PROBLEM 1

40/40

1.1> For a boolean variable

$$E(s_k \mid b_k, a_k, h_k n_k, \theta)$$

$$= P(s_k=1 \mid b_k, a_k, h_k, n_k, \theta)\cdot 1 + P(s_k=0 \mid b_k, a_k, h_k, n_k, \theta)\cdot 0$$

$$= P(s_k=1 \mid b_k, a_k, h_k, n_k, \theta).$$

Current value of $\theta_{s|ij} = \theta' = P(s_k=1 \mid b_k=i, a_k=j)$.

To calculate the maximum value of $\theta_{s|ij}$ we write

$$\frac{d}{d\theta'_{s|ij}} \sum_{k=1}^{K} E_{P(s_k|b_k a_k h_k n_k \theta)} \left[\log P(b_k, a_k, h_k, n_k, s_k \mid \theta') \right] = 0$$

$$\cancel{\frac{d}{d\theta_{s|ij}} \sum_{k=1}^{K} P(s_k=1, b_k a_k h_k n_k \theta) \log}$$

$$\frac{d}{d\theta'_{s|ij}} \sum_{k=1}^{K} E_{P(s_k|b_k a_k h_k n_k \theta)} \Big[\log P_{\theta'}(b_k) + \log P_{\theta'}(a_k) + \log P_{\theta'}(s_k \mid b_k a_k) + \log P_{\theta'}(h_k \mid s_k) + \log P_{\theta'}(n_k \mid s_k) \Big] = 0$$

$$\cancel{\frac{d}{d\theta'} \sum_{k=1}^{K} P(s_k=1 \; b_k a_k h_k n_k \theta)}$$

Since only the term $P_\theta(s_k \mid b_k a_k)$ depend on θ', hence this expression reduces to,

$$\frac{d}{d\theta'} \sum_{k=1}^{K} E_{P(s_k|b_k a_k h_k n_k \theta)} \left[\log P_{\theta'}(s_k \mid b_k a_k) \right] \delta(b_k=i, a_k=j) = 0$$

$$\frac{d}{d\theta'}\Big[\sum_{k=1}^{K} \{ P(s_k=1 / b_k\, a_k\, h_k\, n_k\, \theta).\ \log P_{\theta'}(s_k=1 / b_k=i, a_k=j)$$

$$+ \sum P(s_k=0 / b_k\, a_k\, h_k\, n_k\, \theta) \log P_{\theta'}(s_k=0 / b_k=i, a_k=j)\}\Big]\ \delta(b_k=i, a_k=j) = 0$$

now,
$$P(s_k=1 / b_k\, a_k\, n_k\, h_k\, \theta) = E(s_k)$$
$$P(s_k=0 / b_k\, a_k\, n_k\, h_k\, \theta) = 1 - E(s_k).$$

$$P_{\theta'}(s_k=1 / b_k=i, a_k=j) = \theta'$$
$$P_{\theta'}(s_k=0 / b_k=i, a_k=j) = 1-\theta'.$$

$$\therefore \frac{d}{d\theta'}\sum_{k=1}^{K}\Big[\delta(b_k=i, a_k=j)\, E(s_k) \log\theta' + \delta(b_k=i, a_k=j)\,(1-E(s_k))\log(1-\theta')\Big] = 0$$

$$\sum_{k=1}^{K}\Big[\delta(b_k=i, a_k=j)\, E(s_k)\frac{1}{\theta'} - \frac{\delta(b_k=i, a_k=j)(1-E(s_k))}{(1-\theta')}\Big] = 0$$

$$\sum_{k=1}^{K}\delta(b_k=i, a_k=j)\, E(s_k)\,(1-\theta') = \sum_{k=1}^{K}\delta(b_k=i, a_k=j)\,(1-E(s_k))\,\theta'$$

$$\sum_{k=1}^{K}\delta(b_k=i, a_k=j)\, E(s_k) - \cancel{\theta'\sum_{k=1}^{K} E(s_k)\,\delta(b_k=i, a_k=j)}$$
$$= \theta'\sum_{k=1}^{K}\delta(b_k=i, a_k=j) - \cancel{\theta'\sum_{k=1}^{K} E(s_k)\,\delta(b_k=i, a_k=j)}$$

$$\Rightarrow \theta' = \frac{\sum_{k=1}^{K}\delta(b_k=i, a_k=j)\, E(s_k)}{\sum_{k=1}^{K}\delta(b_k=i, a_k=j)}.$$

