HMC CS 158, Fall 2017
Problem Set 10 Exercises: Learning Theory

Goals:

- To practice applying probability inequalities in the context of learning theory.
- To practice characterizing the VC-dimension for various hypothesis spaces.

1 Learning Theory [10 pts]

You are hired by CNN to help design the sampling procedure for making their electoral predictions for the next presidential election in the (fictitious) country of Elbania. The country of Elbania is organized into states, and there are only two candidates running in this election: One from the Elbanian Democratic party, and another from the Labor Party of Elbania. The plan for making our electorial predictions is as follows: We will sample m voters from each state and ask whether they are voting Democrat. We will then publish, for each state, the estimated fraction of Democrat voters. In this problem, we will work out how many voters we need to sample in order to ensure that we get good predictions with high probability.

Specifically, we will say that our prediction for a state is "highly inaccurate" if the estimated fraction of Democrat voters differs from the actual fraction of Democrat voters within that state by more than a tolerance factor γ. CNN knows that their viewers will tolerate some small number of states estimates being highly inaccurate; however, their credibility would be damaged if they reported highly inaccurate estimates for too many states. So, rather than trying to ensure that all states estimates are within γ of the true values (which would correspond to no state's estimate being highly inaccurate), we will instead try only to ensure that the number of states with highly inaccurate estimates is small.

To formalize the problem, let there be n states, and let m voters be drawn IID from each state. Let the actual fraction of voters in state i that voted Democrat be ϕ_i. Also let X_{ij} ($1 \leq i \leq n$, $1 \leq j \leq m$) be a binary random variable indicating whether the j^{th} randomly chosen voter from state i voted Democrat:

$$X_{ij} = \begin{cases} 1, & \text{if the } j^{\text{th}} \text{ example from the } i^{\text{th}} \text{ state voted Democrat} \\ 0, & \text{otherwise} \end{cases}$$

We assume that the voters correctly disclose their vote during the survey. Thus, for each value of i, we have that X_{ij} are drawn IID from a $Bernoulli(\phi_i)$ distribution. Moreover, the X_{ij}'s (for all i, j) are all mutually independent.

After the survey, the fraction of Democrat votes in state i is estimated as

$$\hat{\phi}_i = \frac{1}{m} \sum_{j=1}^{m} X_{ij}.$$

Parts of this assignment are adapted from course material by Andrew Ng (Stanford) and Andrea Danyluk (Williams).

Also let $Z_i = \mathbb{I}\left[\left[|\hat{\phi}_i - \phi_i| > \gamma\right]\right]$ be a binary random variable that indicates whether the prediction in state i was highly inaccurate. The fraction of states on which our predictions are highly inaccurate is given by $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. We will prove a upper bound on the probability $P(\overline{Z} \geq \tau)$ of being highly inaccurate on at least a fraction τ of the states.

Throughout this problem, make sure to justify your steps to receive full credit.

(a) **(2 pts)** Hoeffding's inequality is perhaps the most important inequality in learning theory. It allows us to bound the probability that sums of bounded random variables[1] are too large or too small.

> **Theorem** (Hoeffding's inequality)
> Let $X_1, \ldots, X_n$ be n independent and identically distributed (iid) random variables drawn from a $Bernoulli(\phi)$ distribution, i.e., $P(X_i = 1) = \phi$, and $P(X_i = 0) = 1 - \phi$. Let $\hat{\phi} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the mean of these random variables, and let any $\gamma > 0$ be fixed. Then $P(|\phi - \hat{\phi}| > \gamma) \leq 2\exp(-2\gamma^2 n)$.

Let ψ_i be the probability that $Z_i = 1$. Using Hoeffding's inequality, find an upper bound on ψ_i.

(b) **(4 pts)** One of the most basic of all probability inequalities is known as Markov's inequality.

> **Theorem** (Markov's inequality)
> For any nonnegative random variable X and $\tau > 0$, then $P(X \geq \tau) \leq \frac{\mathbb{E}[X]}{\tau}$.

Prove Markov's inequality.

(c) **(4 pts)** Use your results above to prove a reasonable closed-form upper bound on $P(\overline{Z} \geq \tau)$.

Hint: You should find that for fixed n and $\tau > 0$, the bound goes to zero as $m \to \infty$. For fixed m and $\tau > 0$, the bound stays constant as $n \to \infty$. That is, the bound decreases as we sample more voters per state, and the bound remains constant as we sample more states.

2 VC-Dimension [8 pts]

(a) **(4 pts)** (Adapted from Mitchell 7.5) Consider the space of instances X corresponding to all points in the x, y plane. What is the VC-dimension of the hypothesis space defined by H_c = circles in the x, y plane, with points inside the circle are classified as positive examples? Justify your answer (e.g. with one or more diagrams).

(b) **(4 pts)** This problem investigates a few properties of the VC dimension, mostly relating to how $VC(H)$ increases as the set H increases. For each part of this problem, you should state whether the given statement is true, and justify your answer with either a formal proof or a counter-example.

  i. **(2 pts)** Let two hypothesis classes H_1 and H_2 satisfy $H_1 \subseteq H_2$. Prove or disprove: $VC(H_1) \leq VC(H_2)$.

  ii. **(2 pts)** Let $H_1 = H_2 \cup H_3$. Prove or disprove: $VC(H_1) \leq VC(H_2) + VC(H_3)$.

[1] Hoeffding's inequality only requires that Z_i be bounded by the interval $[a, b]$, but we state the specific case for $Z_i \sim Bernouilli(\phi)$, $a = 0$, $b = 1$ here.