Foundations of ML & AI

Theodoros Evgeniou - Nicolas Vayatis

Exercise Set No 1

Exercice 1 (PAC learnability in the zero error case)

Consider \mathcal{H} a finite hypothesis space of decision functions $h: \mathcal{X} \to \mathcal{Y}$. Assume that the optimal element h^* in the sense of the error L(h) is an element of \mathcal{H} and that for any sample D_n , the ERM \hat{h}_n satisfies $\hat{L}_n(\hat{h}_n) = 0$.

1. Prove an upper bound of the following probability: for any $\varepsilon > 0$

$$\mathbb{P}\{\widehat{L}_n(\widehat{h}_n) = 0 \text{ and } L(\widehat{h}_n) > \varepsilon\}$$

which will depend on $K = |\mathcal{H}|$ the cardinality of \mathcal{H} , n the sample size and ε .

Hint: Apply a union bound and then use the definition of conditional probability to upper bound $\mathbb{P}\{\widehat{L}_n(h) = 0 \text{ and } L(h) > \varepsilon\}$ by $\mathbb{P}\{\widehat{L}_n(h) = 0 \mid L(h) > \varepsilon\}$ for any $h \in \mathcal{H}$.

2. Show that \widehat{f}_n will be ε -close to h^* with probability $1-\delta$ for a sample size of

$$n(\varepsilon, \delta) = \frac{\log K - \log \delta}{\varepsilon}$$
.

Exercice 2 (Hoeffing's lemma)

1. Consider Z a random variable such that : $\mathbb{E}(Z) = 0$ and $\mathbb{P}(Z \in [a, b]) = 1$ almost surely. Prove the following upper bound : for any s > 0,

$$\mathbb{E}(e^{sZ}) \le \exp(s^2(b-a)^2/8) .$$

Hint: Use the convexity of the exponential function and the order 2 Taylor expansion of the function

$$\varphi(s) = \log\left(\frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb}\right).$$

2. Consider Z_1, \ldots, Z_n IID over [a, b] and $\overline{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$. Show that we have, for any t > 0,

$$\mathbb{P}\{\overline{Z}_n - \mathbb{E}(Z_1) > t\} \le \exp(-2nt^2/(b-a)^2)$$

and

$$\mathbb{P}\{\overline{Z}_n - \mathbb{E}(Z_1) < -t\} \le \exp(-2nt^2/(b-a)^2)$$

Hint: Use Chernoff's bounding method

$$\mathbb{P}(Z > t) \le \inf_{s>0} \left\{ e^{-st} \mathbb{E}\left(e^{sZ}\right) \right\} .$$