HW1

September 20, 2016

A. Probability tools

1. Let $\delta = f(t)$, then $t = f^{-1}(\delta)$.
   $\Pr[X > t] \leq f(t) \Rightarrow \Pr[X > f^{-1}(\delta)] \leq \delta \Rightarrow \Pr[X \leq f^{-1}(\delta)] > 1 - \delta$.

2.

$$
\begin{aligned}
E[X] &= \sum_{n \geq 1} n \Pr[X = n] && \text{(Definition)} \\
&= \sum_{n \geq 1} n (\Pr[X \geq n] - \Pr[X \geq n+1]) && \text{(X is non-negative integer)} \\
&= (\Pr[X \geq 1] - \Pr[X \geq 2]) + 2(\Pr[X \geq 2] - \Pr[X \geq 3]) + \\
&\cdots + n(\Pr[X \geq n] - \Pr[X \geq n+1]) + \cdots \\
&= \Pr[X \geq 1] + \Pr[X \geq 2] + \cdots + \Pr[X \geq n+1] + \cdots \\
&= \sum_{n \geq 1} \Pr[X \geq n]
\end{aligned}
$$

B. Label bias

1. $\hat{p_+} = \frac{1}{m} \sum_{x \sim S} \mathbf{1}_{f(x)=+1}$

   Because f is a labeling function, and p_+ is a probability function, irrelevant to the value of sample, if we change the labeling of f to $\{0, 1\}$, there is no difference. Thus, using corollary of Hoeffding's Inequality, we obtain

$$\Pr[|p_+ - \hat{p_+}| \geq \epsilon] \leq e^{-2m\epsilon^2}$$

   Set $e^{-2m\epsilon^2} = \delta$, solve $\epsilon = \sqrt{\frac{\log(2/\delta)}{2m}}$, we obtain

$$\Pr[|p_+ - \hat{p_+}| \leq \sqrt{\frac{\log(2/\delta)}{2m}}] \geq 1 - \delta$$

C. Learning in the presence of noise

1. (a) Suppose there are n_1 positive samples (before randomly flipped) and n_2 negative samples in R, then

$$\epsilon' = \frac{n_2}{n_1 + n_2}.$$

   After randomly flipped,

$$
\begin{aligned}
\epsilon &= \frac{n_2 + \eta n_1}{n_1 + n_2} = \epsilon' + \eta(1 - \epsilon') \leq \epsilon' + \eta'(1 - \epsilon') \\
&\Rightarrow \epsilon' \geq \frac{\epsilon - \eta'}{1 - \eta'}
\end{aligned}
$$

   As said in lecture, assume four smallest rectangle along sides of R such that $\Pr[r_i] \geq \epsilon'/4$. Therefore,

$$
\begin{aligned}
\Pr[R' \text{ misses region } r_i] &\leq 1 - \epsilon'/4 \\
&\leq 1 - \frac{\epsilon - \eta'}{4(1 - \eta')}.
\end{aligned}
$$

(b)

$$
\begin{aligned}
\Pr_{S\sim D^m}[error(R') > \epsilon] &\leq \Pr_{S\sim D^m}[\cup_{i=1}^{4} R' \text{ misses } r_i] \\
&\leq \sum_{i=1}^{4} \Pr_{S\sim D^m}[R' \text{ misses } r_i] \\
&\leq \sum_{i=1}^{4} (1 - \frac{\epsilon - \eta'}{4(1-\eta')})^m \\
&\leq 4exp(-\frac{m(\epsilon - \eta')}{4(1-\eta')})
\end{aligned}
$$

Thus, for $\delta > 0$, the upper bound:

$$4exp(-\frac{m(\epsilon - \eta')}{4(1-\eta')}) \leq \delta \Rightarrow m \geq 4\frac{1-\eta'}{\epsilon - \eta'} \log(\frac{4}{\delta})$$

2. Bonus

   (a) Let x and x' denote the label of a training point received by a learner and the one given by the hypothesis h, then

   $$
   \begin{aligned}
   d(h^*) &= \Pr[x \neq x'] && \text{(definition)} \\
   &= \eta && \text{(flip rate)}
   \end{aligned}
   $$

   (b)

   (c) $d(h) - d(h') = \eta + (1-2\eta)error(h) - \eta > (1-2\eta)\epsilon \geq (1-2\eta')\epsilon \triangleq \epsilon'$

   (d) Suppose the labels are in $\{0, 1\}$, and use *Hoeffding's Inequality*, we obtain

   $$
   \begin{aligned}
   &\Pr[\hat{d}(h^*) - d(h^*) > \epsilon'/2] \leq exp(-2m(\frac{\epsilon'}{2})^2) \\
   &\qquad \leq \delta/2 && \text{(for } m \geq \frac{2}{\epsilon'^2}\log\frac{2}{\delta}) \\
   &\Leftrightarrow \Pr[\hat{d}(h^*) - d(h^*) \leq \epsilon'/2] > 1 - \delta/2
   \end{aligned}
   $$

   (e) Use *generalization bound on finite H*, we obtain

   $$
   \begin{aligned}
   &\Pr\left[d(h) - \hat{d}(h) \leq \sqrt{\frac{\log|H| + \log\frac{2}{\delta}}{2m}}\right] \geq 1 - \delta \geq 1 - \delta/2 \\
   &\Rightarrow \Pr[d(h) - \hat{d}(h) \leq \epsilon'/2] \geq 1 - \delta/2 && (m \geq \frac{2}{\epsilon'^2}(\log|H| + \log\frac{2}{\delta}))
   \end{aligned}
   $$

   (f) Since $m \geq \frac{2}{\epsilon'^2}(\log|H| + \log\frac{2}{\delta})$, the previous three results all hold.

   $$
   \begin{aligned}
   \hat{d}(h) - \hat{d}(h^*) &= (\hat{d}(h) - d(h)) + (d(h) - d(h^*)) + (d(h^*) - \hat{d}(h^*)) \\
   &\geq \left(-\frac{\epsilon'}{2}\right) + \epsilon' + \left(-\frac{\epsilon'}{2}\right) \\
   &\geq 0
   \end{aligned}
   $$

   with probability $1 - \delta$ for $\forall\delta > 0$ because all the previous results guarantee this.