A. Probability review

Let \( h: X \rightarrow \{0, 1\} \) be a hypothesis and let \( S \) denote an i.i.d. sample of size \( m \). For any \( \epsilon > 0 \), the following two-sided inequality holds:

\[
\Pr_S(\left| \hat{R}_S(h) - R(h) \right| > \epsilon) \leq 2e^{-2m\epsilon^2}.
\]

Show that the variance of \( \hat{R}_S(h) \) satisfies \( \text{var}[\hat{R}_S(h)] \leq \frac{\log(2e)}{2m} \). (Hint: use the identity \( E[X^2] = \int_0^\infty \Pr[X^2 > t]dt \).)

**Solution:** For any \( u > 0 \), we have

\[
\text{var}[\hat{R}(h)] = E[(\hat{R}(h) - R(h))^2]
= \int_0^u \Pr[(\hat{R}(h) - R(h))^2 > t]dt + \int_u^\infty \Pr[(\hat{R}(h) - R(h))^2 > t]dt
\leq u + \int_u^\infty 2e^{-2mt}dt
= u + \frac{e^{-2mu}}{m} := f(u).
\]

The function \( f(u) \) is a convex function of \( u \), with its minimum value attained when

\[
f'(u_0) = 0 \iff 1 - 2e^{-2mu_0} = 0 \iff u_0 = \frac{\log 2}{2m}.
\]

Plug in \( u_0 \) to get \( f(u_0) = \frac{\log(2e)}{2m} \). \( \square \)

B. PAC learning

1. Consider the concept class \( C \) formed by threshold functions on the real line, \( C = \{[c, \infty): \forall c \in \mathbb{R}\} \cup \{(-\infty, c]: \forall c \in \mathbb{R}\} \). Give a PAC-learning algorithm for \( C \). The analysis is similar to that of the axis-aligned
rectangles given in class, but you should carefully present and justify your proof.

**Solution:** Let $S = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ denote the labeled sample of size $m$. Without loss of generality, assume that the true concept is $[c, \infty)$ for some unknown $c \in \mathbb{R}$. Define

\[
\hat{l} = \max \{x_i : (x_i, y_i) \in S, y_i = -1\}, \\
\hat{r} = \min \{x_i : (x_i, y_i) \in S, y_i = 1\}.
\]

By definition, $\hat{l} \leq c \leq \hat{r}$.

The algorithm returns the concept $R_S = [\hat{c}, \infty)$ with $\hat{c} = (\hat{l} + \hat{r})/2$. The error region of $R_S$ is the interval $[\hat{c}, c)$ when $\hat{c} < c$, and $[c, \hat{c})$ otherwise. In both cases, the error region is a subset of $(\hat{l}, \hat{r})$. Therefore,

\[
\Pr[R(R_S) > \epsilon] \leq \Pr[R((\hat{l}, \hat{r})) > \epsilon] \leq (1 - \epsilon)^m \leq e^{-\epsilon m}.
\]

Setting $\delta$ to be greater than or equal to the right-hand side leads to $m \geq \frac{1}{\epsilon} \log(\frac{1}{\delta})$.

2. Give a PAC-learning algorithm for the concept class $C_2$ on $\mathbb{R}^2$ that is formed by intersections of axis-aligned half-spaces: $C_2$ consists of concepts of the following forms:

\[
\{(x, y) : x \geq c_x, y \geq c_y\}, \\
\{(x, y) : x \geq c_x, y \leq c_y\}, \\
\{(x, y) : x \leq c_x, y \geq c_y\}, \\
\{(x, y) : x \leq c_x, y \leq c_y\},
\]

where $c_x, c_y \in \mathbb{R}$. You should carefully justify all steps of your proof.

**Solution:** The algorithm returns the tightest concept in $C_2$ containing points labeled with 1. The proof is similar to the proof of learning axis-aligned rectangles.