Chapter 20
Sample Error (Variance)

Recall that the sample error is bounded by

$$\mathscr{S}\left(\hat{h}_N\right) := \mathscr{R}\left(\hat{h}_N\right) - \mathscr{R}\left(\hat{h}\right) \leq \left|\mathscr{R}\left(\hat{h}_N\right) - \mathscr{R}_N\left(\hat{h}_N\right)\right| + \left|\mathscr{R}_N\left(\hat{h}\right) - \mathscr{R}\left(\hat{h}\right)\right|. \quad (20.1)$$

The two terms on the right hand side of (20.1) are the sampling errors for $\mathscr{R}\left(\hat{h}_N\right)$ and $\mathscr{R}\left(\hat{h}\right)$ using Monte Carlo. Chapter **??** tells us that these errors for finite sample size[1] N can be bounded probalistically.

20.1 Sampling error for a function in $\mathscr{H}$

Lemma 20.1. *Assume that $\mathscr{H}$ is M-bounded in the sense of Assumption 17.1. For any $h \in \mathscr{H}$, there holds*

$$\mathbb{P}\left[\left|\mathscr{R}_N(h) - \mathscr{R}(h)\right| > t\right] \leq 2e^{-2N\frac{t^2}{M^4}}, \quad \forall t \geq 0.$$

Proof. The proof is obvious using the M-boundedness of $\mathscr{H}$, the i.i.d. assumption 16.1 on the training set S, and the result of Exercise 14.3. □

20.2 Sampling error for finite $\mathscr{H}$

We now extend Section 20.1 to hypothesis space $\mathscr{H}$ containing finite number of functions $h_1, \ldots, h_{\mathscr{N}}$. The task at hand is to probabilistly bound the worst error among $h_1, \ldots, h_{\mathscr{N}}$.

Lemma 20.2. *Let $\mathscr{H} = \{h_1, \ldots, h_{\mathscr{N}}\}$, H is M-bounded, and the training set S is i.i.d. in the sense of Assumption 16.1. There holds:*

[1] Again, this is the beauty of cencontration of measures for non-asymptotic theory.

$$\mathbb{P}\left[\left|\mathscr{R}_N(h)-\mathscr{R}(h)\right|>\varepsilon\right]\le 2\mathscr{N}e^{-2N\frac{\varepsilon^2}{M^4}},\quad \forall\varepsilon\ge 0.$$

Proof. We start with the following observation

$$\sup_{h\in\mathscr{H}}\left|\mathscr{R}_N(h)-\mathscr{R}(h)\right|>\varepsilon \iff \exists i\le m:\left|\mathscr{R}_N(h_i)-\mathscr{R}(h_i)\right|>\varepsilon,$$

and hence the following identity of on the equivalent events

$$\begin{aligned}\left\{\sup_{h\in\mathscr{H}}\left|\mathscr{R}_N(h)-\mathscr{R}(h)\right|>\varepsilon\right\}&=\left\{\exists i\le m:\left|\mathscr{R}_N(h_i)-\mathscr{R}(h_i)\right|>\varepsilon\right\}\\&=\cup_{i=1}^{\mathscr{N}}\left\{\left|\mathscr{R}_N(h_i)-\mathscr{R}(h_i)\right|>\varepsilon\right\}.\end{aligned}$$

This leads to

$$\begin{aligned}&\mathbb{P}\left[\sup_{h\in\mathscr{H}}\left|\mathscr{R}_N(h)-\mathscr{R}(h)\right|>\varepsilon\right]=\mathbb{P}\left[\exists i\le m:\left|\mathscr{R}_N(h_i)-\mathscr{R}(h_i)\right|>\varepsilon\right]\\&=\mathbb{P}\left[\cup_{i=1}^{\mathscr{N}}\left\{\left|\mathscr{R}_N(h_i)-\mathscr{R}(h_i)\right|>\varepsilon\right\}\right]\le\sum_{i=1}^{\mathscr{N}}\mathbb{P}\left[\left|\mathscr{R}_N(h_i)-\mathscr{R}(h_i)\right|>\varepsilon\right]\le 2\mathscr{N}e^{-2N\frac{\varepsilon^2}{M^4}},\end{aligned}$$

where we have used the union bound in Lemma 12.3 in the second last inequality, and the i.i.d. nature of S together with Lemma 20.1 in the last inequality. □

20.3 Sampling error when $\mathscr{H}$ is a ball with radius R

Let $\mathscr{H}:=\{h\in\mathbb{C}(X):\|h\|_\infty\le R\}$. Due to the continuity of the sampling error in Exercise 17.2, i.e.,

$$\left|\mathscr{E}(h_1)-\mathscr{E}(h_2)\right|\le 4M\left\|h_1-h_2\right\|_\infty,\quad\forall h_1,h_2\in\mathscr{H},$$

we have

$$\mathscr{E}(h_c)\ge\mathscr{E}(h)-4MR,$$

where h_c is the center of the ball and h is an arbitrary function inside the ball. Thus

$$\mathscr{E}(h_c)\ge\sup_{h\in\mathscr{H}}\mathscr{E}(h)-4MR,$$

which, in turn, shows that if the following implication holds true:

$$\sup_{h\in\mathscr{H}}\left|\mathscr{E}(h)\right|\ge\varepsilon\Rightarrow\left|\mathscr{E}(h_c)\right|\ge\varepsilon-4MR.$$

