یادگیری ماشین   فاطمه داوری   400422079

تمرین 2.1

$$S = \{(x_i, f(x_i))\}_{i=1}^{m} \subseteq (\mathbb{R}^d \times \{0,1\})^m$$

داشتیم در درس معادله‌ی (۲.۳) را،

$$h_S(x) = \begin{cases} y_i & : \text{if } \exists i \in [m] \text{ such that } x_i = x \\ 0 & \text{else} \end{cases}$$

باید نشان دهیم با توجه به یک مجموعه آموزشی $S=(x_i, y_i)$ یک چندجمله‌ای $P(S)$ وجود دارد، $h_S(x)=1$ است اگر و تنها اگر $P_S(x) \geq 0$ و جایی که $h_S(x)=0$ است $P_S(x)<0$ است.

$(x-x_i)$ یک بردار است و $\|x-x_i\|$ یک عدد است برای هر i. اگر بخواهیم ازای $h_S(x)=1$ است، $P_S(x) \geq 0$ باشد می‌توانیم تعریف کنیم $P_S(x) = \prod_{\substack{i=1 \\ y_i=1}}^{m} \|x-x_i\|$ که وقتی $x=x_i$ است برابر صفر و در غیر اینصورت مثبت است. $(\|x-x_i\| \geq 0)$

البته باید در نظر بگیریم که باید برای $h_S(x)=0$ ، $P_S(x)<0$ باشد پس می‌توانیم تعریف کنیم

$$P_S(x) = -\prod_{\substack{i=1 \\ y_i=1}}^{m} \|x-x_i\|$$

$$u = (u_1, u_2, \dots, u_m), \quad v = (v_1, \dots, v_m) \Rightarrow \|u-v\| = \sqrt{(u_1-v_1)^2 + \dots + (u_m-v_m)^2} \geq 0$$

نائمه دلاوری

۴۰۰۴۲۲۰۷۹

2.2

$$L_{D,f}(h) \overset{?}{=} E(L_S(h))$$

$$L_S(h) = \frac{1}{m}\sum_{i=1}^{m} 1_{h(x_i)\neq f(x_i)}$$

به دلیل خطی بودن امید $E(\Sigma) = \Sigma E$

$$E(L_S(h)) = E_{S\sim D^m}[L_S(h)] = \frac{1}{m}\sum E_{S\sim D^m}[1_{h(x_i)\neq f(x_i)}]$$

$$= \frac{1}{m}\sum_i E_{S\sim D^m}[1_{h(x)\neq f(x)}]$$

$\xrightarrow{\text{دلیل}}$ به دلیل iid بودن x_i ها یعنی چون x_i ها مستقل هستند و تصادفی انتخاب شده‌اند به جای x_i می‌نویسیم x

$$= \frac{m}{m}\sum E_{S\sim D^m}[1_{h(x)\neq f(x)}]$$

$\xrightarrow{\text{دلیل}}$ چون $\sum_i = m$

$$= E_{S\sim D^m}[1_{h(x)\neq f(x)}]$$

$$= E_{X\sim D}[1_{h(x)\neq f(x)}]$$

$\xrightarrow{\text{دلیل}}$ چون S از یک مجموعه m تایی می‌آید وقتی به جای x_i ، X را در نظر می‌گیریم یعنی فقط یک x را در نظر گرفتیم پس می‌توانیم به جا $S\sim D^m$ بنویسیم $X\sim D$

$$= L_{D,f}(h)$$

یادگیری ماشین فاطمه بلدری ۴۰۰۴۲۲۰۷۹

<u>2.3</u>

A: الگوریتمی که کوچکترین مستطیل نزدیک به همه مثال های مثبت در داده های آموزش را برمی گرداند است پس A تمام نقاطی که در جعبه مثبت دارند را درست تشخیص می دهد. ما بقیه جاهایی که منفی هستند را نیز درست تشخیص می دهد چون طبق فرض سوال شرط realizability برقرار است یعنی یک $h^* \in H$ وجود دارد که خطای حقیقی صفر دارد. پس می توان گفت A یک ERM است یعنی می توان با آن تابعی پیدا کرد که کمترین خطای تجربی را داشته باشد.

تابع احتمالی که روی X وجود دارد و با آنرا نشان می دهیم : D ، $X \sim D$

$R^* = R(a_1^*, b_1^*, a_2^*, b_2^*)$ $\longrightarrow$ مستطیلی است که جعبه ها را ایجاد می کند

A الگوریتمی است که وقتی که داده S را به آن می دهیم خروجی آن $R(S)$ است که نشان می دهد.

$R \subseteq R^*$

$R_1 = R\{a_1^*, a_1, a_2^*, b_2^*\}$, $R_2 = R\{b_1, b_1^*, a_2^*, b_2^*\}$

$R_3 = R\{a_1^*, b_1^*, a_2^*, a_2\}$, $R_4 = R\{a_1^*, b_1^*, b_2, b_2^*\}$

نقطه سوم $\varepsilon \in (0,1)$, $i \in \{1,2,3,4\}$

$N_i = \{S|_x : S \cap R_i = \phi\}$

$$D^m(\{S|_x : L(A(S)) > \varepsilon\}) \leq D^m\left(\bigcup_{i=1}^{4} N_i\right)$$

$$\leq \sum_{i=1}^{4} D^m(N_i)$$

فاطمه نادری.

ادامه 2-3

طبق بسط تیلور

$e^{x} \simeq 1+x$

$$D^m(N_i) = \left(1-\frac{\varepsilon}{4}\right)^m \le e^{-\frac{m\varepsilon}{4}}$$

$$\Rightarrow D^m\left(\{S|_x : L_{D,f}(A(S)) \ge \varepsilon\}\right) \le 4e^{-\frac{m\varepsilon}{4}}$$

$$4e^{-\frac{m\varepsilon}{4}} < \delta \Rightarrow -\frac{m\varepsilon}{4} \le \log\frac{\delta}{4}$$

$$\Rightarrow m \ge \frac{-4\log\left(\frac{4}{\delta}\right)}{\varepsilon}$$

داشتیم در قسمت اول سوال که

$$h = \begin{cases} 1 & \text{if } a_1 \le x_1 \le b_1 ,\ a_2 \le x_2 \le b_2 \\ 0 & \text{o.w.} \end{cases}$$

در حالت d بعدی داریم

$$h = \begin{cases} 1 & \text{if } a_i \le x_i \le b_i \quad \forall i \in d \\ 0 & \text{o.w} \end{cases}$$

و در حالت d بعدی می‌توان $2d$ تا R تعریف کرد

$$m \ge \frac{-d\log\left(\frac{d}{\delta}\right)}{\varepsilon}$$

حالت قبل ۲ بعدی بود و ما ۴ تا R داشتیم یعنی ۲ برابر

پس زمانی که d بعدی است ما $2d$ تا R داریم.

مقدار m به d، δ، ε بستگی دارد پس یک چند جمله‌ای است.