2

$$E_{S|x\sim D^m}\left[L_S(h)\right] = L_{(D,f)}(h)$$

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

$$= \frac{1}{m}\sum_{i=1}^{m} E\left[1_{h(x_i)\neq f(x_i)}\right]$$

$$\overset{i.i.d}{=} \frac{1}{m}\sum P(h(x)\neq f(x))$$

$$= \cancel{\frac{1}{m}\cdot m}\left[P(h(x_i)\neq f(x_i))\right]$$

$$= P(h(x_i)\neq f(x_i)) = L_{D,f}(h)$$

3

۱) همانطور که در صورت سوال گفته شده، A مستطیل برای داده‌های با لیبل مثبت است پس تمام داده‌های آموزشی با لیبل مثبت را درست تشخیص می‌دهد.

حال با استفاده از فرض Realizability می‌توان گفت که تمام داده‌های منفی را نیز درست تشخیص می‌دهد زیرا وجود دارد یک h^* در مجموعه فرضیات که $P(h^*(x)\neq f(x))=0$ است

پس روی "S" خطای تجربی صفر می‌شود که ERM ثابت می‌شود.

۲) ابتدا با استفاده از فرضیات مسئله:

1) $X \sim D$

2) $R^* = (a_1^*, b_1^*, a_2^*, b_2^*)$

3) f : True label

4) $A(S) = R(S)$

خروجی الگوریتم

می‌دانیم که: $\forall S ; R(S) \subseteq R^*$;

$$L_{D,f}(R(S)) = P(R^* \neq R(S))$$

حال می‌دانیم که:

$$\begin{cases} R_1 = R(a_1^*, a_1, a_2^*, b_2^*) \\ R_2 = R(b_1, b_1^*, a_2^*, b_2^*) \\ R_3 = R(a_1^*, b_1^*, a_2^*, d_2) \\ R_4 = R(a_1^*, b_1^*, b_2, b_2^*) \end{cases}$$

برای پیدا کردن خطا با فرض $\varepsilon \in (0,1)$ داریم:

خطا

$$\forall i \in [4] ; \quad D_i = \{ S|_x : S \cap R_i = \emptyset \}$$

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

دست و دلبازی می‌کنیم

$$\xRightarrow[\text{Union bound}]{\text{قضیه}} \leqslant \sum_{i=1}^{4} D^m(D_i)$$

سری تیلور

$$\xRightarrow{\text{یکی از عبارت‌ها را در ۴ ضرب کردیم}} D^m(D_i) = (1 - \varepsilon/4)^m \leqslant e^{-\frac{m\varepsilon}{4}}$$

چهار تا

$$\Rightarrow D^m(\{ S|_x : L_{D,f}(A(S)) \geqslant \varepsilon \}) \leqslant 4 \cdot e^{-\frac{m\varepsilon}{4}}$$

حال با استفاده از sample complexity ،

$$4e^{-\frac{m\varepsilon}{4}} \leq \delta \Rightarrow -\frac{m\varepsilon}{4} \leq \ln \frac{\delta}{4}$$

$$\Rightarrow m \geq \frac{-4\ln(\frac{4}{\delta})}{\varepsilon}$$

---

3) قسمت (2) برای حالت دو بعدی $h = \begin{cases} 1 & \text{if } a \leq x \leq b \text{ و } a_2 \leq x_2 \leq b_2 \\ 0 & \text{o.w} \end{cases}$ بود.

در این قسمت آن را تعمیم به d بعدی می‌دهیم:

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

✓. در حالت ۲ بعدی ۴ مستطیل (R_1, R_2, R_3, R_4) را تعریف کردیم، پس می‌توان نتیجه گرفت که $m \geq m_H(\varepsilon, \delta)$ به صورت:

$$m \geq \frac{-2d \ln(\frac{2d}{\delta})}{\varepsilon}$$

✓ چون بالا ۲ بعد بودن $(d=2)$ بود.

---

4) واضح است که چون sample complexity $(m \geq m_H(\varepsilon,\delta))$ را می‌خواهیم به دست آوریم مقدار m بستگی به $\ln(\frac{1}{\delta})$ و $\frac{1}{\varepsilon}$ و d دارد که می‌دانیم الگوریتم A در d و $\frac{1}{\varepsilon}$ و $\ln(\frac{1}{\delta})$، یک Polynomial است.
