HW #2

1) (a) $P(A|B) = \dfrac{P(B|A)\,P(A)}{P(B)}$

prove for events A & B

$P(A|B)$ is the probability of A marginalized over B

$$P(B|A)P(A) = P(A,B) = P(A|B)\,P(B)$$

$$P(\cancel{B}|A)P(A) = P(A|B)\,P(B)$$

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

(b) rewrite $P(A,B,C)$ as a product of several conditional & 1 unconditional probability

$$P(A,B,C) = P(A,B|C)\,P(C)$$
$$= P(A|B,C)\,P(B|C)\,P(C)$$

(c) $x = \begin{cases} 1 & \text{if } A \\ \phi & \text{else} \end{cases}$ show $E[x] = P(A)$

$$E(x) = \sum_{x \in X} x_i P(x) = (x=1)\,P(x=1) + (x=\phi)\,P(x=\phi)$$
$$= 1\,P(x=1) + \phi$$
$$= 1\,P(A)$$

(d) Two variables are independent iff
$$P(A \cap B) = P(A)\,P(B) \Rightarrow P(A|B) = P(A)$$

$$P(X=0) = \frac{8}{45} + \frac{4}{45} + \frac{1}{15} + \frac{1}{10}$$

$$P(X=Z) = 1 - P(x=\phi)$$

(i) $P(x=\phi|Y) = \cancel{1}\ \ \frac{\phi}{15}$ $\qquad P(x=\phi|Y) = \frac{1}{10}\ \frac{+8}{45} \neq P(x=\phi)$

$P(x=\phi|Y=1) = \frac{1}{10} + \frac{8}{45}$ $\qquad$ No

(ii) $P(X \cap Y|Z) = P(X \cap Z)\,P(Y \cap Z)$

② $\hat{\theta}^{MLE} = \underset{\theta}{\arg\max}\,(L(\hat{\theta}))$

write a hypothetical function $L(\vec{\theta})$, dependent on

$x_1, x_2, \dots x_n$ & the hypothetical $\hat{\theta}$

$p(x_i = 1) = \theta$ $\quad$ $p(x_i = \emptyset) = 1 - \theta$

we want to estimate θ from observation

$n = 10$ $\quad$ s.t. 1 4 $\emptyset$

\# of events

we would estimate $L(\theta) = \frac{6}{10} = \binom{n}{r}(\hat{\theta})^r(1-\hat{\theta})^{n-r}$

\# of successes

since we are maximizing $L(\theta)$ the binomial $\binom{n}{r}$ is not needed

(b) plot attached

(d) $n=5$ $\quad$ $s=3$ $\quad$ as n increases & $\frac{s}{n} = $ const

$n=100$ $\quad$ $s=60$ $\quad$ the peak becomes more broad but

$n=10$ $\quad$ $s=5$ $\quad$ does not move.

(c) $L(\theta) \propto \theta^r(1-\hat{\theta})^{n-r}$

$$L'(\theta) = \emptyset = r\theta^{r-1}(1-\hat{\theta})^{n-r} \cancel{=} (n-r)(1-\hat{\theta})^{n-r-1}\theta^r$$

$$= r\frac{\cancel{\theta^r}}{\theta}\cancel{(1-\hat{\theta})^{n-r}} = \frac{(n-r)\cancel{(1-\hat{\theta})^{n-r}}\cancel{\theta^r}}{(1-\hat{\theta})}$$

$$= \frac{r}{\theta} = \frac{n-r}{1-\theta} = r\cancel{-r}\theta = n\theta - r\cancel{\theta}$$

$$\boxed{\theta = \frac{r}{n}}$$

(e) $\hat{\theta}^{MLE}$ does not change

$$\hat{\theta}^{MAP} = \underset{\hat{\theta}}{\arg\max}\{L(\theta)\, p\, \hat{\theta}\}$$

$$p(\hat{\theta}) = \frac{\hat{\theta}^2(1-\hat{\theta})^2}{0.0\overline{33}}$$

(g) $$\frac{d}{d\theta}\left(L(\theta)p(\hat{\theta})\right)\Big|_{\hat{\theta}^{MAP}} = \phi$$

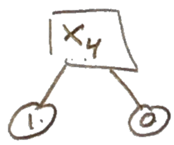
$$\hat{\theta} = \frac{2+r}{4+n}$$

(f) as $n \to \infty$ map $\to$ MLE so the prior becomes less important

map has advantage of using more info, it biases data a bit but if appropriate priors are choosen this is not an issue.
the peak of MAP will ~~get~~ approach MLE as $n \to \infty$, it removes possibility of low statistic events spoiling the model.

(4)

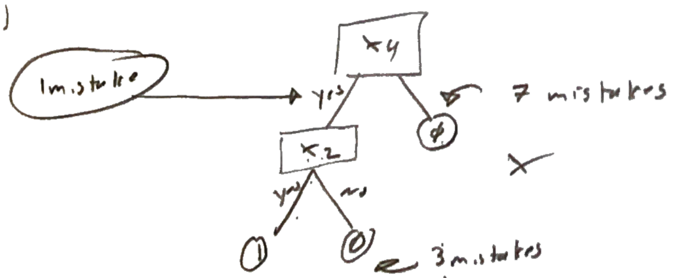
(a) the one leaf tree takes the data and tries to basically guess
we can use say

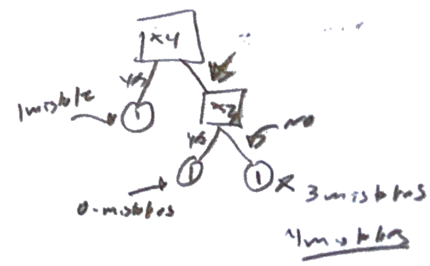


it would get 8 out of 16 correct

<u>8 mistakes</u>

(b)





<u>less mistakes</u>

(c) $$-\text{Entropy} = +\sum_{i=1}^{n} P(x=i)\, \log_2(P(x=i))$$
possible values of x

$$= \frac{14}{16}\log_2\left(\frac{14}{16}\right) + \frac{2}{16}\log_2\left(\frac{2}{16}\right)$$

$$\log_2\left(\left(\frac{14}{16}\right)^{\frac{14}{16}} \cdot \left(\frac{2}{16}\right)^{2/16}\right)$$

Entropy ≈ 0.37677

(d) $$H(x|Y) = -\sum_{i}^{n} P(x=i|Y)\log_2(P(x=i|Y))$$

$$-H(Y) = P(x_4=1)\log 1) + P(x_3=1|x_4=\phi)\log + P(x_3=\phi|x_4=\phi)\log(\,)$$

increased the entropy