We conclude that

$$\mathbb{P}\left[\left\{\sup_{h\in\mathscr{H}}|\mathscr{E}(h)|\geq\varepsilon\right\}\right]\leq\mathbb{P}\left[\left\{|\mathscr{E}(h_c)|\geq\varepsilon-4MR\right\}\right]\leq 2e^{-2N\frac{(\varepsilon-4MR)^2}{M^4}},$$

where we have used Lemma 20.1 in the last inequality. We thus have proved the following result.

Lemma 20.3. *Assume $\mathscr{H}:=\{h\in\mathbb{C}(X):\|h\|_\infty\leq R\}$, and $\mathscr{H}$ is M-bounded in the sense of Assumption 17.1. Then the tail of the worst sampling error decays exponentially, i.e.,*

$$\mathbb{P}\left[\left\{\sup_{h\in\mathscr{H}}|\mathscr{E}(h)|\geq\varepsilon\right\}\right]\leq 2e^{-2N\frac{(\varepsilon-4MR)^2}{M^4}}.$$

20.4 Sampling error when $\mathscr{H}$ is a union of $\mathscr{N}$ balls

Without loss of generality we assume that all the balls have the same radius of $\varepsilon/8M$, i.e., $\mathscr{H}=\cup_{i=1}^{\mathscr{N}}B_{h_i}\left(\frac{\varepsilon}{8M}\right)$.

Lemma 20.4. *Assume $\mathscr{H}$ is M-bounded in the sense of Assumption 17.1. Then the tail of the worst sampling error decays exponentially, i.e.,*

$$\mathbb{P}\left[\left\{\sup_{h\in\mathscr{H}}|\mathscr{E}(h)|\geq\varepsilon\right\}\right]\leq 2\mathscr{N}e^{-N\frac{\varepsilon^2}{2M^4}}.$$

Proof. We proceed with the following observation

$$\sup_{h\in\mathscr{H}}|\mathscr{E}(h)|\geq\varepsilon\iff\exists i\leq\mathscr{N}:\sup_{h\in B_{h_i}\left(\frac{\varepsilon}{8M}\right)}|\mathscr{E}(h)|\geq\varepsilon,$$

which, by the union bound in Lemma 12.3, implies

$$\mathbb{P}\left[\left\{\sup_{h\in\mathscr{H}}|\mathscr{E}(h)|\geq\varepsilon\right\}\right]\leq\sum_{i=1}^{\mathscr{N}}\mathbb{P}\left[\left\{\sup_{h\in B_{h_i}\left(\frac{\varepsilon}{8M}\right)}|\mathscr{E}(h)|\geq\varepsilon\right\}\right],$$

which ends the proof by invoking Lemma 20.3. □

20.5 Sample error for finite dimensional $\mathscr{H}$

Let $W:=\operatorname{span}(\phi_1,\ldots,\phi_n)\subset\mathbb{C}(X)$ and the hypothesis space $\mathscr{H}$ be given as

$$\mathscr{H}:=\{h\in W:\|h\|_\infty\leq R\}.$$

Lemma 20.5. *Assume $\mathscr{H}$ is M-bounded in the sense of Assumption 17.1. Then the tail of the worst sampling error decays exponentially, i.e.,*

$$\mathbb{P}\left[\left\{\sup_{h\in\mathscr{H}}|\mathscr{E}(h)|\geq\varepsilon\right\}\right]\leq 2\mathscr{N}\left(\mathscr{H},\frac{\varepsilon}{8M}\right)e^{-N\frac{\varepsilon^2}{2M^4}},$$

where

$$\mathscr{N}\left(\mathscr{H},\frac{\varepsilon}{8M}\right)\leq\left(\frac{16RM}{\varepsilon}+1\right)^n.$$

Proof. The bound of the covering number of $\mathscr{H}$ using balls with radii $\varepsilon/8M$, $\mathscr{N}\left(\mathscr{H},\frac{\varepsilon}{8M}\right)$, is provided in Proposition 20.1. The assertion is readily available using Lemma 20.4. □

20.6 Sampling error when $\mathscr{H}$ is compact

In this section we assume that $\mathscr{H}$ is a compact subset of $\mathbb{C}(X)$. From Lemma 20.2 we know that the covering number $\mathscr{N}\left(\mathscr{H},\frac{\varepsilon}{8M}\right)$ for $\mathscr{H}$ is finite.[2] The sample error in Lemma 20.5 is still valid, but in this case we leave the covering number $\mathscr{N}\left(\mathscr{H},\frac{\varepsilon}{8M}\right)$ undefined/unestimated.

