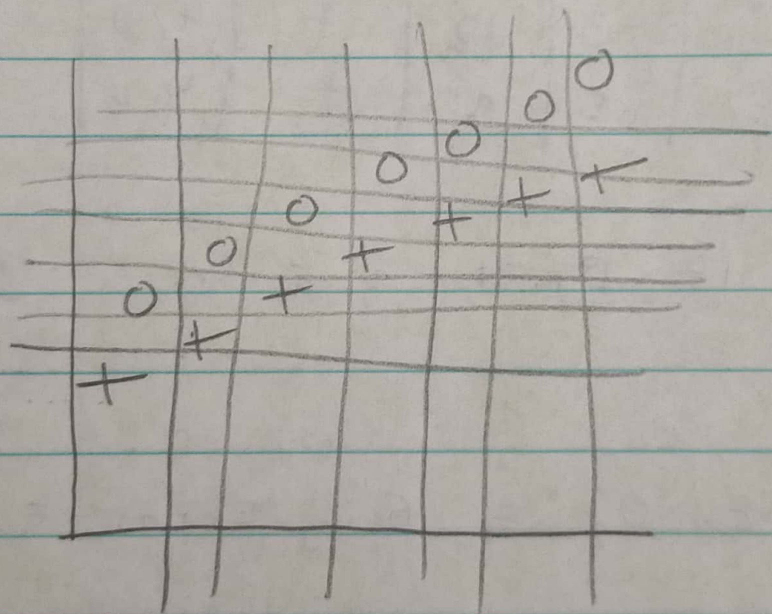
☆ Spring 2018 midterm

Q1]

a] O(n)

-> Because we have to draw O(n) boundaries to seperate each dataset

I.e. if we draw a verticle line & a horizontal line for each dataset, it is only then we can seperate + & -ve



Can we do better than this?

no because we need two lines to isolate each +ve example.

b] If we use x_1,
x_1 has lowest Entropy of 0
ie. highest information gain

c] No. because x_1 seems to be a index column which does not generalize to the entire data. This is because Entropy calculations favour dataset with more possible values. We can used normalized Gaion also known as Gain ratio

$$\text{Gain Ratio} = \frac{\text{Gain}(x_i)}{-\sum \frac{|S[x_j = i]|}{|S|} \log \frac{|S[x_j = i]|}{S}}$$

Q2] a]

$$o_1 = w_1 x_1 + w_3 x_2 + \text{bias}$$
$$o_2 = w_2 x_1 + w_4 x_2 + \text{bias}$$
$$o_3 = w_5 o_1 + w_7 o_2 + \text{bias}$$
$$o_4 = w_6 o_1 + w_8 o_2 + \text{bias}$$

Scanned with CamScanner

b] let y_3 be expected output of O_3
y_4 for O_4

$$\text{Error} = \sum_{i=3}^{4} (y_i - o_i)^2$$

$$\frac{1}{2} \frac{\partial E}{\partial w_5} = -\underbrace{(y_3 - o_3)(o_3)(1-o_3)}_{\delta_3} O_1$$

$$\underline{\delta_3 = (y_3 - o_3)(o_3)(1-o_3)}$$

$$\frac{1}{2} \frac{\partial E}{\partial w_6} = -\underbrace{(y_4 - o_4)(o_4)(1-o_4)}_{\delta_4} O_1$$

$$\underline{\delta_4 = (y_4 - o_4)(o_4)(1-o_4)}$$

$$\frac{1}{2} \frac{\partial E}{\partial w_1} = -\underbrace{\sum_{i=3}^{4} (y_i - o_i)(o_i)(1-o_i)(o_1)(1-o_1)}_{\delta_1} x_1$$

$$\underline{\delta_1 = \sum_{i=3}^{4} (y_i - o_i)(o_i)(1-o_i)(o_1)(1-o_1)}$$

$$\frac{1}{2} \frac{\partial E}{\partial w_2} = -\underbrace{\sum_{i=3}^{4} (y_i - o_i)(o_i)(1-o_i)(o_2)(1-o_2)}_{\delta_2} x_1$$

$$\underline{\delta_2 = \sum_{i=3}^{4} (y_i - o_i)(o_i)(1-o_i)(o_2)(1-o_2)}$$

c]

$w_1 = w_1 - \delta_1 x_1$

$w_2 = w_2 - \delta_2 x_1$

$w_3 = w_3 - \delta_1 x_2$

$w_4 = w_4 - \delta_2 x_2$

$w_5 = w_5 - \delta_3 o_1$

$w_6 = w_6 - \delta_4 o_1$

$w_7 = w_7 - \delta_3 o_2$

$w_8 = w_8 - \delta_4 o_2$

Q3] out of syllabus

Q4]

a]

$$\frac{\partial L}{\partial w_0} = -2 \sum_{i=1}^{m} (y_i - (x_i + z_i) w_1 - w_0)$$

$$\frac{\partial L}{\partial w_1} = -2 \sum_{i=1}^{m} (y_i - (x_i + z_i) w_1 - w_0) x_i$$

$$\frac{\partial L}{\partial w_2} = -2 \sum_{i=1}^{m} (y_i - (x_i + z_i) w_2 - w_0) z_i$$

if $w_1 = w_2$ we can just replace x_i in the formula with $(x_i + z_i)$

$$w_0 = \frac{\sum_{i=1}^{m} y_i - w_1 \sum_{i=1}^{m} (x_i + z_i)}{m}$$

$$w_1 = \frac{m \sum (x_i + z_i) y_i - \sum (x_i + z_i) \sum y_i}{m \sum (x_i + z_i)^2 - (\sum (x_i + z_i))^2}$$

b] we expect logistic regression to produce same params as NB Gaussian when the

- features satisfy conditional independence assumption
- variance is class independent for each feature

& dataset size is not very small

Also, the data ~~is~~ has to be linearly seperable

c] (Not taught yet)

d]
- True - MLE would be same.
~~MLE is~~ Likelihood for both the cases is $p^{20}(1-p)^{80}$.
p is probability of getting a heads
$\therefore \frac{\partial L}{\partial p} = 0$, would yeild same. MLE does not depend on prior knowledge. $p = \frac{20}{100}$
- False

$$MAP_{coin} = \frac{5+20-1}{5+5+20+80-2} = \frac{19}{108}$$

$$MAP_{thumback} = \frac{20+20-1}{20+20+20+80-2} = \frac{39}{138}$$

$MAP_{thumback} > Map_{coin}$