Homework 3

Saturday, April 25, 2015 1:18 AM

P1 1. $Pr(Y=0) = 1/5 + 1/5 = 2/5 \quad Pr(Y=1) = 1/2 + 1/10 = 3/5$

$Pr(Z=0) = 3/10 + 3/20 = 9/20 \quad Pr(Z=1) = 2/5 + 3/20 = 11/20$

$Pr(X=0 \mid Y=0) = 1/5 \,/\, 2/5 = 1/2 \qquad Pr(X=1 \mid Y=0) = 1/5 \,/\, 2/5 = 1/2$

$Pr(X=0 \mid Z=0) = 3/20 \,/\, 9/20 = 1/3 \qquad Pr(X=1 \mid Z=0) = 3/10 \,/\, 9/20 = 2/3$

$Pr(X=0 \mid Y=1) = 1/10 \,/\, 3/5 = 1/6 \qquad Pr(X=1 \mid Y=1) = 1/2 \,/\, 3/5 = 5/6$

$Pr(X=0 \mid Z=1) = 3/20 \,/\, 11/20 = 3/11 \qquad Pr(X=1 \mid Z=1) = 2/5 \,/\, 11/20 = 8/11$

2. $H(X \mid Y=0) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = \log 2$

$H(X \mid Y=1) = -\frac{1}{6}\log\frac{1}{6} - \frac{5}{6}\log\frac{5}{6} = .196$

$H(X \mid Y) = \frac{2}{5}(\log 2) + \frac{3}{5}(.196) = .238$

$H(X \mid Z=0) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = .276$

$H(X \mid Z=1) = -\frac{3}{11}\log\frac{3}{11} - \frac{8}{11}\log\frac{8}{11} = .254$

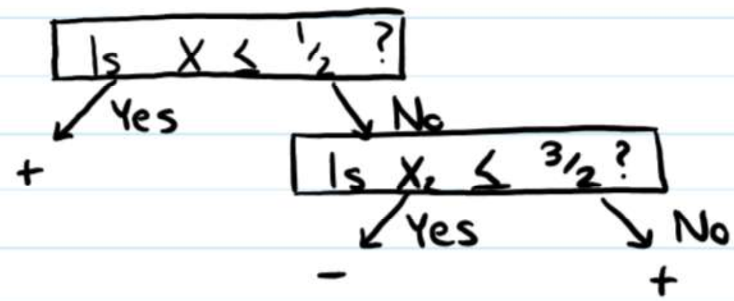
$H(X \mid Z) = \frac{9}{20}(.276) + \frac{11}{20}(.254) = .264$

3. Based on the conditional entropies, Gene A is more informative since $H(X|Y)$ is smaller, you are more certain than $H(X|Z)$.

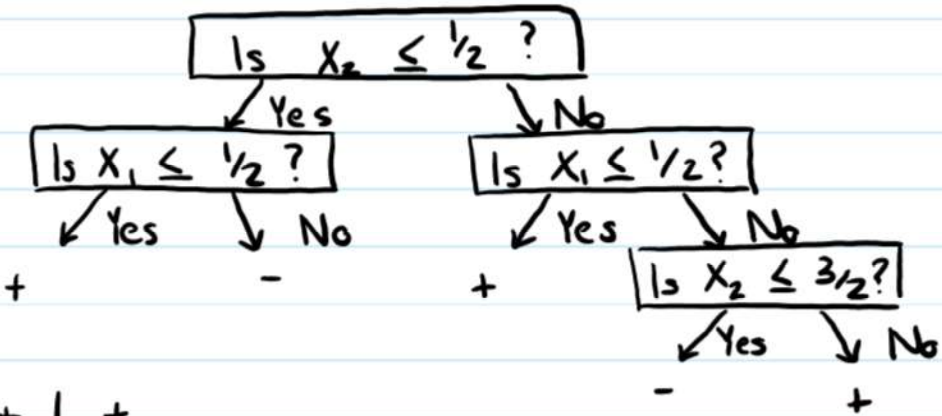
P2 1. False, we can have two trees where one is an extended version of the other as shown in lecture:

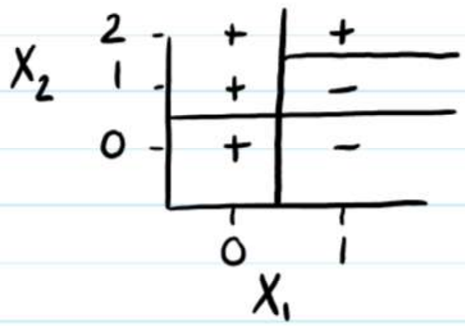
Data $((0,2),+)\ ((1,2),+)\ ((0,1),+)$
$((1,1),-)\ ((0,0),+)\ ((1,0),-)$

Tree 1



Tree 2





Both predicts all correctly.

2. True, because they both got no errors on the same training data set that means they both predicted the same labels for every feature. By definition that makes them equal since for all x feature vectors $\in X$, $T(x) = T'(x)$.

P3 1. $H(X) = -\frac{1}{2}\log\frac{1}{2} - (\frac{1}{2}\cdot\frac{1}{2})\log(\frac{1}{2}\cdot\frac{1}{2}) - (\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2})\log(\frac{1}{8})\ldots$

first flip — second flip — third flip

$$= -(\tfrac{1}{2})^n \log(\tfrac{1}{2})^n = -(\tfrac{1}{2})^n\, n\log(\tfrac{1}{2})$$
$$= -n(\tfrac{1}{2})^n \log(\tfrac{1}{2}) = \left[-\tfrac{1}{2} / (1-\tfrac{1}{2})^2\right]\log(\tfrac{1}{2})$$
$$= -2\log(\tfrac{1}{2})$$

2. $Z = X$ w.p. α, Y w.p. $(1-\alpha)$

$= x_1$ w.p. αp_1, x_2 w.p. $\alpha p_2 \ldots x_m$ w.p αp_m

x_{m+1} w.p. $(1-\alpha)p_{m+1}, \ldots x_{m+n}$ w.p. $(1-\alpha)p_{m+n}$

$$H(Z) = -\alpha p_1 \log(\alpha p_1) \ldots - \alpha p_m \log(\alpha p_m)$$
$$-(1-\alpha)p_{m+1}\log((1-\alpha)p_{m+1})\ldots$$
$$-(1-\alpha)p_{m+n}\log((1-\alpha)p_{m+n})$$

$$= -\alpha p_1 \log\alpha - \alpha p_1 \log p_1 \dots -\alpha p_m \log\alpha - \alpha p_m \log p_m$$
$$- (1-\alpha) p_{m+1} \log(1-\alpha) - (1-\alpha) p_{m+1} \log(p_{m+1}) \dots$$
$$- (1-\alpha) p_{m+n} \log(1-\alpha) - (1-\alpha) p_{m+n} \log(p_{m+n})$$

$$= \alpha H(x) (-p_1 \log\alpha \dots - p_m \log\alpha) +$$
$$(1-\alpha) H(X) (-p_{m+1} \log(1-\alpha) \dots$$
$$- p_{m+n} \log(1-\alpha))$$

P4 1. Next Page

2. Test error = .36