Sample Complexity Bounds in Agnostic Case via Growth Function

Theorem 1. Let D be a distribution on $X \times Y$. Let H be a class of hypotheses (functions mapping X to Y). Let $(x_1, y_1), ..., (x_m, y_m)$ be drawn i.i.d. from D. Assume that

$$m \;\geq\; \frac{K}{\epsilon^2}\Big(\log\big(\Pi_H(m)\big) + \log(1/\delta) \Big)$$

for some constant K. Consider the empirical risk minimization algorithm. Then, with probability at least $1-\delta$, we have $R(h_S^{ERM}) \leq \min_{h \in H} R(h) + \epsilon$.

We will show that, with probability at least $1-\delta$, all $h \in H$ have $|R(h) - \hat{R}(h)| \leq \epsilon$. Let $h^* \in \operatorname{argmin}_{h \in H} R(h)$. Then

$$R(h_S^{ERM}) \leq \hat{R}(h_S^{ERM}) + \epsilon \leq \hat{R}(h^*) + \epsilon \leq R(h^*) + 2\epsilon.$$

Replacing ϵ by $\epsilon/2$ then proves the result (at the expense of increasing m by a factor of 4).

So, consider the event

$$A \;=\; \Big\{ \exists h \in H \text{ with } |R(h) - \hat{R}(h)| > \epsilon \Big\}.$$

Our goal is to show that $\Pr[\, A \,] \leq \delta$.

Consider drawing **two** i.i.d. samples S and S', each with m labeled examples. Let $\hat{R}_S$ and $\hat{R}_{S'}$ respectively denote empirical error wrt samples S and S'. Define the event

$$B \;=\; \Big\{ \exists h \in H \text{ with } |\hat{R}_S(h) - \hat{R}_{S'}(h)| \geq \epsilon/2 \Big\}.$$

Claim 2. $\Pr[\, A \,] \leq 2 \cdot \Pr[\, B \,]$.

Proof. Clearly

$$\Pr[\, B \,] \geq \Pr[\, A \cap B \,] = \Pr[\, A \,] \cdot \Pr[\, B \mid A \,].$$

So it suffices to show that $\Pr[\, B \mid A \,] \geq 1/2$.

Suppose that A holds, and let $h \in H$ satisfy $|R(h) - \hat{R}_S(h)| > \epsilon$. We will show that, for this *same* h, with probability at least $1/2$ we have $|\hat{R}_S(h) - \hat{R}_{S'}(h)| \geq \epsilon/2$, implying that $\Pr[\, B \mid A \,] \geq 1/2$.

What is going on here? Recall that $\mathrm{E}\Big[\, \hat{R}_S(h) \,\Big] = R(h)$. So we have conditioned on the event that the random variable $\hat{R}_S(h)$ is somewhat far (i.e., distance at least ϵ) from its expectation. But in our second sample, $\hat{R}_{S'}(h)$ has good probability (i.e., at least 1/2) of being close (i.e., distance at most $\epsilon/2$) to its expectation. When that happens, we will have $|\hat{R}_S(h) - \hat{R}_{S'}(h)| \geq \epsilon/2$, as required.

To prove this, let us use the Hoeffding bound (Theorem D.1 in the textbook). (We could also use the Chernoff bound as it appeared in the Handout, but Hoeffding is slightly more convenient). We write the Hoeffding bound as follows:

Theorem 3. Let $X_1, ..., X_m$ be independent random variables with $X_i \in [a_i, b_i]$ for $i = 1, ..., m$. Let $Z = \sum_{i=1}^m X_i$. Then, for any $t > 0$, the following inequality holds:

$$\Pr[\, |Z - \mathrm{E}[\, Z \,]| \geq \epsilon \,] \;\leq\; 2 \cdot \exp\Big(-2\epsilon^2 / \sum_{i=1}^m (b_i - a_i)^2 \Big).$$

We will now use this to prove that $\Pr\Big[\, |\hat{R}_{S'}(h) - R(h)| \leq \epsilon/2 \,\Big] \geq 1/2$.

Question 3(a): *...Complete the rest of the argument...* ■

Next, we would like to show that $\Pr[\, B \,]$ is small. The idea is to use a union bound, but we don't want to union bound over all $h \in H$. Instead, we want to union bound over the number of labelings, so we can connect to the growth function. Fortunately, we can rewrite B in terms of labelings.

For any $b \in \Pi_H(S \cup S')$, let

$$\hat{R}_S(b) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{b(x_i) \neq y_i}$$
$$\hat{R}_{S'}(b) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{b(x'_i) \neq y'_i}$$

Define

$$B(b) = \left\{ |\hat{R}_S(b) - \hat{R}_{S'}(b)| \geq \epsilon/2 \right\}.$$

Then the event B that we defined above can be rewritten as

$$B = \bigcup_{b \in \Pi_H(S \cup S')} B(b) = \bigcup_{b \in Y^{2m}} \Big(\big\{ b \in \Pi_H(S \cup S') \big\} \wedge B(b) \Big)$$

Now we can apply a union bound.

$$\begin{aligned} \Pr[B] &\leq \sum_{b \in Y^{2m}} \Pr\big[b \in \Pi_H(S \cup S') \wedge B(b) \big] \\ &= \sum_{b \in Y^{2m}} \Pr\big[b \in \Pi_H(S \cup S') \big] \cdot \Pr\big[B(b) \mid b \in \Pi_H(S \cup S') \big] \end{aligned}$$

We will show that

$$\Pr\big[B(b) \mid b \in \Pi_H(S \cup S') \big] \leq \frac{\delta}{2 \cdot \Pi_H(2m)} \tag{1}$$

for all S, S' and b. It will then follow that

$$\Pr[B] \leq \frac{\delta}{2 \cdot \Pi_H(2m)} \sum_{b \in Y^{2m}} \Pr\big[b \in \Pi_H(S \cup S') \big] = \frac{\delta}{2 \cdot \Pi_H(2m)} \mathrm{E}\big[|\Pi_H(S \cup S')| \big] \leq \delta/2.$$

So let us turn to the proof of (1). We may assume that $b \in \Pi_H(S \cup S')$.

Let $(x_1, y_1), ..., (x_d, y_d)$ be the points in $S \cup S'$ for which $b(x_d) \neq y_d$. For $i = 1, ..., d$, let

$$Z_i = \begin{cases} 1 & \text{if } (x_i, y_i) \in S \\ -1 & \text{if } (x_i, y_i) \in S' \end{cases}$$

Then $\Pr[Z_i = 1] = 1/2$ and $\mathrm{E}[Z_i] = 0$.

Note that $\hat{R}_S(b) - \hat{R}_{S'}(b) = \sum_{i=1}^{d} Z_i/m$. So our goal is to show that

$$\Pr\Big[|\textstyle\sum_{i=1}^{d} Z_i| \geq m\epsilon \Big] \leq \frac{\delta}{2 \cdot \Pi_H(2m)}.$$

Thus, it seems that we are in a scenario where the Hoeffding bound is applicable.

But there is **a small catch!** The catch is that $Z_1, ..., Z_d$ are not independent. This is because it is known that $|S| = |S'| = m$. So, whenever $Z_1 = 1$ this slightly decreases the probability that $Z_2 = 1$, etc. There are several ways to work around this catch.

Bonus question: *By digging through the literature, or coming up with our yown ideas, figure out a way around this catch.*

Now we assume that we have figured out a way around the catch, and use the Hoeffding bound to prove (1).

Question 3(b): *...Complete the rest of the argument...*