Problem 1. Bayes predictor.

The Bayes predictor $f_\star : \mathcal{X} \to \mathcal{Y}$ is defined as:

$$f_\star(x') \in \arg\min_{z \in \mathcal{Y}} \mathbb{E}[\ell(y, z)|x = x'].$$

1. Consider binary classification with $\mathcal{Y} = \{-1, 1\}$ with the loss function

$$\ell(y, z) = \begin{cases} 0 & \text{if } y = z, \\ c_{\text{FP}} > 0 & \text{if } y = -1, z = 1, \\ c_{\text{FN}} > 0 & \text{if } y = 1, z = -1. \end{cases}$$

   Compute a Bayes predictor at $x = x'$ as a function of $\mathbb{E}[y|x = x']$.
   Hint: Show that $f_\star(x') = \mathbf{1}\{\mathbb{E}[y|x = x'] \geq$ *or* $\leq$ *???*$\}$

2. Show that every Bayes predictor minimizes the population risk; that is, $\mathcal{R}(f_\star) \leq \mathcal{R}(g)$ for any $g : \mathcal{X} \to \mathcal{Y}$.

3. *(Bonus)* We consider a learning problem on $\mathcal{X} \times \mathcal{Y}$, with $\mathcal{Y} = \mathbb{R}$ and the absolute loss defined as $\ell(y, z) = |y - z|$. Characterize the Bayes predictors in terms of the conditional distribution of $y|x = x'$, and give one explicit choice using the conditional cumulative distribution function (CDF).

Solution.

Problem 2. Empirical risk minimization and PAC learning.

Let $\mathcal{X}$ be a discrete (finite or infinite) domain, $\mathcal{Y} = \{0, 1\}$, and we use the 0–1 loss. The training sample consists of m i.i.d. samples drawn from a distribution *realizable* by $\mathcal{H} = \{h_z : z \in \mathcal{X}\} \cup \{h_-\}$, where for each $z \in \mathcal{X}$, h_z is the indicator function for z:

$$h_z(x) = \begin{cases} 1 & \text{if } x = z, \\ 0 & \text{otherwise.} \end{cases}$$

h_- is the all-zero function, namely, $\forall x \in \mathcal{X}$, $h_-(x) = 0$.

1. Describe an algorithm that implements the ERM rule for learning $\mathcal{H}$.
2. Show that $\mathcal{H}$ is PAC learnable. Provide an upper bound on the sample complexity.

Solution.

Problem 3. Probability and concentration.

Suppose that we have m coins. Each coin has the same probability of landing on heads, denoted p. Suppose we pick up each of the m coins in turn and, for each coin, perform n tosses. We assume that those mn tosses are mutually independent. Note that the probability of obtaining exactly k heads out of n tosses for any given coin is given by the binomial distribution:

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Define

$$\hat{p}_i := \frac{\text{number of times coin } i \text{ lands on heads}}{n}.$$

1. Compute a formula for the exact probability that at least one coin will have $\hat{p}_i = 0$.

2. For $\epsilon > 0$, apply Hoeffding's inequality[1] and a union bound to prove an upper bound on

$$\Pr\left[\max_{1 \le i \le m} |\hat{p}_i - p| > \epsilon\right]$$

Solution.

[1] https://en.wikipedia.org/wiki/Hoeffding%27s_inequality