Stanford CS 229, Public Course, Problem Set 3

Dylan Price

December 7, 2016

1

a)

By the Hoeffding inequality, we know that

$P(|\varepsilon(\hat{h}_i) - \hat{\varepsilon}_{S_{cv}}(\hat{h}_i)| > \gamma) \leq 2\exp(-2\gamma^2\beta m)$

Let A_i denote the event that $|\varepsilon(\hat{h}_i) - \hat{\varepsilon}_{S_{cv}}(\hat{h}_i)| > \gamma$. Then

$$\begin{aligned}
P(\exists \hat{h}_i \in \{\hat{h}_1 ... \hat{h}_k\}. |\varepsilon(\hat{h}_i) - \hat{\varepsilon}_{S_{cv}}(\hat{h}_i)| > \gamma) &= P(A_1 \cup ... \cup A_k) \\
&\leq \sum_k P(A_i) \\
&\leq \sum_k 2\exp(-2\gamma^2\beta m) \\
&= 2k\exp(-2\gamma^2\beta m)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&P(\neg\exists \hat{h}_i \in \{\hat{h}_1 ... \hat{h}_k\}. |\varepsilon(\hat{h}_i) - \hat{\varepsilon}_{S_{cv}}(\hat{h}_i)| > \gamma) \\
&= P(\forall \hat{h}_i \in \{\hat{h}_1 ... \hat{h}_k\}. |\varepsilon(\hat{h}_i) - \hat{\varepsilon}_{S_{cv}}(\hat{h}_i)| \leq \gamma) \\
&\geq 1 - 2k\exp(-2\gamma^2\beta m)
\end{aligned}$$

Let $\dfrac{\delta}{2} = 2k\exp(-2\gamma^2\beta m)$. Then

$$\begin{aligned}
\frac{\delta}{4k} &= \exp(-2\gamma^2\beta m) \\
\frac{4k}{\delta} &= \exp(2\gamma^2\beta m) \\
\log\frac{4k}{\delta} &= 2\gamma^2\beta m \\
\frac{1}{2\beta m}\log\frac{4k}{\delta} &= \gamma^2 \\
\gamma &= \sqrt{\frac{1}{2\beta m}\log\frac{4k}{\delta}}
\end{aligned}$$

Therefore,

$P\left(\forall \hat{h}_i \in \{\hat{h}_1 ... \hat{h}_k\}. |\varepsilon(\hat{h}_i) - \hat{\varepsilon}_{S_{cv}}(\hat{h}_i)| \leq \sqrt{\dfrac{1}{2\beta m}\log\dfrac{4k}{\delta}}\right) \geq 1 - \dfrac{\delta}{2}$

b)

From part (a), we have that

$P(|\varepsilon(\hat{h}_i) - \hat{\varepsilon}_{S_{cv}}(\hat{h}_i)| \leq \gamma) \geq 1 - \frac{\delta}{2}$, where $\gamma = \sqrt{\frac{1}{2\beta m} \log \frac{4k}{\delta}}$

Because $\hat{h} \in \{\hat{h}_1 ... \hat{h}_k\}$

$P(|\varepsilon(\hat{h}) - \hat{\varepsilon}_{S_{cv}}(\hat{h})| \leq \gamma) \geq 1 - \frac{\delta}{2}$

Let $h^* = \arg \min_{\hat{h}_i \in \{\hat{h}_1 ... \hat{h}_k\}} \varepsilon(\hat{h}_i)$

Then with probability at least $1 - \frac{\delta}{2}$

$$
\begin{aligned}
\varepsilon(\hat{h}) &\leq \hat{\varepsilon}_{S_{cv}}(\hat{h}) + \gamma \\
&\leq \hat{\varepsilon}_{S_{cv}}(h^*) + \gamma && \text{By the definition of } \hat{h} \text{ it has the lowest } \hat{\varepsilon}_{S_{cv}} \text{ of any } \hat{h}_i \in \{\hat{h}_1 ... \hat{h}_k\} \\
&\leq \varepsilon(h^*) + 2\gamma && \text{By the uniform convergence result proved in part (a)} \\
&= \min_{i=1,\dots,k} \varepsilon(\hat{h}_i) + 2\gamma && \text{By the definition of } h^* \\
&= \min_{i=1,\dots,k} \varepsilon(\hat{h}_i) + 2\sqrt{\frac{1}{2\beta m} \log \frac{4k}{\delta}} \\
&= \min_{i=1,\dots,k} \varepsilon(\hat{h}_i) + \sqrt{\frac{2}{\beta m} \log \frac{4k}{\delta}}
\end{aligned}
$$