Solution to Homework 2
CS-526 Learning Theory

---

### Exercise 4.1

<u>1 ⇒ 2</u>: Assume for every $\epsilon, \delta > 0$ there exists $m(\epsilon, \delta)$ such that $\forall m \geq m(\epsilon, \delta)$

$$\mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) > \epsilon) < \delta. \tag{1}$$

Then using the definition of expectation

$$\begin{aligned}\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] &\leq \mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) > \epsilon)\cdot 1 + \mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) \leq \epsilon)\cdot \epsilon \\ &\leq \mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) > \epsilon) + \epsilon \\ &\leq \delta + \epsilon,\end{aligned}$$

where the last inequality follows from the assumption (1). Now set $\delta = \epsilon$. We have for every $\epsilon > 0$ there exists $m(\epsilon, \epsilon)$ such that $\forall m \geq m(\epsilon, \epsilon)$

$$\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] \leq 2\epsilon. \tag{2}$$

So it is valid to pass both sides of (2) to the limit $\lim_{m\to\infty}\lim_{\epsilon\to 0}$, which gives

$$\lim_{m\to\infty}\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] \leq 0.$$

Also by definition $\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] \geq 0$. Thus we conclude $\lim_{m\to\infty}\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] = 0$.
<u>2 ⇒ 1</u>: Assume that $\lim_{m\to\infty}\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] = 0$. For every $\epsilon, \delta \in (0,1)$ there exists some $m_0 \in \mathbb{N}$ such that for every $m \geq m_0$, $\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))] \leq \epsilon\delta$. By Markov's inequality,

$$\begin{aligned}\mathbb{P}_{S\sim\mathcal{D}^m}(L_{\mathcal{D}}(A(S)) > \epsilon) &\leq \frac{\mathbb{E}_{S\sim\mathcal{D}^m}[L_{\mathcal{D}}(A(S))]}{\epsilon} \\ &\leq \frac{\epsilon\delta}{\epsilon} \\ &= \delta.\end{aligned}$$

### Exercise 4.2

Applying Hoeffding's inequality to $L_S(h) = \frac{1}{m}\sum_{i=1}^{m}\ell\big(h, (x_i, y_i)\big)$ yields:

$$\mathbb{P}_{S\sim\mathcal{D}^m}\Big(|L_S(h) - \mathbb{E}\,L_S(h)| > \epsilon\Big) = \mathbb{P}_{S\sim\mathcal{D}^m}\Big(|L_S(h) - L_{\mathcal{D}}(h)| > \epsilon\Big) \leq 2\exp\left(-\frac{2m\epsilon^2}{(b-a)^2}\right).$$

We then use this upper bound in the step where we use the union bound to obtain:

$$\begin{aligned}\mathbb{P}_{S\sim\mathcal{D}^m}(\exists h\in\mathcal{H}:|L_S(h)-L_D(h)|>\epsilon) &\leq \sum_{h\in\mathcal{H}}\mathbb{P}_{S\sim\mathcal{D}^m}(|L_{\mathcal{D}}(h)-L_S(h)|>\epsilon)\\ &\leq 2|\mathcal{H}|\exp\left(-\frac{2m\epsilon^2}{(b-a)^2}\right).\end{aligned}$$

The desired bound on the sample complexity follows from requiring $2|\mathcal{H}|\exp\left(-\frac{2m\epsilon^2}{(b-a)^2}\right)\leq\delta$.

---

Solution to ExtraHomework on Hoeffding inequality
CS-526 Learning Theory

---

1. A function $f$ which is convex on an interval $I\subseteq\mathbb{R}$ satisfies $\forall(a,b)\in I^2,\forall\alpha\in[0,1]:$ $f(\alpha\, a+(1-\alpha)\, b)\leq\alpha f(a)+(1-\alpha)f(b)$. Substituting $f(x)=e^{\lambda x}$ and $\alpha=\frac{b-X}{b-a}\in[0,1]$ into this inequality, we get:

   $$e^{\lambda X}\leq\frac{b-X}{b-a}e^{\lambda a}+\frac{X-a}{b-a}e^{\lambda b}\,.$$

   Taking the expectation on both sides and using $\mathbb{E}[X]=0$, we have

   $$\mathbb{E}[e^{\lambda X}]\leq\frac{b}{b-a}e^{\lambda a}-\frac{a}{b-a}e^{\lambda b}\,.$$

2. With $p=-a/(b-a)$ and $h=\lambda(b-a)$, we have

   $$\begin{aligned}\log\left(\frac{b}{b-a}e^{\lambda a}-\frac{a}{b-a}e^{\lambda b}\right)&=\log(e^{\lambda a})+\log\left(\frac{b}{b-a}-\frac{a}{b-a}e^{\lambda(b-a)}\right)\\&=\lambda a+\log\left(1+\frac{a}{b-a}-\frac{a}{b-a}e^{\lambda(b-a)}\right)\\&=-hp+\log\left(1-p+pe^h\right).\end{aligned}$$

3. Let $\theta(h)=\frac{pe^h}{1-p+pe^h}$. We can compute:

   $$L'(h)=-p+\theta(h)\quad,\quad L''(h)=\theta(h)\big(1-\theta(h)\big)=-\left(\theta(h)-\frac{1}{2}\right)^2+\frac{1}{4}\leq\frac{1}{4}\,.$$

   We can also verify that $L(0)=L'(0)=0$. Plugging these computations back in the equation $L(h)=L(0)+hL'(0)+(h^2/2)L''(\xi)$ yields $L(h)\leq h^2/8$. Combining this upper bound with the previous step gives:

   $$\mathbb{E}[e^{\lambda X}]\leq e^{L(\lambda(b-a))}\leq\exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

4. Let $X_i = Z_i - \mathbb{E}Z_i$ and $\overline{X} = \frac{1}{m}\sum_{i=1}^m X_i$. First using the monotonicity of the exponent function and then Markov's inequality, we have:

$$\mathbb{P}\big(\overline{X} \geq \epsilon\big) = \mathbb{P}\big(e^{\lambda\overline{X}} \geq e^{\lambda\epsilon}\big) \leq e^{-\lambda\epsilon}\ \mathbb{E}\big[e^{\lambda\overline{X}}\big] .$$

As $X_1, \ldots, X_m$ are independent we have $\mathbb{E}[e^{\lambda\overline{X}}] = \prod_{i=1}^m \mathbb{E}[e^{\frac{\lambda X_i}{m}}]$. We have shown in the previous step that $\forall i \in \{1, \ldots, m\} : \mathbb{E}[e^{\lambda X_i/m}] \leq e^{\lambda^2(b-a)^2/(8m^2)}$. We conclude that:

$$\mathbb{P}\big(\overline{X} \geq \epsilon\big) \leq \exp\left( -\lambda\epsilon + \frac{\lambda^2(b-a)^2}{8m}\right) .$$

5. The inequality is obtained by optimizing over $\lambda$ the upper bound of step 4. The exponent $-\lambda\epsilon + \frac{\lambda^2(b-a)^2}{8m}$ is a quadratic (convex) function of $\lambda$. It is minimized when $\lambda = 4m\epsilon/(b-a)^2$. Choosing $\lambda$ this way gives the desired bound, i.e., *Hoeffding's inequality.*