Model selection and validation

1 Let S be an iid sample. Let h be the output of the described Learning algorithm. Note that $L_D(h) = \frac{1}{2}$

Let us calculate the estimate $L_V(h)$. Assume that the parity of S

Fix some fold $\{(x,y)\} \subseteq S$, we distinguish between two cases:

→ The parity of $S \setminus \{x\}$ is 1. It follows that $y = 0$. When being trained using $S \setminus \{x\}$, the algorithm outputs the constant predictor $h(x) = 1$. Hence, the leave-one-out estimate usi this fold is 1

→ The parity of $S \setminus \{x\}$ is 0. It follows that $y = 1$. when being trained using $S \setminus \{x\}$, the algorithm outputs the constant predictor $h(x) = 0$. Hence, the leave-one-out estimate using this fold is 1

2. Consider for example the case in which $H_1 \subseteq H_2 \subseteq \dots H_k$. And $|H_i| = 2^i$ for every $i \in k$, Learning H_k in the Agnostic-Pac model provides the following bound for an ERM hypothisish:

$$L_D(h) \leq \min_{h \in H_k} L_D(h) + \sqrt{\frac{2(k+1+\log(1/\delta))}{m}}$$

$$L_D(h_r) - L_V(h_r) \leq \sqrt{\frac{1}{2\alpha m} \log \frac{4}{\delta}}$$

Applying the union bound, we obtain that with probability at lest $1 - \frac{\delta}{2}$

$$L_D(\hat{h}) \leqslant L_V(\hat{h}) + \sqrt{\frac{1}{2\alpha m}\log\frac{4k}{\delta}} \leqslant L_r(\hat{h}_r) + \sqrt{\frac{1}{2\alpha m}\log\frac{4k}{\delta}}$$

$$\leqslant L_D(h_r) + 2\sqrt{\frac{1}{2\alpha m}\log\frac{4k}{\delta}} \leqslant L_D(h_r) + 2\sqrt{\frac{1}{\alpha m}\log\frac{4k}{\delta}}$$

$$= L_D(h_r) + \sqrt{\frac{2}{\alpha m}\log\frac{4k}{\delta}}$$

In particular, with probability at least $1-\frac{\delta}{2}$ we have

$$L_D(\hat{h}) \leqslant L_D(\hat{h}_j) + \sqrt{\frac{2}{\alpha m}\log\frac{4k}{\delta}}$$

Using similar arguments, we obtain that with probability at least $1-\frac{\delta}{2}$

$$L_D(\hat{h}_j) \leqslant L_D(h^*) + \sqrt{\frac{2}{(1-\alpha)m}\log\frac{4|H_j|}{\delta}}$$

$$\cancel{\leqslant L_D(h^*)}$$

$$\Rightarrow L_D(\hat{h}) \leqslant L_D(h^*) + \sqrt{\frac{2}{\alpha m}\log\frac{4k}{\delta}} + \sqrt{\frac{2}{(1-\alpha)m}\log\frac{4|H_j|}{\delta}}$$

$$L_D(\hat{h}) \leqslant L_D(h^*) + \sqrt{\frac{2}{\alpha m}\log\frac{4k}{\delta}} + \sqrt{\frac{2}{(1-\alpha)m}\left(j+\log\frac{4}{\delta}\right)}$$