Machine Learning Theory (CSC 482A/581B)

Problem Set 4 **Due on Friday, March 29th, 7pm**

Instructions:

- You must write up your solutions individually.
- You may have high-level discussions with 1 other student registered in the course. If you discuss problems with another student, include at the top of your submission: their name, V#, and the problems discussed.
- Please do not search for solutions online. Instead, ask the instructor for hints if you are stuck.
- Your must type up your solutions and are encouraged to use LaTeX to do this. For any problems where you only have a partial solution, be clear about any parts of your solution for which you have low confidence.
- Please submit your solutions via conneX by the due date of Friday, March 29th, 7pm. This is a hard deadline.

Questions:

1. Using a somewhat different proof than we saw in class, it is possible to obtain the following PAC-Bayesian bound for a finite set of hypotheses $\mathcal{F}$ and a training sample of n labeled examples drawn from a distribution P over $\mathcal{X} \times \mathcal{Y}$.

   With probability at least $1 - \delta$, for all distributions $\hat{\Pi}$ over $\mathcal{F}$,

$$\begin{aligned} &\mathsf{E}_{f \sim \hat{\Pi}} \left[\mathsf{E}_{(X,Y) \sim P} \left[\mathbf{1} \left[f(X) \neq Y \right] \right] \right] \\ &\quad \leq 2 \left(\mathsf{E}_{f \sim \hat{\Pi}} \left[\frac{1}{n} \sum_{j=1}^{n} \mathbf{1} \left[f(X_j) \neq Y_j \right] \right] + \frac{\mathrm{D}_{\mathrm{KL}}(\hat{\Pi} \,\|\, \Pi) + \log \frac{1}{\delta}}{n} \right). \end{aligned} \tag{1}$$

   Take the prior distribution Π to be the uniform distribution over $\mathcal{F}$. Suppose that we are in a "lucky" situation where, for the particular training sample, there is a set $\hat{\mathcal{F}}_0 \subseteq \mathcal{F}$ (of cardinality at least 1) for which

$$\frac{1}{n} \sum_{j=1}^{n} \mathbf{1} \left[f(X_j) \neq Y_j \right] = 0 \quad \text{for all } f \in \hat{\mathcal{F}}_0.$$

   In this lucky situation, show that the right-hand side of the bound (1) can be equal to

$$\frac{2}{n} \left(\log \frac{|\mathcal{F}|}{|\hat{\mathcal{F}}_0|} + \log \frac{1}{\delta} \right).$$

   In particular, provide the form of the posterior distribution $\hat{\Pi}$ that realizes this bound.

2. In this question, we explore a modified form of a regret bound for decision-theoretic online learning, called a *quantile bound*. The idea of an ε-quantile bound, for $\varepsilon \in [1/K, 1]$, is to ensure that the cumulative loss of the learning algorithm is not much greater than the cumulative loss of the $\lceil \varepsilon K \rceil^{\text{th}}$ best expert. To describe this bound formally, let $J(L_T, \varepsilon)$ be the $\lceil \varepsilon K \rceil^{\text{th}}$ best expert with respect to the cumulative loss vector $L_T = (L_{1,T}, \ldots, L_{K,T})$, where we define $L_{j,T} = \sum_{t=1}^{T} \ell_{j,t}$ for each $j \in [K]$. For example, if expert 5 is the second-best expert for data L_T and if $\varepsilon = \frac{2}{K}$, then we have $J(L_t, \varepsilon) = 5$.

   Formally, for an ε-quantile bound, the goal is to obtain, for all sequences of loss vectors $\ell_1, \ldots, \ell_T$, an upper bound of the form

$$\sum_{t=1}^{T} p_t \cdot \ell_t - \sum_{t=1}^{T} \ell_{J(L_T,\varepsilon),t} \qquad .$$

   Suppose that we want an ε-quantile bound for a specified value of ε. For a given number of experts K, fraction ε, and number of rounds T, show that if Hedge is run with an appropriate choice of learning rate, then, for all sequences of loss vectors $\ell_1, \ldots, \ell_T$, the cumulative loss of Hedge satisfies

$$\sum_{t=1}^{T} p_t \cdot \ell_t \leq \sum_{t=1}^{T} \ell_{J(L_T,\varepsilon),t} + \sqrt{\frac{T \log \frac{1}{\varepsilon}}{2}} \qquad .$$

3. Bonus question: Solve Problem 2.10 in the "Prediction, Learning, and Games" book.[1] Note that this question relies on using Theorem 2.4 in that book, and it is helpful to also take a look at Corollary 2.4 and its proof. A correct answer to the bonus question will be worth at least as much as either one of the previous two questions.

[1] Please contact the instructor if you need help getting access to this book.