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

ما به دنبال $P(S)$ هستیم که شرط مقابل برقرار باشد

$\exists P_S \quad h_S(x) = 1 \iff P_S(x) \geq 0$

از آنجایی که $h(x)$ براساس ERM ساخته شده پس زمانی که $h(x)=1$ باشد

$P(x)$ هم 1 است بدین ترتیب اگر $P_S(x) = -(x-x_1)\times \dots \times (x-x_i)$

فرض کنیم مثلاً $P_S(x)=0$ و برابر $h(x)=0$ مثلاً چند جمله‌ای $P_S(x)$ منفی خواهد بود

و در این صورت $P_S(x)$ به شدت overfitting خواهد داشت

Date .... Subject ....

ادامه ۱ زیرا برای داده های آموزشی $P_s(x)$ مثبت خواهد بود

و به ازای هر داده ی دیگر $P_s(x) \leq 0$ خواهد بود پس به شدت overfitting

دارند / ریشه های $P_s(x)$ در مجموعه آموزشی است / ($p(x)=1$)

2.4 Exercise 2

p: $\underset{S|_x \sim D^n}{E} [L_S(h)] = L_{(D,f)}(h)$

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

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

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

Scanned with CamScanner

Date .... Subject ....

3) بخش اول: می‌خواهیم ثابت کنیم A یک ERM است یعنی

کمترین خطای تجربی را دارد. برای اثبات کافی است ثابت کنیم برای هر

مستطیل مانند A^* که بزرگتر یا کوچکتر باشد فرض ERM برقرار نیست

۱) اگر $\tilde{A}^*$ مستطیل بزرگتر از A باشد پس تعدادی از نقاط منفی را دربرمی‌گیرد

پس $L_S(\tilde{A}^*) > L_S(A)$ خواهد شد و A^* ERM نخواهد بود

۲) اگر $\tilde{A}^*$ مستطیل کوچکتر از A باشد یکسری نقاط مثبت هست که دربرنمی‌گیرد

پس $L_S(\tilde{A}^*) > L_S(A)$ خواهد شد و $\tilde{A}^*$ ERM نخواهد بود

پس از موارد ۱ و ۲ می‌توان نتیجه گرفت الگوریتم A دارای

خاصیت ERM است.

Scanned with CamScanner

۳ قسمت ب

همانطور که در قسمت قبل ثابت کردیم الگوریتم A دارای خاصیت ERM است

پس کوچکترین مستطیل است پس حتماً $R(S) \subseteq \hat{R}^*$ است

همانطور که در قسمت راهنمایی اشاره شد R به طور مشابه برای R_2 و R_3 و R_4 به همین صورت

تعریف می‌شود و به طور کلی R_i $i \in [4]$

مجموعه F را برای همه هایی تعریف می‌کنیم که در R_i نباشد

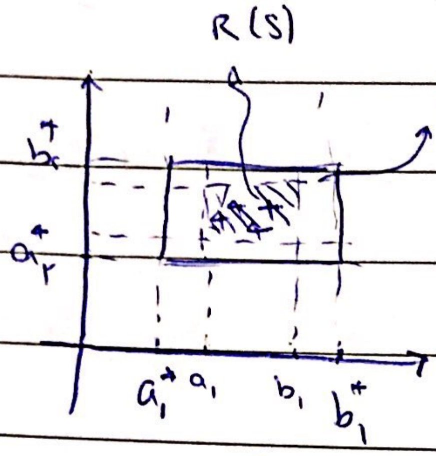
$$F_i = \{S|_x : S|_x \cap R_i = \Phi\}$$

مطلق تعریف داریم برای هر R_i چگالی احتمال $\frac{\epsilon}{4}$ است

Date .... Subject ....

Union bound

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



$$D^m(F_i) = \left(1 - \frac{\varepsilon}{4}\right)^m \leq \exp\left(-m\frac{\varepsilon}{4}\right)$$

$$= e^{-\frac{\varepsilon}{4} \times m}$$

$$4\exp\left(-m\frac{\varepsilon}{4}\right) \leq \delta \Rightarrow \frac{4}{e^{\frac{m\varepsilon}{4}}} \leq \delta \rightarrow \frac{4}{\delta} \leq e^{\frac{m\varepsilon}{4}}$$

$$\ln\frac{4}{\delta} \leq \frac{m\varepsilon}{4} \rightarrow m \geq \frac{4\ln 4/\delta}{\varepsilon}$$

Scanned with CamScanner

Date .... Subject ....

بخش دوم در حالت R^d داریم

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

همانند قسمت قبل عمل می‌کنیم اما وقتی قبل ۴ بود اما اینجا $2d$ خواهد بود

قبلا R^2 بود

پس همانند قبل داریم $\left[\frac{2d\log(2d/\delta)}{\epsilon}\right]$