COLT— Exercise Sheet 6

Exercise 1. Let $\mathcal{C}$ be a class of hypothesis from X to $\{0,1\}$. Show that if $\mathcal{C}$ is efficiently agnostic PAC-learnable, then it is efficiently PAC-learnable.

Exercise 2. Hoeffding's inequality states that if $X_1, \ldots, X_m$ are independent random variables with $a_i \leq X_i \leq b_i$ almost surely, then for every $\epsilon > 0$, for $S_m = \sum_{i=1}^{m} X_m$ and $\Delta^2 = \sum_{i=1}^{m}(b_i - a_i)^2$

we have

$$\mathbb{P}(S_m - \mathbb{E}(S_m) \geq \epsilon) \leq e^{\frac{2\epsilon^2}{\Delta^2}}$$
$$\mathbb{P}(S_m - \mathbb{E}(S_m) \leq -\epsilon) \leq e^{\frac{2\epsilon^2}{\Delta^2}}$$

Consider $\mathcal{H}$ being classifiers with the 0-1 loss. Use Hoeffding's inequality to estimate:

- The function $\mathbf{m}^R(\epsilon, \delta) = min\{m \mid \mathbb{P}_{S\sim D^m}(\textsc{Rep}_D(\mathcal{H}, S) \leq \epsilon) \geq 1 - \delta\}$ for $|\mathcal{H}| = 1$.
- The function $\mathbf{m}^R(\epsilon, \delta)$ for $|\mathcal{H}|$ finite.

Derive the estimate on the function $\epsilon^R(m, \delta) = min\{\epsilon \mid \mathbf{m}^R(\epsilon, \delta) \leq m\}$.

Exercise 3. Generalize the above assignment to other loss functions. What assumptions do you need for that to work?

Exercise 4. Let $D_1, \ldots, D_m$ be a sequence of distributions over an example space X. Let $\mathcal{C}$ be a finite hypothesis class over X. Fix $\mathbf{c} \in \mathcal{C}$. We consider samples S consisting of m examples, where each x_i is drawn with the distribution D_i (the examples are not independent), and then labeled with $\mathbf{c}$.

Now, for the evaluation we consider the average distribution:

$$D = \frac{D_1 + \ldots + D_m}{m}.$$

That is we consider the generalisation error w.r.t. D:

$$\mathcal{R}_D(\mathbf{h}) = \mathbb{P}_{x\sim D}(\mathbf{c}(x) \neq \mathbf{h}(x)).$$

Show that for every $\epsilon > 0$, we have

$$\mathbb{P}_{S\sim(D_1,\ldots,D_m)}\Big(\exists \mathbf{h} \in \mathcal{C}(\mathcal{R}_D(\mathbf{h}) \geq \epsilon \wedge \widehat{\mathcal{R}}_S(\mathbf{h}) = 0)\Big) \leq |\mathcal{C}| \cdot e^{-\epsilon \cdot m}.$$