1.2)

A) ~~P(f)x~~

$$P(f, a, s, h, n) = P(f)\,P(a)\,P(s/fa)\,P(h/s)\,P(n/s)$$

No. of parameters to estimate for each term

$P(f) \rightarrow P(f=1) \rightarrow 1$

$P(a) \rightarrow P(a=1) \rightarrow 1$

$P(s/fa) \rightarrow P(s=1/f=1, a=1)$

$P(s=1/f=0, a=1)$

$P(s=1/f=1, a=0)$

$P(s=1/f=0, a=0) \rightarrow 4$

$P(h/s) \rightarrow P(h=1/s=1)$

$P(h=1/s=0) \rightarrow 2$

$P(n/s) \rightarrow P(n=1/s=1)$

$P(n=1/s=0) \rightarrow 2$

Total no. of parameters to estimate:

$1+1+4+2+2 = 10$

B) For example 1, we calculate

$P(h=1/f=1, a=1, S=1, N=1)$

$$= \frac{P(h=1, f=1, s=1, n=1, a=1)}{P(h=1, f=1, s=1, n=1, a=1) + P(h=0, f=1, s=1, n=1, a=1)}$$

$$= \frac{p(f=1)\,p(a=1)\,p(s=1|a=1,f=1)\,p(h=1|s=1)\,p(n=1|s=1)}{p(f=1)\,p(a=1)\,p(s=1|a=1,f=1)\,p(h=1|s=1)\,p(n=1|s=1) + p(f=1)\,p(a=1)\,p(s=1|a=1,f=1)\,p(h=0|s=1)\,p(n=1|s=1)}$$

$$= \frac{0.5^5}{2\times 0.5^5} = 0.5$$

For example 8,

$$p(n=1|f=1,a=1,s=1,h=1)$$

$$= \frac{p(n=1,f=1,a=1,s=1,h=1)}{p(n=1,f=1,a=1,s=1,h=1) + p(n=0,f=1,a=1,s=1,h=1)}$$

$$= \frac{p(f=1)\,p(a=1)\,p(s=1|a=1,f=1)\,p(h=1|s=1)\,p(n=1|s=1)}{p(f=1)\,p(a=1)\,p(s=1|a=1,f=1)\,p(h=1|s=1)\,p(n=1|s=1) + p(f=1)\,p(a=1)\,p(s=1|a=1,f=1)\,p(h=1|s=1)\,p(n=0|s=1)}$$

$$= \frac{0.5^5}{2\times 0.5^5} = 0.5.$$

c) $P(f=1) = \dfrac{\sum_k \delta(f=1)}{\sum_k \delta(1)} = \dfrac{4}{8} = 0.5$

$P(a=1) = \dfrac{\sum \delta(a=1)}{\sum_k \delta(1)} = \dfrac{4}{8} = 0.5$

$P(s=1 | f=1, a=1) = \dfrac{\sum \delta(s=1, f=1, a=1)}{\sum (f=1, a=1)} = \dfrac{3}{3} = 1.0$

$P(s=1 | f=1, a=0) = \dfrac{\sum \delta(s=1, f=1, a=0)}{\sum (f=1, a=0)} = \dfrac{1}{1} = 1.0$

$P(s=1 | f=0, a=1) = \dfrac{\sum \delta(s=1, f=0, a=1)}{\sum (f=0, a=1)} = \dfrac{1}{1} = 1.0$

$P(s=1 | f=0, a=0) = \dfrac{\sum \delta(s=1, f=0, a=1)}{\sum (f=0, a=0)} = \dfrac{0}{3} = 0.0$

