• bind() function which wraps around the original function.

ML Exams

Spring 2023

1] $P_\theta(x) = \theta e^{-\theta x}$, $x \in [0, \infty]$

Likelihood $= \prod_{k=1}^{d} \theta e^{-\theta x^k}$

Log Likelihood $= \sum_{k=1}^{d} \left[\log\theta - \theta x^k \right]$

$\Rightarrow \frac{\partial L(\theta, D)}{\partial \theta} = \sum_{k=1}^{d} \left[\frac{1}{\theta} - x^k \right]$

$\Rightarrow$ MLE θ,

$\frac{\partial L(\theta, D)}{\partial \theta} = 0$

$\Rightarrow \sum_{k=1}^{d} \frac{1}{\theta} = \sum_{k=1}^{d} x^k$

$\Rightarrow \theta = \frac{d}{\sum_{k=1}^{d} x^k}$, d is number of datasets

$\Rightarrow \theta = \frac{n}{\sum_{k=1}^{n} x^k}$

b]

Entropy x_1 = $-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$

Entropy x_2 = $-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$

Entropy x_3 = $-\frac{1}{2}\log\frac{1}{2} - 1\log 1 < 1$

Take $x_3 \rightarrow$ $1 \rightarrow +$

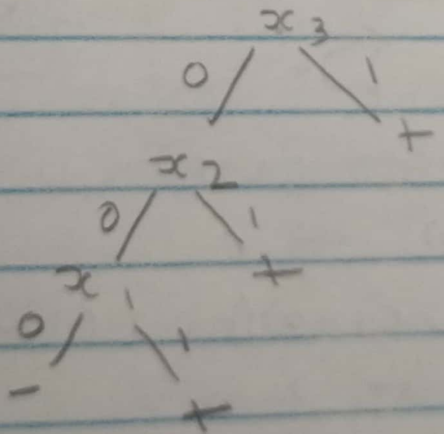
$0 \rightarrow + \; - \; -$

for level 2

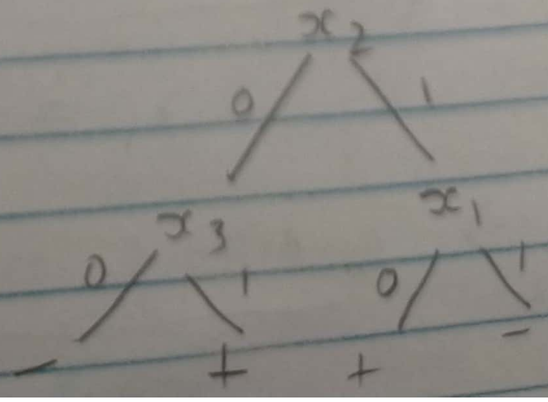
Entropy x_1 = $0 - \frac{1}{2}\log\frac{1}{2}$

Entropy x_2 = $0 - \frac{1}{2}\log\frac{1}{2}$

Pick x_2 or x_3



Optimal tree



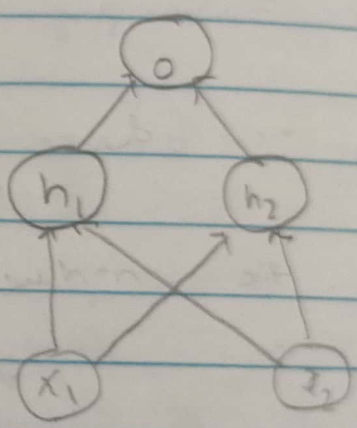
2] Out of syllabus

3] a) False. There might be noise in data or we may have high priors.

b] False, we cannot find such a weight setting because we dont know which K features need to be true. It can be random K among n.

1 out of 2 true is a XOR

c] $\frac{\partial E}{\partial w_{oi}} = \frac{2}{2}(y-o)\frac{\partial(y-o)}{\partial w_{oi}}$



$\Rightarrow (y-o)(o)(1-o)\frac{\partial \sum_{j=1}^{2} w_{oj} h_j}{\partial w_{oi}}$

$\Rightarrow -\underbrace{(y-o)(o)(1-o)}\, h_i$

∂_o can be

$\partial_o = (y-o)(o)(1-o)$

$\frac{\partial E}{\partial w_{i1}} = -(y-o)(o)(1-o)(h_i)(1-h_i)\frac{\partial \sum_{i=1}^{2} w_{i1} x_i}{\partial w_i x_i}$

$\Rightarrow +\underbrace{(y-o)(o)(1-o)(h_1)(1-h_1)}\, x_i$

$\partial_1 =$

$\frac{\partial E}{\partial w_{i2}} \Rightarrow \underbrace{(y-o)(o)(1-o)(h_2)(1-h_2)}_{\partial_2} X_i$

$w_{01} = w_{01} - \alpha \partial_0 h_1$
$w_{02} = w_{02} - \alpha \partial_0 h_2$

$w_{11} = w_{11} - \alpha \partial_1 x_1$
$w_{12} = w_{12} - \alpha \partial_1 x_2$

$w_{21} = w_{21} - \alpha \partial_2 x_1$
$w_{22} = w_{22} - \alpha \partial_2 x_2$

d) 3.

Q4] c] LR will better than GNB

-> GNB has assumption that variance of all parameters are same. This assumption can be violated
Also, the data features may not be indepent, this is another violation.
~~NB generally~~

GNB perform better

NB converges faster than LR.
For smaller datasets (100s of examples) NB ~~can~~ is likely to perform better.
For missing data GNB can perform better