Machine Learning Exercises - chapter 11

Mohammadjavad Abbaspour

11.1. Model selections and validation

Failure of k-fold cross validation: Let S be an i.i.d. and h be the output of the described l_A. h is a constant function and the independantly of idents of S show that: $L_p(h) = \frac{1}{2}$.

Now we calcute the estimate $L_V(h)$. fix i old $\{(x_i, y_i)\} \subseteq S$.

we have 2 cases:

1) Parity of $S \setminus \{x_i\}$ is 1, then $h(x_i) = 1$, so the leave-one-out estimate using i old is 1.

2) Parity of $S \setminus \{x_i\}$ is 0, then $h(x_i) = 0$, so the leave one out estimate using i old is 1.

The estimate error of h is 1. (with Averaging over the folds)

the defrences between the estimate and the true error is $\frac{1}{2}$.

11.2. Assume that $H_1 \subseteq H_2 \subseteq H_3 \subseteq \dots \subseteq H_k$. $|H_i| = 2^i$. $(\forall i \in k)$

$$L_D(h) \leq \min L_D(H) + \sqrt{\frac{2(k+1, \log(1/\delta))}{m}}.$$

Asume that $j \rightarrow$ minimal index which contains $h^* \in \arg_{h \in H} L_D(h)$.

By Hoeffding and probability at least $1 - \frac{\delta}{2k}$:

$$|L_D(\hat{h}_r) - L_V(\hat{h}_r)| \leq \sqrt{\frac{1}{2am} \log \frac{4}{\delta}}.$$

and with applying union bound we have: $(1 - \frac{\delta}{2}$ prob$)$

$$L_D(\hat{h}) \leq L_V(\hat{h}) + \sqrt{\frac{1}{2am} \log \frac{4k}{\delta}}$$

$$\leq L_V(\hat{h}_r) + \sqrt{\frac{1}{2am} \log \frac{4k}{\delta}}$$

$$\leq L_D(\hat{h}_r) + 2\sqrt{\frac{1}{2am} \log \frac{4k}{\delta}} = L_D(\hat{h}_r) + \sqrt{\frac{2}{am} \log \frac{4k}{\delta}}.$$

Now we have: $L_D(\hat{h}_j) \leq L_D(h^*) + \sqrt{\frac{2}{(1-\alpha)m} \log \frac{4|H_j|}{\delta}} =$ (with prob $1 - \frac{\delta}{2}$)

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

And with prob $1 - \delta$ we have:

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

At last with comparing over two bounds, the optimal index j is smaller that k. So the model selection is much better that that. (the logarithmic improvement)