تمرین ۱

فرض کنید که S یک نمونه مستقل و هم توزیع باشد .طبق صورت سوال، یک الگوریتم یادگیری داریم که خروجی آن تابع h است .

$$h(x) = \begin{cases} 1 & \text{the parity of the labels on the training set is 1} \\ 0 & \text{OtherWise} \end{cases}$$

چون که h یک تابع ثابت است بنابراین داریم که

$L_{\mathrm{D}}(h) = 1/2$

میخواهیم که $L_{\mathrm{V}}(h)$ را تخمین بزنیم.

فرض میکنیم که

the parity of S is 1

دو حالت را در نظر میگیریم.

حالت اول:

the parity of $S \setminus X$ is 1

نتیجه میدهد که $y = 0$

الگوریتم به عنوان خروجی تابع ثابت $h(x) = 1$ را میدهد. بنابراین

the leave-one-out estimate using this fold is 1

حالت دوم:

the parity of $S \setminus X$ is 0

نتیجه میدهد که $y = 1$

الگوریتم به عنوان خروجی تابع ثابت $h(x) = 0$ را میدهد بنابراین

the leave-one-out estimate using this fold is 1

Averaging over the folds, the estimate of the error of h is 1. Consequently, the difference between the estimate and the true error is 1/2. The case in which the parity of S is 0 is analyzed analogously .

تمرین ۲

برای مثال این حالت را در نظر بگیرید

$\mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq \cdots\cdots \subseteq \mathcal{H}_{\mathrm{k}}$

و فرض کنید برای هر

$i \in k$

$|\mathcal{H}_{\mathrm{i}}| = 2^i$

باند بالا برای $\mathrm{L_D(h)}$ اینگونه است

$$\mathrm{L_D(h)} \leq \min_{h\in\mathcal{H}_k} L_{\mathrm{D}}(h) + \sqrt{\frac{2(k+1+log(1/\delta)}{m}}$$

فرض میکنیم j هست minimal index که شامل h^* است

$$h^* \in \mathrm{argmin}_{h\in\mathcal{H}_k} L_{\mathrm{D}}(h)$$

$r \in [k]$ را فیکس میکنیم

بنا به نامساوی Hoeffding با احتمال حداقل $1-\delta/2k$ داریم

$$|L_D(\hat{h}_{\mathrm{r}}) \text{ - } L_V(\hat{h}_{\mathrm{r}})| \leq \sqrt{(1/2\alpha m) log 4/\delta}$$

با احتمال حداقل $1\text{-}\delta/2$ نامساوی زیر به طور همزمان برای هر $i \in [k]$ بر قرار است

$$L_D(\hat{h}) \leq L_V(\hat{h}) + \sqrt{(1/2\alpha m) log 4k/\delta} \leq L_V(\hat{h}_{\mathrm{r}}) + \sqrt{(1/2\alpha m) log 4k/\delta} \leq L_D(\hat{h}_{\mathrm{r}}) + 2\sqrt{(1/2\alpha m) log 4k/\delta} = L_D(\hat{h}_{\mathrm{r}}) + \sqrt{(2/\alpha m) log 4k/\delta}$$

به ویژه با احتمال حداقل $1\text{-}\delta/2$ داریم که

$$L_D(\hat{h}) \leq L_D(\hat{h}_{\mathrm{j}}) + \sqrt{(2/\alpha m) log 4k/\delta}$$

با احتمال حداقل $1\text{-}\delta/2$ به دست میاد که

$$L_D(\hat{h}_{\mathrm{j}}) \leq L_D(h^*) + \sqrt{(\tfrac{2}{(1-\alpha)m}) log \tfrac{4|\mathcal{H}_{\mathrm{j}}|}{\delta}} = \mathrm{L_D}(h^*) + \sqrt{(\tfrac{2}{(1-\alpha)m}) log \tfrac{4|\mathcal{H}_{\mathrm{j}}|}{\delta}}$$

با دو نا مساوی اخیر داریم که با احتمال $1\text{-}\delta$

$$L_D(\hat{h}) \leq L_D(h^*) + \sqrt{(\tfrac{2}{\alpha m}) log \tfrac{4k}{\delta}} + \sqrt{(\tfrac{2}{(1-\alpha)m}) log \tfrac{4|\mathcal{H}_{\mathrm{j}}|}{\delta}}$$

نتیجه میگیریم که :

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