20.7 Sample error

We are now in the position to estimate the sample error $\mathscr{S}\left(\hat{h}_N\right)$ by bounding the two sampling errors on the right hand side of (20.1). To be concrete we consider the case when $\mathscr{H}$ is a compact subset of $\mathbb{C}(X)$, and this in fact covers the other cases except[3] the case in Section 20.3.

Since $\hat{h}$ is a deterministic function, we can use the sampling error estimation in Section 20.1 to conclude that

$$\left|\mathscr{R}_N\left(\hat{h}\right)-\mathscr{R}\left(\hat{h}\right)\right|\leq\frac{\varepsilon}{2}$$

with the probability at least

$$1-2e^{-N\frac{\varepsilon^2}{2M^4}}.$$

Nothing less will do!

Now, as remarked in Section 16.2, $\hat{h}_N$ is a random function and we have to employ the worst case error to bound $\left|\mathscr{R}\left(\hat{h}_N\right)-\mathscr{R}_N\left(\hat{h}_N\right)\right|$. Lemma 20.5 says that

$$\left|\mathscr{R}\left(\hat{h}_N\right)-\mathscr{R}_N\left(\hat{h}_N\right)\right|\leq\frac{\varepsilon}{2}$$

[2] The subject of estimating covering numbers is standard (but technical) in functional analysis, we refer the readers to [13].

[3] The reason is that balls in infinite dimensional spaces are not compact!

with the probability at least

$$1 - 2\mathscr{N}\left(\mathscr{H}, \frac{\varepsilon}{16M}\right) e^{-N\frac{\varepsilon^2}{8M^4}}.$$

Combining these results we conclude that the sample error is bounded by any ε, i.e.,

$$\mathscr{S}\left(\hat{h}_N\right) \le \varepsilon$$

with the probability at least

$$\left[1 - 2e^{-N\frac{\varepsilon^2}{2M^4}}\right] \times \left[1 - 2\mathscr{N}\left(\mathscr{H}, \frac{\varepsilon}{16M}\right) e^{-N\frac{\varepsilon^2}{8M^4}}\right]$$
$$\ge 1 - 2\left[\mathscr{N}\left(\mathscr{H}, \frac{\varepsilon}{16M}\right) + 1\right] e^{-N\frac{\varepsilon^2}{8M^4}}.$$

In summary we have proved the following result.

Theorem 20.1 (Sample error estimation). *Suppose $\mathscr{H}$ is a compact subset of $\mathbb{C}(X)$ and $\mathscr{H}$ is M-bounded. The following estimation of the sample error*

$$\mathscr{S}\left(\hat{h}_N\right) \le \varepsilon, \quad \varepsilon > 0,$$

holds true with probability at least

$$1 - 2\left[\mathscr{N}\left(\mathscr{H}, \frac{\varepsilon}{16M}\right) + 1\right] e^{-N\frac{\varepsilon^2}{8M^4}}.$$

20.8 Appendix

Definition 20.1 (ε-net). Let $\mathscr{H}$ be a metric space with a metric $\|\cdot\|$. A subset W of $\mathscr{H}$ is called a ε-net of $\mathscr{H}$ if any point in $\mathscr{H}$ is within ε-distance from a point in W, i.e.,

$$\forall h \in \mathscr{H}, \exists h_0 \in W : \|h - h_0\| \le \varepsilon.$$

Equivalently, a subset W of $\mathscr{H}$ is called a ε-net of $\mathscr{H}$ if and only if $\mathscr{H}$ can be covered by balls of radius ε with centers in W.

Definition 20.2 (Covering number). Let $\mathscr{H}$ be a metric space and $\eta > 0$. The covering number $\mathscr{N}(\mathscr{H}, \eta)$ is defined as the minimal number such that there exist $\mathscr{N}(\mathscr{H}, \eta)$ balls in $\mathscr{H}$ with radius η covering $\mathscr{H}$.

Proposition 20.1 (Coverning number for a ball in finite dimensional space). *Let n be the dimension of a Banach space W and $\mathscr{H} := \{h \in W : \|h\|_\infty \le R\} = B_0(R)$. Then for $0 < \eta < R$, we have*

$$\mathscr{N}(\mathscr{H}, \eta) \le \left(\frac{2R}{\eta} + 1\right)^n.$$

Since $\mathscr{N}(B_0(R), \eta) = \mathscr{N}(B_0(1), \eta/R)$, it is sufficient to consider unit ball in W.

Proposition 20.2 (Finite cover of compact sets). *Let $\mathscr{H}$ be a compact subset of $\mathbb{C}(X)$, then there exists a finite cover for $\mathscr{H}$.*