 299: Statistical learning theory. *Tutorial*, 2011.

4 Proofs

4.1 Proof of lemma 1.10

Proof. To prove this we use the fact that, the variance of a non-negative random variable is non-negative. Thus,

$$\mathbb{E}[Z]^2 \leq \mathbb{E}\left[Z^2\right].$$

Using this fact,

$$\begin{aligned}
\mathbb{E}[Z]^2 \leq \mathbb{E}\left[Z^2\right] &= \int_0^{+\infty} \mathbb{P}\left(Z^2 \geq t\right) dt \\
&\leq \int_0^z 1dt + \int_z^{+\infty} \mathbb{P}\left(Z \geq \sqrt{t}\right) dt \\
&\leq z + \int_z^{+\infty} Ce^{-2nt} dt \\
&= z + \frac{C}{2n} e^{-2nz}
\end{aligned}$$

$$\Rightarrow \mathbb{E}[Z] \leq \sqrt{\inf_{z \in \mathbb{R}^+} \left\{ z + \frac{C}{2n} e^{-2nz} \right\}}$$

Now, since this bound holds for any $z \in \mathbb{R}^+$, we minimize it with respect to z to find a tighter bound.

$$\frac{\partial}{\partial z}\left(z + \frac{C}{2n} e^{-2nz}\right) = 1 + \frac{C}{2n}(-2n)e^{-2nz} = 0$$

$$\Rightarrow z = \frac{1}{2n} \ln C \Rightarrow \mathbb{E}[Z] \leq \sqrt{\frac{\ln e \times C}{2n}}$$

■

5 Answers

Here answers to some of the questions are included. The answers are mostly by the authors, and might be buggy. Therefore, read cautiously!

5.1 Answers to example 1.5

To answer this we use generalization of *the finite class lemma.* We use Lemma **??** to prove each of the following cases. In fact, the only thing that we need to do, is to count the number of the distinct values in $\mathcal{F}(Z^n)$.

1. For the class of functions of the form $\mathbf{1}\{X \geq t\}$, the possible configuration of the values is shown

$$
n+1 \text{ cases}: \begin{cases} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 1 & 1 \end{cases}
$$

Just plugging-in in the Lemma **??**, it gives us,

$$
R_n(\mathcal{F}_1) \leq 2\sqrt{\frac{\log|\mathcal{F}(Z^n)|}{n}} = 2\sqrt{\frac{\log(n+1)}{n}}, \quad \forall n
$$

2. Now consider the class of the functions of the form, $\mathbf{1}\{s \geq X \geq t\}$. Now, we want to count the number of distinct values could be produced by this function. This class consists of binary strings of lengths n, with consecutive sequence of 1's, and the rest being zero. For example,

$$
\underbrace{0,0,0,\dots,0,0,0,\overbrace{1,1,1,\dots,1,1,1}^{k \text{ consecutive 1's}},0,0,0,\dots,0,0,0}_{\text{length } n}
$$

The number of distinct such binary sequences could easily be counted. The count equals to the number of the ways we can put two separating partitions between object (and before the first object, and after the last object), which equals to $\binom{n}{2}$, plus one for everything being zero. Then,

$$
|\mathcal{F}(Z^n)| = \binom{n}{2} + 1 \Rightarrow \log|\mathcal{F}(Z^n)| = \frac{n(n+1)}{2} + 1 \leq 2n^2
$$

$$
R_n(\mathcal{F}_2) \leq 2\sqrt{\frac{\log(2n^2)}{n}} \leq 2\sqrt{\frac{2\log(n) + \log 2}{n}}
$$

3. We use the Sour's lemma. The number of distinct subsets of size at most k elements, among n elements, equal to,

$$
\sum_{i=1}^{k} \binom{n}{i}
$$

which, based on Lemma 1.3, is upper-bounded by $\left(\frac{en}{k}\right)^k$. Using the finite class lemma, this gives us the following Rademacher bound,

$$
R_n(\mathcal{F}_3) \leq 2\sqrt{\frac{k\log(ne/k)}{n}}, \quad \forall n \geq k
$$

5.2 Answer to example 1.6

5.2.1 First part:

We start with the definition of the Rademacher complexity, and bound it from above. We use the property given in the problem that, for any function $f \in \mathcal{F}$, there exits a function $f_j \in \mathcal{F}'$ such that

$\sup_{z \in Z} |f(z) - f_j(z)| \leq \epsilon$. Given an arbitrary $f \in \mathcal{F}$,

$$
\begin{aligned}
\left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i f(Z_i) \right| &= \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i \left(f(Z_i) - f_j(Z_i) + f_j(Z_i) \right) \right| \\
&\leq \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i \left(f(Z_i) - f_j(Z_i) \right) \right| + \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i f_j(Z_i) \right| \\
&\leq \epsilon + \max_{1 \leq j \leq k} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i f_j(Z_i) \right|
\end{aligned}
$$

