COS511: Theoretical Machine Learning

Homework 1

Kevin Chen

December 11, 2017

Problem. *2a*

Solution. Let us repeat steps 1-7 of the proof of learning bounds from lecture 4, to begin to bound the probability that h_A is consistent and ϵ-bad, by showing that $Pr[h_A$ is consistent and ϵ-bad$] \leq \sum_{h \in B}(1-\epsilon)^m$. These steps have been omitted from the proof for brevity.

For every $x, 1 + x \leq e^x$, so

$$\sum_{h \in B}(1-\epsilon)^m \leq \sum_{h \in B} e^{-em}$$

We know $\epsilon = \frac{1}{m}(ln(1/g(h)) - ln(1/\delta))$ by problem statement.

Thus $-\epsilon m = ln(1/g(h)) - ln(1/\delta)$ by algebra.

And $-\epsilon m = ln(g(h)) + ln(\delta)$ by log rules.

Therefore,

$$\sum_{h \in B} e^{-em} = \sum_{h \in B} e^{ln(g(nh))+ln(\delta)} = \delta \sum_{h \in B} g(h)$$

And since B is a subset of H by definition,

$$\delta \sum_{h \in B} g(h) \leq \delta \sum_{h \in H} g(h)$$

And since $\sum_{h \in H} g(h) \leq 1$ by problem statement,

$$\delta \sum_{h \in H} g(h) \leq \epsilon$$

Thus we have shown that,

$$\forall h_A \in H, Pr[err_D(h_A) > \epsilon] \leq \delta$$

Which means, with probability $\geq 1 - \delta$, and $\forall h \in H$ if h is consistent, then

$$err_D(h) \leq \frac{ln(\frac{1}{g(h)}) + ln(\frac{1}{\delta})}{m}$$

QED

Problem. *2b*

Solution. Since we know that $err_D(h) \leq \frac{ln(\frac{1}{g(h)})+ln(\frac{1}{\delta})}{m}$, to prove that $err_D(h) \leq O(\frac{|h|+ln(\frac{1}{\delta})}{m})$, we need to choose a g such that $O(|h|) = ln(\frac{1}{g(h)})$.

So suppose we construct $O(|H|)$ for some scalar k

$$k|h| = O(|H|) = ln(\frac{1}{g(h)})$$

Then

$$e^{-k|h|} = g(h)$$

.

And now we need $\sum_{h \in H} g(h) = \sum_{h \in H} e^{-k|h|} \leq 1$.

Since H has $2^{|h|}$ bitstrings for each $|h|$,

$$\sum_{h \in H} e^{-k|h|} \leq \sum_{|h|=1} 2^{|h|} e^{-k|h|}$$

By geometric series formula,

$$\sum_{|h|=1} 2^{|h|} e^{-k|h|} = \frac{2}{e^k} \frac{1}{1 - \frac{2}{e^k}} = \frac{1}{\frac{e^k}{2} - 1}$$

Now, if we set k to 4,

$$\frac{1}{\frac{e^k}{2} - 1} = 1 \leq 1$$

Therefore, we have found our k.

Then, $g(h) = e^{-ln(4)|h|} = \frac{1}{4|h|}$.

Choosing g to be $\frac{1}{4|h|}$ satisfies that $g(h) \in (0, 1]$ for all h.

QED.

Problem. *2c*

Solution. In (b), the bound shows that simpler hypotheses probabilistically result in less generalization error than more complex hypotheses, since lowering $|h|$ lowers error.

In (a), the bound shows that more training data (more priors) probabilistically result in less generalization error, since increasing m lowers error.