★ Fall 2022 Midterm

Q1] $y \in \{A, B\}$

$P(X = x_i \mid Y = A) = \frac{x}{\lambda^2_{A_i}} e^{-\frac{x^2}{2\lambda^2_{A_i}}}$

a] dataset from $\{(x_1^1, x_2^1, y^1) \cdots (x_1^m, x_2^m, y^m)\}$

Likelihood,

$L(\theta \mid D) = \prod_{k=1}^{m} P(Y = \text{class}) \prod_{i=1}^{2} P(x_i = x \mid y = \text{class})$

Let θ be $P(Y = A)$

$P(Y = \text{class}) = \theta^{\{y=A\}} (1-\theta)^{\{y=B\}}$

$P(X_i = x \mid Y = \text{class}) = \left(\frac{x}{\lambda^2_{A,i}} e^{-\frac{x^2}{2\lambda^2_{A,i}}}\right)^{\{y=A\}} \times \left(\frac{x}{\lambda^2_{B_i}} e^{-\frac{x^2}{2\lambda^2_{B,i}}}\right)^{\{y=B\}}$

Log likelihood,

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

$$+ \sum_{k=1}^{m} \sum_{i=1}^{2} \left[\{y^k = A\} \left(\ln x^k - 2 \ln \lambda_{A_i} - \frac{{x^k}^2}{2\lambda^2_{A_i}} \right) + \{y^k = B\} \left(\ln x^k - 2 \ln \lambda_{B_i} - \frac{{x^k}^2}{2\lambda^2_{B_i}} \right) \right]$$

b] $\dfrac{\partial l(\theta \mid D)}{\partial \lambda_{A_i}} = 0$ (MLE for λ_{A_i})

$$\Rightarrow \sum_{k=1}^{m} \left[\{y^k = A\} \left(-\frac{2}{\lambda_{A_i}} + \frac{(x_i^k)^2}{\lambda_{A_i}^3} \right) \right] = 0$$

let $\sum_{k=1}^{m} \{y^k = A\} = d_A$

& $d_{B \{y^k = B\}} = d - d_A$

$$\Rightarrow -2 \frac{\sum_{k=1}^{m} \{y^k = A\}}{\lambda_{A_i}} + \frac{\sum_{k=1}^{m} (x_i^k)^2 \{y^k = A\}}{\lambda_{A_i}^2} = 0$$

$$\Rightarrow 2 d_A = \frac{\sum_{k=1}^{m} (x_i^k)^2 \{y^k = A\}}{\lambda_{A_i}^2}$$

$$\Rightarrow \lambda_{A_i} = \sqrt{\frac{\sum_{k=1}^{m} (x_i^k)^2 \{y^k = A\}}{2 d_A}}$$

Similarly,

$$\Rightarrow \lambda_{B_i} = \sqrt{\frac{\sum_{k=1}^{m} (x_i^k)^2 \{y^k = B\}}{2 (d - d_A)}}$$

$i \in \{1, 2\}$

c) $\theta = \frac{d_A}{d} = \frac{3}{7}$

$\lambda A_1 = \sqrt{\frac{4^2+3^2}{2\times 3}} = \frac{5}{\sqrt{3}}$

$\lambda A_2 = \sqrt{\frac{3^2+8^2+4^2}{2\times 3}}$

$\lambda B_1 = \sqrt{\frac{6^2+3^2+2^2+5^2}{2\times 4}}$

$\lambda B_2 = \sqrt{\frac{2^2+5^2+1^2+4^2}{2\times 4}}$

Q3] a]

	Bias	variance
i]	increases	decreases
(ii)	increases	decreases
iii]	increases	decreases
iv]	decr-	increases

d) No. because we ~~are creating random~~ might have not reached optimal solution for test data. since we create stupid classifiers it is highly unlikely to overfit the training set but it might reduce the error for test set.

