فاطمه داوری

$\varepsilon, \delta \in (0,1)$

اگر k قسمت با سایز $m_H(\frac{\varepsilon}{2})$ انتخاب کنیم و الگوریتم A را روی این k قسمت اجرا کنیم خروجی الگوریتم را $\hat{h}_1, \dots, \hat{h}_k$ می‌نامیم.

می‌دانیم با احتمال حداقل $1-\delta/2$ داریم: $\min_{i\in[k]} L_D(\hat{h}_i) \le \min L_D(h) + \frac{\varepsilon}{2}$ است.

$\hat{H} = \{\hat{h}_1, \dots, \hat{h}_k\}$ الگوریتم ERM را اگر روی $\hat{H}$ با اندازه $\left\lceil \frac{2\log(4k/\delta)}{\varepsilon^2} \right\rceil$ اجرا کنیم

آنگاه، با احتمال حداقل $1-\delta/2$: $L_D(\hat{h}) \le \min_{i\in[k]} L_D(h_i) + \frac{\varepsilon}{2}$ است.

بنابراین با احتمال حداقل $1-\delta$ داریم:

$$L_D(\hat{h}) \le \min_{i\in[k]} L_D(h_i) + \frac{\varepsilon}{2}$$

$$\le \min_{h\in H} L_D(h) + \varepsilon$$

×۳

فاطمه داوری

۵-۴

همانطور که ثابت کردیم در تمرین ۳:

$|\mathcal{X}| = n$

$VCDim(\mathcal{B}) = VCDim(L(\mathcal{B},T)) = \log_2 2^n = n$

$\mathcal{B} = \{h_{j,b,\theta} : j \in [d],\ b \in \{\pm 1\},\ \theta \in \mathbb{R}\}$

$h_{j,b,\theta}(x) = b\,\text{sign}(\theta - x_j)$

$\forall j \in [d] \quad \mathcal{B}_j = \{h_{b,\theta} : b \in \{\pm 1\},\ \theta \in \mathbb{R}\}$

$h_{b,\theta}(x) = b\,\text{sign}(\theta - x_j)$

$VCDim(\mathcal{B}_j) = 2$

طبق تمرین ۱۱ فصل ۶:

$VCDim(\bigcup_{i=1}^{r} H_i) \leq 4d\log(2d) + 2\log(r)$

$\mathcal{B} = \bigcup_{j=1}^{d} \mathcal{B}_j$

$VCDim(\bigcup_{j=1}^{d} \mathcal{B}_j) \leq 4 \times 2\log_2(4) + 2\log(d)$

$= 16 + 2\log(d)$

Scanned with CamScanner

فاطمه داوری

11-1

با در نظر گرفتن مجموعه آموزشی S و تابع h تابع ثابت باشد

$L_D(h) = \frac{1}{2}$

فرض می‌کنیم m زوج باشد که S را داشته باشیم و اگر نقطه (x,y) را از آن در نظر بگیریم $y=1$ به ترتیب m است.

اگر $S|\{x\}$ برابر ۱ باشد یعنی به ترتیب m برابر صفر است $(y=0)$ بنابراین $h(x)=1$ باشد طبق صورت سوال
و اگر $S|\{x\}$ برابر صفر باشد یعنی $y=1$ است و $h(x)=0$ می‌باشد

بنابراین خطا برابر ۱ و تفاوت آن با $L_D(h)$ برابر $\frac{1}{2}$ است.

estimation error

$e_{est} - L_D(h) = 1 - \frac{1}{2} = \frac{1}{2}$

فاطمه باوری

11-2

درنظر می‌گیریم:

$H_1 \subseteq H_2 \subseteq \dots \subseteq H_k$ $\quad |H_i| = 2^i$ $\quad \delta \in (0,1)$

بنابر الگوریتم ERM

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

$$L_D(\hat{h}_r) - L_V(\hat{h}_r) \leq \sqrt{\frac{1}{2\alpha m}\log\frac{4}{\delta}} \qquad r \in [k]$$ با احتمال $1-\frac{\delta}{2k}$

با احتمال $1-\delta/2$

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

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

اگر برای یک j داشته باشیم: $h^* \in \arg\min_{h \in H} L_D(h)$ آنگاه با احتمال $1-\delta/2$

