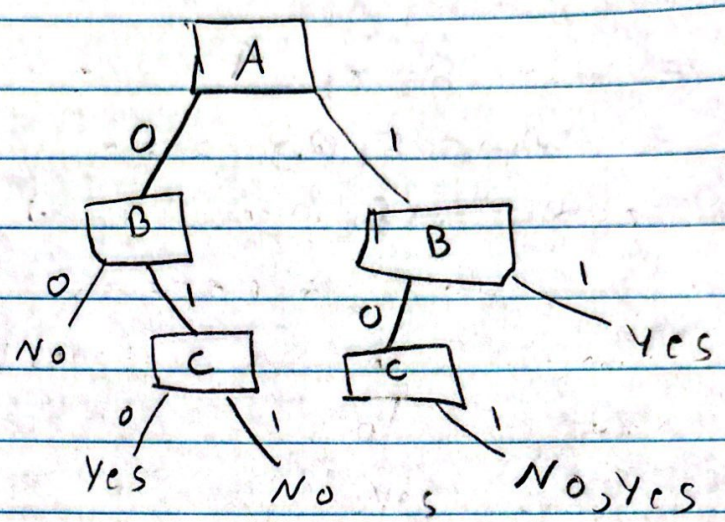
★ Spring 2015 - midterm

Q1]

a)



Not Possible, becasc 101 ~~is~~ has duplicate clases

b] Entropy A $= \frac{1}{2}\left[-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right] +$

$\frac{1}{2}\left[-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right]$

Entropy B $= \frac{1}{2}\left[-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{1}{3}\right)$

$+ \frac{1}{2}\left[-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right]$

Entropy C $= \frac{1}{2}\left[-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right]$

$+ \frac{1}{2}\left[\cdots\right]$

All have same Entropy so same Information gain.

c) True. But the number of ~~nodes~~ leaves can be equal to the size of dataset. If a data is linearly seperable that means we can always find thresholds that can seperate each individual data point i.e leaf node.

d) Not necessary, we can build simple & complex tree which can both represent the same concept

for example

$x_1 \wedge -x_2 \wedge x_3$

↳ we can either spilt it to create tree with 3 nodes

or we can create a truth table & form a tree with 2^3 leaves i.e. one leaf for each value in the truth table.

Q2]

True, use w0 = k-1, all other weights = 1 ,
if f(x) >0 then True else false

a] True / No.

if n = 2 & k = 1 it becomes a XOR gate. XOR gate cannot be represented by linear function. Because we e.g. k out of n true can be any random k our of n which we dont know, so we cannot find a weight setting which will match the input which are positive or negative.

b) let the linear function with n-1 features be

$$\sum_{i=1}^{n-1} w_i x_i > 0$$

Answer is NO

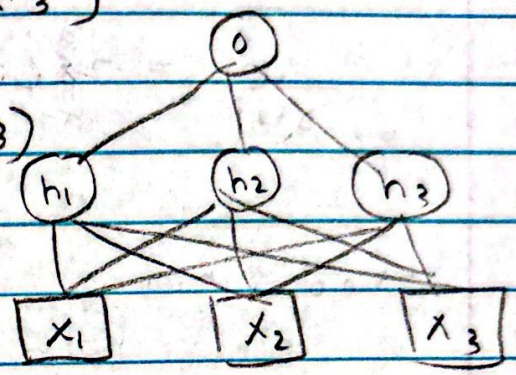
Because skipping one feature is loss of information & it wont represent same function.

what if left out feature was the most valuable?

Q3] ☆ Assuming input to be $-1, +1$ not $0, 1$

a] Truth table can be represented as

$(x_1 \wedge -x_2 \wedge x_3) \vee (-x_1 \wedge x_2 \wedge -x_3] \vee$
$(-x_1 \wedge -x_2 \vee x_3)$



$h_1 = (x_1 \wedge -x_2 \wedge -x_3)$

$w_0 = -2.5$

$w_1 = 1, w_2 = -1, w_3 = -1$

$h_2 = (-x_1 \wedge x_2 \wedge -x_3]$

$w_0 = -2.5$

$w_1 = -1, w_2 = 1, w_3 = -1$

$h_3 = (-x_1 \wedge -x_2 \wedge x_3)$

$w_0 = -2.5$

$w_1 = -1, w_2 = -1, w_3 = 1$

$o = h_1 \vee h_2 \vee h_3$

$w_0 = 2.5$

$w_1 = 1, w_2 = 1, w_3 = 1$

h_1, h_2, h_3 are threshold functions
if $f(x) > 0$, 1 else -1

o is threshold function
if $f(x) > 0$ 10, else -5

b) $o = w_0 + \sum_{i=1}^{n} w_i (x_i + x_i^2 + x_i^3)$

$$\text{Loss} = \sum_{k=1}^{d} (y - o)^2$$

$$\frac{\partial L}{\partial w_i} = -2 \sum_{k=1}^{d} (y - o)(x_i + x_i^2 + x_i^3)$$

Algorithm

Loop until convergence

for i in feature n

$$w_i = w_i - \sum_{k=1}^{d} 2\alpha (y - o)(x_{ik} + x_{ik}^2 + x_{ik}^3)$$

end for

end loop

where o is $= w_0 + \sum_{i=1}^{n} w_i (x_i + x_i^2 + x_i^3)$

~~where~~ α is learning rate

Q-5) a] P of k failure & 1 success

$= p^k (1-p)$

b) Likelihood,

$$L(D|\theta) = \prod_{k=1}^{m} p^y (1-p)^{1-y}$$

$y=1$ = succes dat, $y=0$ = failure

Log Likelihood,

$$\log L(D|\theta) = \sum_{k=1}^{m} \left[y \ln p + (1-y) \ln(1-p) \right]$$

MLE,

$$\Rightarrow \frac{\partial L}{\partial p} = 0$$

$$\Rightarrow \sum_{k=1}^{m} \left[\frac{y}{p} - \frac{(1-y)}{1-p} \right] = 0$$

$$\Rightarrow \sum_{k=1}^{m} \left[\frac{y - \cancel{yp} - p + \cancel{yp}}{p(1-p)} \right] = 0$$

$$\Rightarrow \sum_{k=1}^{m} y = mp$$

$$\Rightarrow p = \frac{\sum_{k=1}^{m} y}{m}$$

or $\boxed{p = \sum_{i=1}^{m} \frac{k_i \{y = \text{dat}\}}{m}}$

Q6]

a] $$w_0 = \ln\left(\frac{\theta}{1-\theta}\right) - \left(\sum_{i=1}^{2} \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2}\right)$$

$$\theta = \frac{2}{7}$$

For x_1,

$$\mu_{11} = \frac{0+2}{2} = 1, \quad \mu_{10} = \frac{1+2+3+4+5}{5} = 3$$

$$\sigma_1 = \sum_{i=1}^{7} \frac{(x_{1i} - \mu_{11})^2}{7}$$

For x_2,

$$\mu_{21} = \frac{9+5}{2} = 7, \quad \mu_{20} = \frac{3+4+5+6+7}{3}$$

$$\sigma_2 = \sum_{i=1}^{7} \frac{(x_{1i} - \mu_{21})^2}{7}$$

$$w_1 = \frac{\mu_{11} - \mu_{10}}{\sigma_1^2}$$

$$w_2 = \frac{\mu_{21} - \mu_{20}}{\sigma_2^2}$$

☆ As prof said we don't need to calculate answer, just write the expression

b] Just same as, except $\sigma_1 = \sigma_2$

Q8] Refer to Fall 2020 midterm 1 solutions Q 6]