Yes because this stupid model has $d_E = \infty$. since it is stupid it is highly unlikely to overfit & it should fit the overall data.

e) • L2 regularization

Adding a $\frac{\lambda |w_i|^2}{2}$ cost to the loss function to prevent high values of weights or to prevent some weights to dominate the prediction.

• Drop random nodes during training to prevent model from learning the test set

• Dropping a layer or using minimum layers that can accurately fit the training & validation data

• Early stopping using validation set

Question 4]

a] Entropy x_1 $= 0 - \frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}$

Entropy x_2 $= \left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) \times 2$

Entropy x_3 $= \left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) \times 2$

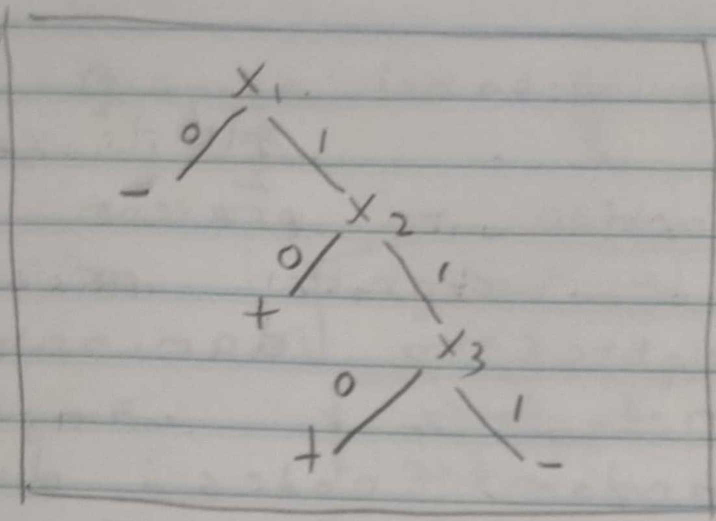
Pick X_1 for root.

2^{nd} layer + + −

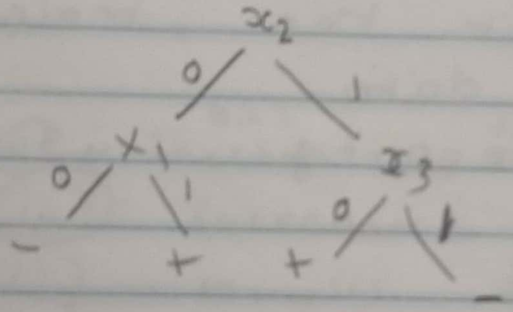
Entropy x_2 $= 0 \left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right)$

Entropy x_3 $=$ same as x_2

So, pick x_2



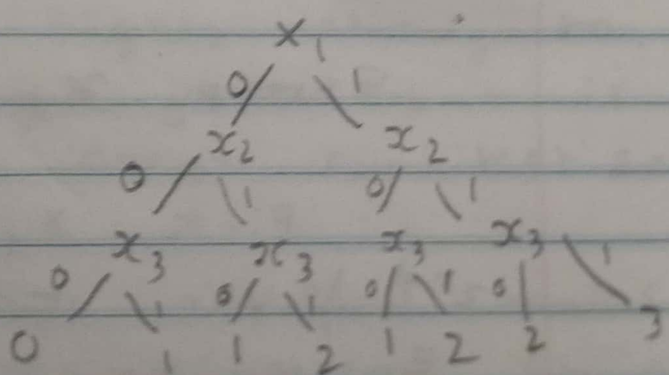
b) This is not optimal as we can create tree with height/depth 2.



c)

x_1	x_2	x_3	0
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	2
1	0	0	1
1	0	1	2
1	1	0	2
1	1	1	3

we would need 8 leaves



d] we would need 2^d leaf nodes
This is because every feaure can
take 2 values ~~and~~, both of which
result in a different answer so
~~a~~ each layer of tree would be
just one feature ~~that~~.
~~&~~ d features would represent
~~a~~ tree of debt d
which has 2^d leaves