# Problem 3

**Due on November 16th**

Consider the RANDOMIZED HYPOTHESIS ELIMINATION algorithm defined for a finite hypothesis set $H$ as follows:

---

**Algorithm 1** RANDOMIZED HYPOTHESIS ELIMINATION

---

**Initialization:** Let $V_1 = H$.
**For all** $t = 1, 2, \ldots, n$, **repeat**

- Choose an $h_t \in V_t$ uniformly at random.
- Observe $\mathbf{x}_t$.
- Predict $\widehat{y}_t = h_t(\mathbf{x}_t)$.
- Update $V_{t+1} = \{h \in V_t : h(\mathbf{x}_t) = y_t\}$.

---

The following theorem establishes a performance guarantee for this algorithm:

**Theorem 3.** *Under the realizability assumption, the total expected loss of* RANDOMIZED HYPOTHESIS ELIMINATION *over $n$ rounds satisfies*

$$\mathbb{E}\left[\sum_{t=1}^{n} \ell(\mathbf{x}_t, y_t, \widehat{y}_t)\right] \leq 1 + \ln |H|.$$

Your homework is to answer the following questions about RANDOMIZED HYPOTHESIS ELIMINATION and the other algorithms we discussed in class:

**Question 1:** When is the bound of Theorem 3 better than the bound of Theorem 2 presented in the lecture? (What is the size of the smallest hypothesis class for which the former is better than the latter?)

**Question 2:** What is the computational complexity of RANDOMIZED HYPOTHESIS ELIMINATION? (Hints: How expensive is it to generate a random hypothesis and test if it is in the version space? How many random hypotheses do we need to generate until we find a consistent one?)

**Question 3:** Construct a hypothesis class and a sequence of examples for which HYPOTHESIS ELIMINATION (the non-randomized one) makes exactly $|H| - 1$ mistakes. (Hint: a possible good idea is choosing the $\mathbf{x}_t$'s from the natural numbers $1, 2, \ldots, |H|$ and consider the hypothesis class $H$ consisting of functions of the form $h_k(x) = \mathbb{I}\{x \leq k\}$ (i.e., $h_k(x) = 1$ if $x \leq k$ and $h_k(x) = 0$ otherwise) for all $k = 1, 2, \ldots, |H|$.)

**Question 4:** Construct a hypothesis class and a sequence of examples for which HALVING makes exactly $\log_2 |H|$ mistakes. (Hint: The same hypothesis class as above and the same pool of $\mathbf{x}_t$'s may work here too.)

**Question 5*:** Prove Theorem 3! (See hints in the slides and below.)

## Hints for Question 5

It is enough to consider a sequence $(\mathbf{x}_t, y_t)_{t=1}^n$ such that in every round $t$, there is exactly one hypothesis $g_t \in H$ making a mistake. (For the curious reader, this follows from the argument used for proving Corollary 4.5 in Cesa-Bianchi and Lugosi, 2006.) At any round $t$, what is the maximal probability of the algorithm choosing the hypothesis $t$? Hint: choosing a hypothesis $h_t$ uniformly at random from a set $V$ means that each hypothesis is $V$ is chosen with equal probability $\frac{1}{|V|}$.

# References


Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games.* Cambridge University Press, New York, NY, USA.