$P(h=1 | s=1) = \dfrac{\sum_k \delta(h=1, s=1)\, E(h_k=1)}{\sum \delta(s=1)}$

$= \dfrac{4+0.5}{5} = \dfrac{9}{10} = 0.9$

$P(h=1 | s=0) = \dfrac{1}{3} = 0.33$

$P(n=1 | s=1) = \dfrac{\sum_k \delta(n=1, s=1)\, E(n_k=1)}{\sum (s=1)} = \dfrac{3+0.5}{5} = 0.7$

$P(n=1 | s=0) = \dfrac{\sum \delta(n=1, s=0)}{\sum \delta(s=0)} = \dfrac{1}{3} = 0.33$

o> Using the updated values.

$$P(h=1 \mid f=1, a=1, s=1, n=1)$$

$$= \frac{0.5 \times 0.5 \times 1 \times 0.9 \times 0.7}{0.5 \times 0.5 \times 1 \times 0.9 \times 0.7 + 0.5 \times 0.5 \times 1 \times 0.1 \times 0.7}$$

$$= \frac{\cancel{0.5} \times \cancel{0.5} \times 1 \times 0.9 \times \cancel{0.7}}{\cancel{0.5} \times \cancel{0.5} \times 1 \times 1 \times \cancel{0.7}}$$

$$= 0.9$$

$$P(n=1 \mid f=1, a=1, s=1, h=1)$$

$$= \frac{0.5 \times 0.5 \times 1 \times 0.9 \times 0.7}{0.5 \times 0.5 \times 1 \times 0.9 \times 0.7 + 0.5 \times 0.5 \times 1 \times 0.9 \times 0.3}$$

$$= \frac{0.5 \times 0.5 \times 1 \times 0.9 \times 0.7}{0.5 \times 0.5 \times 1 \times 0.9 \times 1} = 0.7$$

PROBLEM 2

54/60

1>

(a) We might prefer Decision trees when the data is not conditionally independent, or when we want a non-linear decision boundary, or when we want to generate new examples (i.e. calculate $P(x, y)$.

(b) We might prefer Logistic Regression over Naive Bayes because Logistic regression does not make assumptions about conditional independence and the variances of distributions of $P(X_i/Y=y_k)$ being independent of k.

(c) We minimize sum of squared error because this is mathematically equivalent to a MCLE estimate, or maximizing $P(data/w)$ when we assume the value of y to be normally distributed around $\vec{W}\vec{x}$.

2> Gaussian Naive Bayes: Yes.

From Conditional independence,

$P(X_1, X_2, X_3 / Y) = P(X_1/Y)\, P(X_2/Y)\, P(X_3/Y)$.

$P(X_1, X_2, X_3, Y) = P(X_1, X_2, X_3 / Y)\, P(Y)$

$= P(X_1/Y)\, P(X_2/Y)\, P(X_3/Y)\, P(Y)$

From training examples, we can calculate MLE or MAP estimates of $P(X_i/Y)$. ~~we~~ We can calculate

$$P(Y=y_k) = \frac{\delta(y=y_k)}{\text{total no. of examples}}$$

(b) Logistic Regression: No.
Logistic regression only calculates
$P(Y|X_1, X_2, X_3)$
It does not give information about $P(Y)$
or $P(X_1, X_2, X_3|Y)$

(c) Linear Regression: No.
Linear Regression also calculates only $P(Y|X)$
It does not tell $P(X|Y)$ and $P(Y)$.

3)
(a) True.
The no. of real examples constitute the exponent of $P(X_i|Y)$. This becomes very large as data points tend to infinity. Hence the value of exponents from priors become immaterial, as they are much smaller. So, MLE and MAP estimates converge

0/4

(b) False.
In an extreme case, at each level, there is one example for $(Y=y_k)$ and all other examples for $(Y \neq y_k)$. If we keep going down the tree, for n examples, after level n, we will have only 1 example, and tree cannot grow further, and must be terminated.

(c) False.

Decision tree has a non linear decision boundary which is a superset of the linear boundary generated by Logistic regression. So, it is possible that they have the same linear boundary for a given set of train examples.

Example??

