درخت تقسیم ۱۸.۲:

این الگوریتم ابتدا گره‌ای را برمی‌دارد که بیشترین اطلاعات را می‌دهد. حال

information gain ویژگی اول برابر است با:

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

و برای ویژگی دوم و سوم داریم:

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

پس در نتیجه الگوریتم $x_1=0$ را انتخاب می‌کند و اما این به آن معنی است که درخت ما

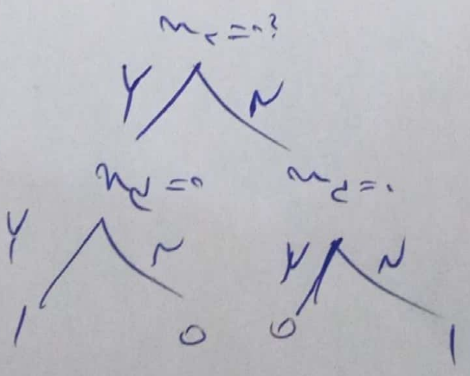
$((1,1,0,0,1)),((1,1,1,1)),((1,0,0,1,1))$ یک زیر درخت می‌شود حال برای مثال

$x_2=0$ یا $(1,1,1,1),(1,0,0,1,1)$ که خوب این اشتباه. پس توی هر حالتی حداقل

یکی از نمونه‌ها مون غلط پیش‌بینی می‌شود که چون چهارتا نمونه داریم پس

$\frac{1}{4}$ حداقل خطا است.

۲)



11.1. $h$ را الگوریتم که تعریف کردیم در سوال بگیریم و $C$ را فضای نمونه فرض کنیم می دانیم $L_D(h) = \frac{1}{2}$ می شود فرض کنیم ما خطای $L_V(h)$ را محاسبه کرده ایم از طرفی برابری را در $C$ فرض کرده ایم حال خوشه های $S_i = \{(x, y)\}$ را انتساب می کنیم حالا یا $y=1$ است یا $y=0$ است.

حال $S_1 < S_2$ باشد $y=0$ در نتیجه ما آموزش می دهیم مجموعه ای که $y=0$ است پس $h(x)=0$ پیش بینی می کند که $e_{est}$ آن برابر با یک است.

حال $S_1 > S_2$ برای $y=1$ باشد در نتیجه ما روی مجموعه ی $y=1$ آموزش می دهیم پس $h(x)=1$ را پیش بینی می کند و $e_{est}$ آن برابر با ۱ است.

پس در نتیجه ~~[illegible]~~ میانگین خطای تخمین خوشه ها برابر ۱ است و در نتیجه تفاضل خطای حقیقی از خطای تخمین برابر با $1 - \frac{1}{2} = \frac{1}{2}$ است.

11.2. مجموعه فرضیات های $H_1, H_2, \ldots, H_k$ را در نظر بگیریم

ensenble

۱۰.۱. $\epsilon,\delta\in(0,1)$ فرض کنیم و برداریم $k$ تکه های از سایز $m_{\mathcal{H}}(\epsilon/2,\delta_0)$

حالا $A$ را برای هر تکه قرار دهیم و بدست آوریم $\hat{h}_1,\dots,\hat{h}_k$. حالا توجه کنیم که احتمال

که $\min_{i\in[k]} L_D(\hat{h}_i) \le \min L_D(h)+\frac{\epsilon}{2}$ هست حداقل $1-\delta_0^k \ge 1-\frac{\delta}{2}$. حالا قرار دهیم

ERM رو کلاس $\hat{H}$ با مجموعه آموزشی از سایز تکه‌های حداقل $\left[\frac{2\log(4k/\delta)}{\epsilon^2}\right]$.

حالا با توجه به این که خروجی فرضیه ها با $\hat{h}$ است. حالا با احتمال حداقل $1-\frac{\delta}{2}$ داریم $L_D(\hat{h}) \le \min_{i\in[k]} L_D(h_i)+\epsilon/2$. حالا با قرار دادن union bound ما بدست می آوریم با احتمال حداقل $1-\delta$ داریم:

$$L_D(\hat{h}) \le \min_{i\in k} L_D(h_i)+\epsilon/2 \le \min_{h\in H} L_D(h)+\epsilon$$

۱۰.۴. (a. فرض کنیم $\mathcal{X}$ یک مجموعه‌ی متناهی با سایز $m$ باشد و فرض کنیم $B$ کلاس تمام تابع های از $X$ به $\{0,1\}$ باشد. حال $L(B,T)=B$ و باهم متناهی هستند.

در نتیجه داریم:

$$VCdim(B)=VCdim(L(B,T))=\log_2 2^m$$

b. —

c. با توجه به راهنمایی برای هر $i \in [\frac{Tk}{2}]$ فرض کنیم $x_i = [i/k] A_i$ ما نتیجه می‌گیریم که

مجموعه $C = \{x_i : i \in [\frac{Tk}{2}]\}$ به وسیله‌ی $L(B_d, T)$ shatter می‌شود.

فرض کنیم $I \subseteq [\frac{Tk}{2}]$ در نتیجه $I = I_1 \cup \dots \cup I_{T/2}$ که هر $I_t$ یک زیرمجموعه‌ی $\{(t-1)k+1, \dots, tk\}$

است. برای هر $t \in [T/2]$ فرض کنیم $j_t$ متناظر با ستون $A$ باشد ($A_{i,j} = 1$ iff $(t-1)k + i \in I_t$)

فرض کنیم

$$h(x) = \operatorname{sign}\Big(\big(h_{j_1, -1, \frac{1}{2}} + h_{j_1, 1, \frac{3}{2}} + h_{j_2, -1, \frac{3}{2}} + h_{j_2, \dots} + \dots + h_{j_{T/2-1}, -1, \frac{T}{2} - \frac{3}{2}} + h_{j_{T/2-1}, 1, \frac{T}{2} - \frac{1}{2}} + h_{j_{T/2}, -1, \frac{T}{2} - \frac{1}{2}}\big)(x)\Big).$$

در نتیجه $h(x_i) = 1$ اگر $i \in I$ باشد در نتیجه $h \in L(B_d, T)$.