For a given set of samples Z^n,

$$
\begin{aligned}
R_n(\mathcal{F}(Z^n)) &= \mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i f(Z_i) \right| \\
&\leq \mathbb{E} \left\{ \epsilon + \max_{1 \leq j \leq k} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i f_j(Z_i) \right| \right\} \\
&= \epsilon + \mathbb{E} \max_{1 \leq j \leq k} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i f_j(Z_i) \right| \\
&\leq \epsilon + 2L \sqrt{\frac{\log N_\infty(\mathcal{F}(Z^n), \epsilon)}{n}}
\end{aligned}
$$

To be more accurate, the last bound above is, $\epsilon + 2(L+\epsilon)\sqrt{\frac{\log N_\infty(\mathcal{F}(Z^n),\epsilon)}{n}}$. Since the covering functions aren't necessarily bounded by L (but bounded by $L + \epsilon$ instead). But for any set of covering function f_j which is $|f_j(z)| > L$, we can limit it to $\pm L$, and it will still be a covering:

$$
f'(z) = \begin{cases} L & f_j(z) > L \\ -L & f_j(z) < -L \\ 0 & \text{otherwise} \end{cases}
$$

This will bound the covering functions to L and will give the desired bound. The Rademacher average over the whole class,

$$
R_n(\mathcal{F}) = \sup_{z^n \in \mathsf{Z}^n} R_n(\mathcal{F}(z^n)) \leq \epsilon + 2L \sqrt{\frac{\log N_\infty(\mathcal{F}, \epsilon)}{n}}
$$

Now we can tighten the bound for an arbitrary value of $\epsilon > 0s$,

$$
R_n(\mathcal{F}) \leq \inf_{\epsilon \in \mathbb{R}^+} \left\{ \epsilon + 2L \sqrt{\frac{\log N_\infty(\mathcal{F}, \epsilon)}{n}} \right\}
$$

5.2.2 Second part:

Before starting the proof, we state Holder's inequality without proof. We will use this theorem during the proof.

We show that one can find a family of functions $\mathcal{G}$ with size $(2/\epsilon)^n$ such that for any functions $f \in \mathcal{F}$, there exists $g \in \mathcal{G}$, and $||f - g||_\infty \leq \epsilon$. To show this, it is enough to show it for special case of $\mathcal{G}$, though there might be better answers. For that, we define the following class functions:

$$\mathcal{G}(\epsilon) = \left\{ \langle w.z \rangle | z \in \mathsf{Z}, w \in \mathcal{W}(\epsilon)^d \right\},$$

in which $\mathcal{W}(\epsilon) = \{\pm \epsilon k | k \in [1, ..., 1/\epsilon]\}$. Based on the above definition,

$$\forall w \in [-1, 1], \exists w' \in \mathcal{W},\ s.t.\ |w - w'| \leq \epsilon$$

$$\Rightarrow \forall w \in [-1, 1]^d, \exists w' \in \mathcal{W}^d,\ s.t.\ ||w - w'||_\infty = \max_{1 \leq j \leq d} |w - w'| \leq \epsilon$$

Also note that, $|\mathcal{G}(\epsilon)| = (2/\epsilon)^d$. Now, for any arbitrary function $f \in \mathcal{F}$, we choose the function $g \in \mathcal{G}$ which has smallest $||f - g||_\infty$. For this function, for a given $z \in \mathsf{Z}$,

$$\begin{aligned} \forall z \in \mathsf{Z} \quad |f(z) - g(z)| &\leq |w.z - w'.z| \\ &\leq |w - w'|_\infty |z|_1 \quad \text{(Holder's inequality)} \\ &\leq \epsilon \times 1 \Rightarrow ||f - g||_\infty = \sup_{z \in \mathsf{Z}} |f(z) - g(z)| \leq \epsilon \end{aligned}$$

$$\Rightarrow \forall f \in \mathcal{F}, \exists g \in \mathcal{G},\ s.t.\ ||f - g||_\infty \leq \epsilon$$

This end our proof that, $N_\infty(\mathcal{F}, \epsilon) \leq (2/\epsilon)^d$.

Now we use the result of the previous part, and plug-in $N_\infty(\mathcal{F}, \epsilon)$,

$$\begin{aligned} R_n(\mathcal{F}) &\leq \inf_{\epsilon \in \mathbb{R}^+} \left\{ \epsilon + 2\sqrt{\frac{\log N_\infty(\mathcal{F}, \epsilon)}{n}} \right\} \\ &\leq \inf_{\epsilon \in \mathbb{R}^+} \left\{ \epsilon + 2\sqrt{\frac{d \ln \frac{2}{\epsilon}}{n}} \right\} \end{aligned}$$

If we choose $\epsilon = \frac{2}{n}$,

$$R_n(\mathcal{F}) \leq \epsilon + 2\sqrt{\frac{d \ln \frac{2}{\epsilon}}{n}} = \frac{2}{n} + 2L\sqrt{\frac{d \ln n}{n}}$$

Thus,

$$R_n(\mathcal{F}) \leq \frac{2}{n} + 2\sqrt{\frac{d \ln n}{n}} \leq C_2 \sqrt{\frac{d \ln n}{n}}, \text{ for } C_2 \text{ big enough.}$$

