Exercises

Machine Learning: Foundations and Applications — Paul J. Atzberger
MATH 260 — http://atzberger.org/

Can choose to complete 3 out the following 6 problems.

1. Show that the concept class of hyper-rectangles $[a_1, b_1] \times [a_2, b_2] \times [a_n, b_n] \in \mathbb{R}^n$ in PAC learnable. Hint: Start by considering $n = 2$ and showing this is learnable and then work from there.

2. Consider the concept class of concentric circles of the form $\{(x, y)|x^2 + y^2 \leq r^2\}$ for $r > 0, r \in \mathbb{R}$. Show this is (ϵ, δ)-PAC-learnable from a training data set of size $m \geq (1/\epsilon)\log(1/\delta)$.

3. Consider a data collection (x, y) with $x \in \mathcal{X}$ a discrete set and $y \in \{0, 1\}$ with distribution $\mathcal{D}_{x,y}$ when for x the assigned labels y are stochastic (non-deterministic). Consider the Bayes Classifier given by $h_{\text{bayes}}(x) = \text{argmax}_y \Pr[y|x]$. Show the Generalization Error $R(\cdot)$ has the property that $R^* = R(h_{\text{bayes}}) \leq R(h)$ for any hypothesis.

4. Show that the strategy of *Empirical Risk Minimization* (ERM) for a discrete hypothesis space $\mathcal{H}$ satisfies the bound

$$R(h^{ERM}) - R(h^*) \leq 2 \sup_{h \in \mathcal{H}} |R(h) - \hat{R}(h)|. \tag{1}$$

   Here, $h^{ERM} = \text{argmin}_h \hat{R}(h)$ and $h^* = \text{argmin}_h R(h)$. State why this result is potentially useful.

5. Consider a collection of i.i.d. random variables $X_1, X_2, \ldots, X_N$ with finite mean μ and variance σ^2. From the Law of Large Numbers we know that $S_N = \sum_{k=1}^{n} X_k$ will have $\mu_N = \frac{1}{N} S_N \to \mu$. Using the Cheyshev inequality derive an estimate for at least how many samples N are required so that for the estimate of the mean $\mu_N = \frac{1}{N} S_N$ satisfies for a given $\epsilon > 0$ and $\delta > 0$

$$\Pr\{|\mu_N - \mu| > \epsilon\} \leq \delta. \tag{2}$$

   In the case we make the special choice of $\epsilon = m\sigma$ state the lower bound on N.

6. Suppose we observe a collection of flips of a coin with the unknown probability of heads p^*. We make an estimate $\tilde{p} = \mu_N$ of this probability where $X = 1$ for heads and $X = 0$ for tails. To be confident that our estimate $\tilde{p}$ is within 0.1 of the true probability p^* would occur 99% of the time we use this approach, how many samples N do we need to make sure to observe?