10.1

ابتدا به اندازه k قسمت از $m_H(\varepsilon/2)$ برمیداریم که $0<\varepsilon<1$ و هربار A رو روی قسمت ها اعمال میکنیم تا به مجموعه $\hat{H} := \{\hat{h}_1, \ldots, \hat{h}_k\}$ برسیم . با توجه به اینکه احتمال $\min_{i\in[k]} L_D(\hat{h}_i) \le \min L_D(h) + \varepsilon/2$ میباشد یک ERM روی H اعمال میکنیم و داده های آموزشی آخرین قسمت k تایی با سایز زیر باشند:

$$\left\lceil \frac{2\log(4k/\delta)}{\epsilon^2} \right\rceil$$

با وجود h داریم : $1 - \delta/2, L_D(\hat{h}) \le \min_{i\in[k]} L_D(h_i) + \varepsilon/2$

و با اعمال union bound میرسیم که با احتمال $1-\delta$

$$\begin{aligned} L_{\mathcal{D}}(\hat{h}) &\le \min_{i\in[k]} L_{\mathcal{D}}(h_i) + \epsilon/2 \\ &\le \min_{h\in\mathcal{H}} L_{\mathcal{D}}(h) + \epsilon \ . \end{aligned}$$

10.4

a.

فرض کنید X یک مجموعه محدود به اندازه n باشد. فرض کنید B کلاس همه توابع از X تا {0، 1} باشد. سپس $L(B, T) = B$ و هر دو متناهی هستند. بنابراین، برای هر $T \ge 1$:

$VCdim(B) = VCdim(L(B, T)) = \log 2^n = n$

b.

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

اگر

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

بنابراین:

$j\in[d], \mathrm{let} B_j = \{h_{b,\theta} : b\in\{-1,1\}, \theta\in R\}$

$VCdim(B_j) = 2$

c.

با توجه به راهنمایی، برای هر $i \in [Tk/2]$ فرض کنید $x_i = \lceil i/k \rceil A_{i,\rightarrow}$.

ما ادعا می کنیم که $C = \{x_i : i \in [Tk/2]\}$ توسط $L(B_d,T)$ shatter می شود فرض کنید $I \subseteq [Tk/2]$.

سپس $I = I_1 \cup \ldots \cup I_{T/2}$، که I_t زیر مجموعه از $\{(t-1)k+1,\ldots,tk\}$ است. For each $t \in [T/2]$. t .

فرض کنید j ستون مربوطه A است (یعنی $A_{i,j} = 1$ if $(t-1)k + i \in I_t$.(i.e.,

بنابراین:

$$h(x) = \mathrm{sign}((h_{j_1,-1,1/2} + h_{j_1,1,3/2} + h_{j_2,-1,3/2} + h_{j_2,1,5/2} + \cdots$$
$$+h_{j_{T/2-1},-1,T/2-3/2} + h_{j_{T/2-1},1,T/2-1/2} + h_{j_{T/2},-'1,T/2-1/2})(x))$$

اگر باشد انگاه $h(x_i) = 1$ if $i \in I$ سپس خواهیم داشت $h \in L(B_d,T)$

11.1

فرض کنید S یک i.i.d باشد. فرض کنید h خروجی الگوریتم یادگیری توصیف شده باشد. میدانیم که مستقل از(S) , $L_D(h) = 1/2$ زیرا h یک تابع ثابت است.

فرض کنید $L_V(h)$ را محاسبه کنیم. فرض کنید که برابری S 1باشد. مقداری برابر $\{(x, y)\} \subseteq S$ را ثابت کنید. ما بین دو حالت تمایز قائل می شویم:

•برابری$S \setminus \{x\}$ 1است. نتیجه این است که $y = 0$. هنگامی که با استفاده از $S \setminus \{x\}$ آموزش می‌بینید، الگوریتم پیش‌بینی‌کننده ثابت $h(x) = 1$ را خروجی می‌دهد. بنابراین، leave one out با استفاده از این فولد 1 است.

•برابری $S \setminus \{x\}$ صفر است. نتیجه این است که $y = 1$. هنگامی که با استفاده از $S \setminus \{x\}$ آموزش می‌بینید، الگوریتم ثابت $h(x) = 0$ را خروجی می‌دهد. بنابراین، leave one out با استفاده از این فولد 1 است.

به‌طور میانگین بر روی دسته ها، تخمین خطایh برابر 1است. در نتیجه، تفاوت بین برآورد و خطای واقعی 0.5 است. موردی که در آن برابریS صفر است به طور مشابه درست است.

18.2

آنتروپی باینری را با H نشان می دهیم. الگوریتم اولین ریشه رامیگیرد با جستوجوی بیشترین Inforamtion gain و IG برای فیچراول

$$H\left(\frac{1}{2}\right)-\left(\frac{1}{2}H\left(\frac{1}{2}\right)+\frac{1}{2}H\left(\frac{1}{2}\right)\right)=0.$$

وبرای فیچر دوم و سوم :

$$H\left(\frac{1}{2}\right)-\left(\frac{3}{4}H\left(\frac{2}{3}\right)+\frac{1}{4}H(0)\right)\approx 0.22$$

بنابراین الگوریتم $x1 = 0$ را انتخاب به عنوان ریشه انتخاب میکند اما این بدان معنی است که سه مثال ,((1,1,1),1),((1,1,0),0) و ((1,0,0),1) از یک زیر درخت پایین می روند و مهم نیست که اکنون چه سؤالی بپرسیم، نمی‌توانیم هر سه نمونه را به طور کامل طبقه‌بندی کنیم. به عنوان مثال، اگر $x_2 = 0$ باشد (پس از آن باید یک پیش‌بینی ارائه کنیم)، یا ((1, 1, 0), 0) یا ((1, 1, 1), 1)لیبل اشتباه می‌شوند. بنابراین در هر صورت حداقل یک نمونه به اشتباه برچسب گذاری می شود. از آنجایی که در مجموعه آموزشی 4 نمونه داریم، نتیجه این است که خطای آموزش حداقل 1/4 است.

بنابراین یک درخت ۵ راسی خواهیم داشت که ریشه بررسی میکند $x_2=0$ است یا خیر بعد داده های رووس بعدی با سوال $x_3=0$ افراز میشوند .