1

$$h_S(x) = \begin{cases} y_i & , \quad \exists i \in [m] \ s.t \ x_i = x \\ 0 & , \quad o.w \end{cases}$$

اگر چند جمله‌ای $p(s)$ را طوری بیابیم که در نقاط $x_i=1$ ، $p(s)\geq 0$ و در غیر این صورت $p(s)<0$ باشد؛ بیش برازش (over fitting) رخ می‌دهد. پس اگر $p(s)$ به صورت زیر باشد

$$P_s(x) = -\left|(x-x_1)(x-x_2)\cdots(x-x_i)\right|$$

و $x_i , \ldots , x_1$ عضو $S$ باشند که دارای لیبل 1 هستند در جایی که $h_s(x)=1$ است، $P_s(x)=0$ می‌شود و در غیر این صورت $p(s)$ منفی خواهد بود.

---

2

$$E_{S\sim D^m}\left[L_S(h)\right] = \frac{1}{m}\sum_{i=1}^{m} E_{S\sim D^m}\left[\mathbb{1}_{h(x_i)\neq f(x_i)}\right]$$

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

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

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

(3) (a)

برای اثبات ERM بودن A باید نشان دهیم برای هر مستطیل مانند $A^*$ که بزرگتر یا کوچکتر از A باشد، فرض ERM برقرار نیست.

(I) اگر $A^*$ کوچکتر از A باشد، تعدادی از + ها را در بر نمی‌گیرد و $L_S(A^*) > L_S(A)$ خواهد شد.

(II) اگر $A^*$ بزرگتر از A باشد پس تعدادی از - ها را در بر خواهد گرفت پس $L_S(A^*) > L_S(A)$ می‌باشد.

با توجه به (I) و (II) نتیجه می‌گیریم A، ERM است.

---

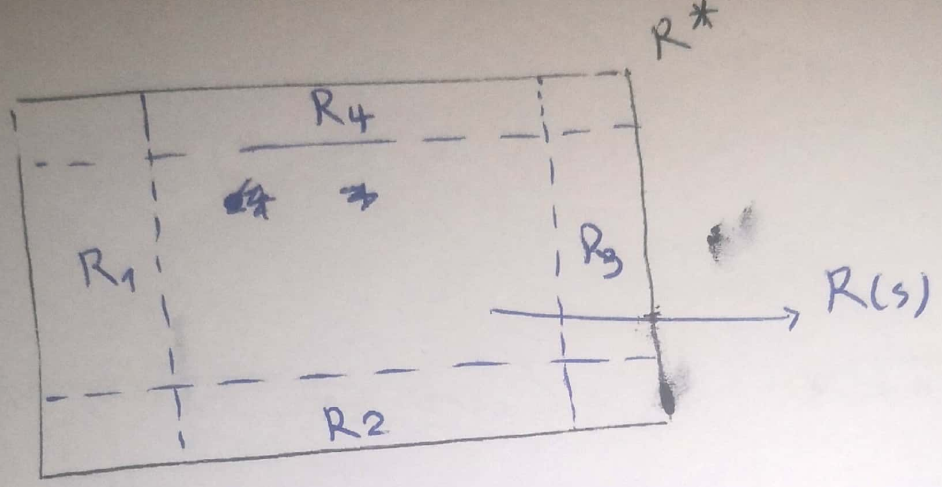
(3) (b)

با توجه به فرض سؤال مستطیل حاصل از A را با $R(S)$ نمایش دهیم می‌دانیم $R(S) \subseteq R^*$ است و داریم

$$L_{(D,f)}(R(S)) = D(R^* - R(S))$$

به ازای هر $\epsilon \in (0,1)$، ~~region~~ مجموعه فرضیات زیر را تعریف می‌کنیم:

$$F_i = \{ S|_x : S_x \cap R_i = \emptyset \}$$




$$D^m\left(\{s : L_{(D,f)}(A(s)) > \epsilon\}\right) \le D^m\left(\bigcup_{i=1}^{4} F_i\right) \le \sum_{i=1}^{4} D^m(F_i)$$

احتمال این که یک نمونه داخل $F_i$ باشد (برای $i$ مشخص) برابر است با
این که هیچ نمونه‌ای در $R_i$ نباشد $= (1-\epsilon/4)$ . چون $m$ نمونه وجود
دارد پس $(1-\frac{\epsilon}{4})^m$ پس

$$D^m(F_i) = \left(1-\frac{\epsilon}{4}\right)^m \le e^{-m\epsilon/4}$$

$$\longrightarrow D^m\left(\{s : L_{(D,f)}(A(s)) > \epsilon\}\right) \le 4e^{-m\epsilon/4} < \delta$$

$\vdots$

$$m \ge \frac{4\log(4/\delta)}{\epsilon}$$

(c) (3)

با تعمیم قسمت قبل به $\mathbb{R}^d$ داریم:

$$h_{(a_1,b_1,\dots,a_d,b_d)}(x_1,\dots,x_d)=\begin{cases}1 & \forall i\in[d],\ a_i\le x_i\le b_i\\ 0 & \text{o.w}\end{cases}$$

مجموعه فرض ها به صورت زیر تعریف می شود

$$H=\{h_{(a_1,b_1,\dots,a_d,b_d)}:\forall i\in[d],\ a_i\le b_i\}$$

الگوریتم ERM گفته شده برای $d$ بُعد هم صدق می کند (و به ازای اضافه شدن هر بعد 2 region جدید اضافه می شود.)

محدوده ی $m$ به صورت زیر بدست می آید.

$$m\geq\frac{2d\log\left(\frac{2d}{\delta}\right)}{\varepsilon}$$