It can be shown that $C_2 = 4$ satisfies the above property. Now we prove this. For any $d \geq 1$, and for any $n \geq 1$, we have

$$\frac{2}{n} \leq 2\sqrt{\frac{1}{n}} \leq 2\sqrt{\frac{d \ln n}{n}} \Rightarrow \frac{2}{n} + 2\sqrt{\frac{d \ln n}{n}} \leq 4\sqrt{\frac{d \ln n}{n}},$$

$$\Rightarrow R_n(\mathcal{F}) \leq 4\sqrt{\frac{d \ln n}{n}}.$$

Which proves,

$$R_n(\mathcal{F}) = O\left(\sqrt{\frac{d \ln n}{n}}\right)$$

5.2.3 Third part:

Given the assumption, we know,

$$\ln N_\infty(\mathcal{F}, \epsilon) \leq C\epsilon^{-\frac{1}{\alpha}}.$$

By plugging this into the result of the first part,

$$\begin{aligned} R_n(\mathcal{F}) &\leq \inf_{\epsilon \in \mathbb{R}^+} \left\{ \epsilon + 2L\sqrt{\frac{\log N_\infty(\mathcal{F}, \epsilon)}{n}} \right\} \\ &\leq \inf_{\epsilon \in \mathbb{R}^+} \left\{ \epsilon + 2L\sqrt{\frac{C\epsilon^{-\frac{1}{\alpha}}}{n}} \right\} \end{aligned}$$

Now, choosing $\epsilon = n^{-\frac{\alpha}{2\alpha+1}}$,

$$\epsilon + 2L\sqrt{\frac{C\epsilon^{-\frac{1}{\alpha}}}{n}} = n^{-\frac{\alpha}{2\alpha+1}} + 2L\sqrt{C}n^{-\frac{\alpha}{2\alpha+1}} = \left(1 + 2L\sqrt{C}\right) n^{-\frac{\alpha}{2\alpha+1}}$$

$$\Rightarrow R_n(\mathcal{F}) \leq C' n^{-\frac{\alpha}{2\alpha+1}} \text{ for } C' \geq 1 + 2L\sqrt{C}$$

5.3 Answer to example 1.9

5.3.1 First part:

To prove this bound, we use the notion of the VC-dimension. We know,

$$\mathbb{E}R_n(\mathcal{F}(Z^n)) \leq C\sqrt{\frac{V(\mathcal{F})}{n}}$$

We define $\mathcal{F}$ to be set of indicators on half-interval. It is easy to show that, the VC-dimension for a class of functions consisting of half-intervals is one. Also, we can treat the CDF function, as empirical risk for a set of half-interval functions. By this assumption, we know that, using the symmetrization trick we can find the following bound (proof in Lemma **??**):

$$\mathbb{E}\left[\sup_{x \in \mathbb{R}} \left|\hat{F}_n(x) - F_n(x)\right|\right] \leq 2\mathbb{E}R_n(\mathcal{F}(Z^n))$$

Using these facts, we can find the following bound,

$$\mathbb{E}\left[\sup_{x \in \mathbb{R}} \left|\hat{F}_n(x) - F_n(x)\right|\right] \leq 2\mathbb{E}R_n(\mathcal{F}(Z^n)) \leq \frac{C}{\sqrt{n}}, \quad \text{for some } C > 0$$

5.3.2 Second part:

To prove this, we first prove a lemma.

We use the Lemma 1.10. Defining $Z = \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_n(x) \right|$, we know that,

$$\mathbb{E}\left[Z\right] \leq \sqrt{\frac{\ln\left(e \times 2\right)}{2n}} \leq \sqrt{\frac{2\ln\left(e\right)}{2n}} = \frac{1}{\sqrt{n}}.$$

5.4 Answer to example 1.11

5.4.1 First part:

To prove this bound, we use the notion of the VC-dimension. We know,

$$\mathbb{E}R_n(\mathcal{F}(Z^n)) \leq C\sqrt{\frac{V(\mathcal{F})}{n}}$$

We define $\mathcal{F}$ to be set of indicators on half-interval. It is easy to show that, the VC-dimension for a class of functions consisting of half-intervals is one. Also, we can treat the CDF function, as empirical risk for a set of half-interval functions. By this assumption, we know that, using the symmetrization trick we can find the following bound (proof in Lemma **??**):

$$\mathbb{E}\left[\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_n(x) \right| \right] \leq 2\mathbb{E}R_n(\mathcal{F}(Z^n))$$

Using these facts, we can find the following bound,

$$\mathbb{E}\left[\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_n(x) \right| \right] \leq 2\mathbb{E}R_n(\mathcal{F}(Z^n)) \leq \frac{C}{\sqrt{n}}, \qquad \text{for some } C > 0$$

5.4.2 Second part:

We use the Lemma 1.10. Defining $Z = \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_n(x) \right|$, we know that,

$$\mathbb{E}\left[Z\right] \leq \sqrt{\frac{\ln\left(e \times 2\right)}{2n}} \leq \sqrt{\frac{2\ln\left(e\right)}{2n}} = \frac{1}{\sqrt{n}}.$$

References

[1] Maxim Raginsky. Lecture notes: Ece 299: Statistical learning theory. *Tutorial*, 2011.