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

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

یادگیری ماشین　　　　فاطمه داوری

18-2

$$IG(X) = H(Y) - H(Y|X)$$

$$H(Y) = -\left[P(Y=0)\cdot\log_2 P(Y=0) + P(Y=1)\log_2 P(Y=1)\right]$$

$$= -\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right] = 1$$

X_1:

$$P(Y|X_1) = -P(X_1=0)\left[P(Y=0|X_1=0)\cdot\log_2 P(Y=0|X_1=0) + P(Y=1|X_1=0)\log_2 P(Y=1|X_1=0)\right]$$

$$- P(X_1=1)\left[P(Y=0|X_1=1)\log_2 P(Y=0|X_1=1) + P(Y=1|X_1=1)\log_2 P(Y=1|X_1=1)\right]$$

$$= -\frac{1}{4}\left(1\,\cancel{\log_2(1)} + 0\,\cancel{\log_2 0}\right) - \frac{3}{4}\left(\frac{1}{3}\log_2\frac{1}{3} + \frac{2}{3}\log_2\frac{2}{3}\right)$$

$$= -\frac{3}{4}(-0.53 - 0.4) = 0.7$$

X_2:

$$P(Y|X_2) = -P(X_2=0)\left[P(Y=0|X_2=0)\log_2 P(Y=0|X_2=0) + P(Y=1|X_2=0)\log_2 P(Y=1|X_2=0)\right]$$

$$- P(X_2=1)\left[P(Y=0|X_2=1)\log_2 P(Y=0|X_2=1) + P(Y=1|X_2=1)\log_2 P(Y=1|X_2=1)\right]$$

$$= -\frac{1}{2}\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right] - \frac{1}{2}\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right]$$

$$= \frac{1}{2} + \frac{1}{2} = 1$$

فائزه داوری.

ادامه ۴-۱۸

X_3:

$$P(Y|X_3) = -P(X_3=0)\Big[P(Y=0|X_3=0)\log_2 P(Y=0|X_3=0) + P(Y=1|X_3=0)\log_2 P(Y=0|X_3=0)\Big]$$

$$-P(X_3=1)\Big[P(Y=0|X_3=1)\log_2 P(Y=0|X_3=1) + P(Y=1|X_3=1)\log_2 P(Y=1|X_3=1)\Big]$$

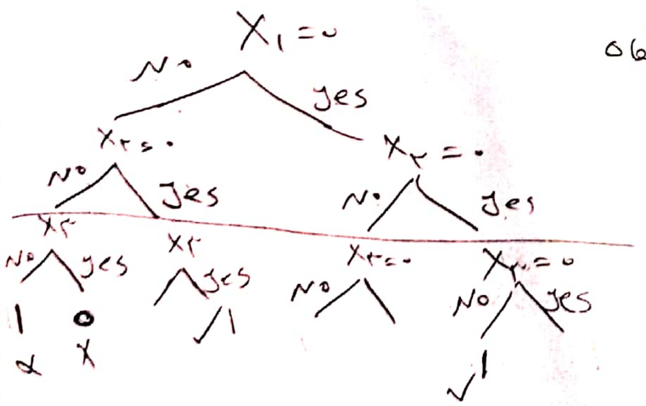
$$= -\frac{1}{2}\Big(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\Big] - \frac{1}{2}\Big[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\Big] = 1$$

$$\Rightarrow IG(X_1) = H(Y) - H(Y|X_1) = 1 - 0.7 = 0.3$$

$$IG(X_2) = H(Y) - H(Y|X_2) = 1 - 1 = 0$$

$$IG(X_3) = H(Y) - H(Y|X_3) = 1 - 1 = 0$$

$\Rightarrow$ X_1 انتخاب می‌شود که در ریشه قرار می‌گیرد



اگر هر یک از X_1، X_2 را قرار دهیم به عنوان متغیر سطح دوم آنگاه چون متغیر سوم نمی‌تواند همه ۴ نمونه train را درست پیش‌بینی کند بنابراین خطای آموزش حداقل $\frac{1}{4}$ است

ب