Decision Tree:

(a) we want to talk about future 1 (we choose $x_1 = 0$)

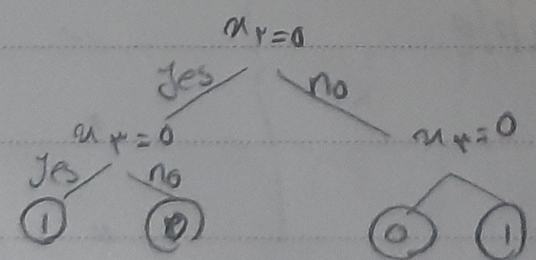
$$\Rightarrow H(1/۲) - (\frac{۳}{4} H(\frac{۲}{۳}) + \frac{1}{۴} H(0)) \approx ۰,۲۲$$

for future 2 & 3 we have

$$\Rightarrow H(1/۲) - (1/۲\, H(1/۲) + 1/۲\, H(1/۲)) = 0$$

so we understand that no matter what question will ask we won't able to classify all three example

(b) example tree:



Model selection & validation

① $S \sim$ iid sample , $h \sim$ learning algorithm $\Rightarrow L_D(h) = \frac{1}{Y}$

if we follow $y=0$ by training $S/\{u\}$ so the predictors is $h(x)=1$

by consider $S/\{u\}=1$.

if we consider $S/\{u\}=0$ so we get that $h(x)=0$

$\Rightarrow$ So the eutimate error of h is 1 & $|L_D(h) - L_S(h)| = \frac{1}{Y}$

and in the case of $S=0$ is analyzed analogously.

(Y)

boosting

① the probability that $\min_{i\in[k]} L_D(\hat{h}_i) \le \min L_D(h) + \varepsilon/۲$ is at least $1-\delta_0^k \ge 1-\delta/۲$

ERM over $\hat{H}$ with the training data with size of $\left\lceil \frac{۲\log(۴k/\delta)}{\varepsilon^۲} \right\rceil$

by applying the union bound with probablity at least $1-\delta$

$$L_D(\hat{h}) \le \min L_D(h_i) + \varepsilon/۲$$
$$\le \min L_D(h) + \varepsilon$$

---

② if $g(x)=1 \Rightarrow h(x)$ is $۰,۵$ or $۱,۵$

if $g(x)=-1 \Rightarrow h(x)$ is $-۰,۵$ or $-۱,۵$

---

③ $\varepsilon_t \Rightarrow$ $\sum_{i=1}^{m} D_i^{(t)} \exp(-w_t y_i h_t(x_i)) = \varepsilon_t \cdot \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} = \sqrt{\varepsilon_t(1-\varepsilon_t)}$

$$\Rightarrow \sum D_j \cdot \exp(-w_t y_j \cdot h_t(x_j)) = \varepsilon_t \cdot \sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} + (1-\varepsilon_t) \cdot \sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} = ۲\sqrt{\varepsilon_t(1-\varepsilon_t)}$$