2/4

4) $P(X,Y|Z) = P(X,Y,Z)/P(Z)$

$$= \frac{P(X|Y,Z)\ P(Y,Z)}{P(Z)}$$

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

It is given that

$$P(X|Y,Z)\ \cancel{P(Y|Z)} = P(X|Z)\ \cancel{P(Y|Z)}$$

$$P(X|Y,Z) = P(X,Z)$$

Hence proved that X is conditionally independent of Y, given Z.

5) (a) We can write

$$P(X_1, X_2, X_3, X_4) = P(X_1)\ P(X_2|X_1)\ P(X_3|X_1)\ P(X_4|X_2,X_3).$$

$P(X_1) \rightarrow 1$

$P(X_2|X_1) \rightarrow 2$

$P(X_3|X_1) \rightarrow 2$

$P(X_4|X_2,X_3) \rightarrow 4$

Total parameters = 9

(b) $P(x_1=1, x_2=0, x_3=1, x_4=0)$

$$= P(x_1=1)\, P(x_2=0 \mid x_1=1)\, P(x_3=1 \mid x_1=1)\, P(x_4=0 \mid x_2=0, x_3=1)$$

(c) $P(x_1=1, x_4=0)$

$$= P(x_1=1, x_2=0, x_3=0, x_4=0)$$
$$+ P(x_1=1, x_2=1, x_3=0, x_4=0)$$
$$+ P(x_1=1, x_2=01, x_3=1, x_4=0)$$
$$+ P(x_1=1, x_2=0, x_3=1, x_4=0)$$

$$= P(x_1=1)\Big[P(x_2=0 \mid x_1=1)\, P(x_3=0 \mid x_1=1)\, P(x_4=0 \mid x_2=0, x_3=0)$$
$$+ P(x_2=1 \mid x_1=1)\, P(x_3=0 \mid x_1=1)\, P(x_4=0 \mid x_2=1, x_3=0)$$
$$+ P(x_2=1 \mid x_1=1)\, P(x_3=1 \mid x_1=1)\, P(x_4=0 \mid x_2=1, x_3=1)$$
$$+ P(x_2=0 \mid x_1=1)\, P(x_3=1 \mid x_1=1)\, P(x_4=0 \mid x_2=0, x_3=1)\Big]$$

(d) $P(x_2=1 \mid x_3=0) = \dfrac{P(x_2=1, x_3=0)}{P(x_3=0)}$

$$= \frac{P(x_2=1, x_3=0)}{P(x_2=1, x_3=0) + P(x_2=0, x_3=0)}$$

$P(x_2=1, x_3=0)$

$= \cancel{P(x_1=1)\, P(x_2=1 \mid x_1=1)\, P(x_3=0 \mid x_1=1)\, P(x_4=0}$

$= P(x_1=0, x_2=1, x_3=0, x_4=0)$

$+ P(x_1=1, x_2=1, x_3=0, x_4=0)$

$+ P(x_1=0, x_2=1, x_3=0, x_4=1)$

$+ P(x_1=1, x_2=1, x_3\,0, x_4=1)$

$= P(x_1=1) [P(x_2=1 \mid x_1=1)\, P(x_3=0 \mid x_1=1) \,[P(x_4=0 \mid x_2=1, x_3=0)$

$+ P(x_4=1 \mid x_2=1, x_3=0)]$

$+ P(x_1=0)\, P(x_2=1 \mid x_1=0)\, P(x_3=0 \mid x_1=0)$

$[P(x_4=0 \mid x_2=1, x_3=0) + P(x_4=1 \mid x_2=1, x_3=$

$= P(x_1=1)\, P(x_2=1 \mid x_1=1)\, P(x_3=0 \mid x_1=1)$

$+ P(x_1=0)\, P(x_2=1 \mid x_1=0)\, P(x_3=0 \mid x_1=0)$

$P(x_2=0, x_3=0)$

$= P(x_1=1)\, P(x_2=0 \mid x_1=1)\, P(x_3=0 \mid x_1=1)$

$+ P(x_1=0)\, P(x_2=0 \mid x_1=0)\, P(x_2=0 \mid x_1=0)$.