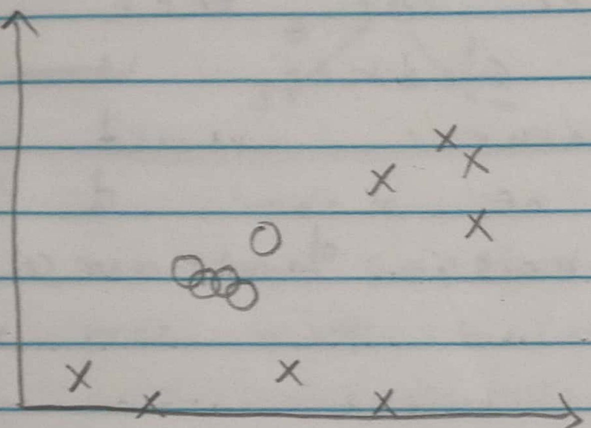
★ Mid term fall 2023

1] solved in other book

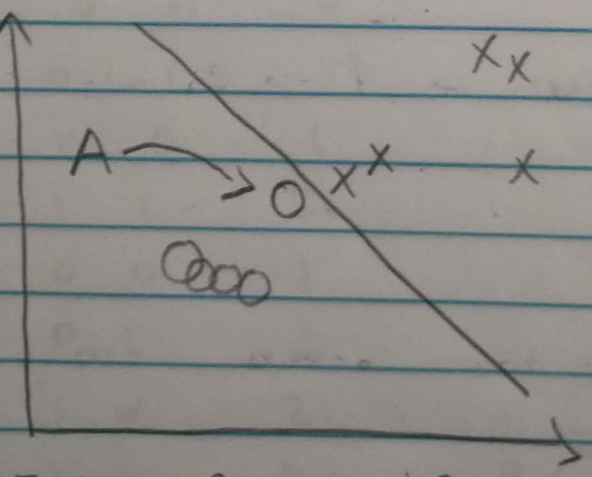
Q 2] Classification & Regression

a)



LR -> No, it won't have zero training error, linear line cannot seperate the X & O

3-nearest neighbors: Yes, it will have 0 train error, because closest points from each point belong to the same class



LR -> Yes, Possible

KNN -> No, it will have non zero error because of point A.

c] Loss $= \frac{1}{m} \sum_{i=1}^{m} (y^{(i)} - w_1 x^{(i)} - w_0)^4$

$$\frac{\partial L}{\partial w_0} = \underbrace{-\frac{4}{m} \sum_{i=1}^{m} (y^i - w_1 x^i - w_0)^3}_{\text{Gradient w.r.t } w_0}$$

$$\frac{\partial L}{\partial w_1} = \underbrace{-\frac{4}{m} \sum_{i=1}^{m} (y^i - w_1 x^i - w_0)^3 x_i}_{\text{Gradient w.r.t } w_1}$$

d) writ pseudo code for batch gradient descent

→ Repeat Until convergence

for i in dataset

$w_0 = w_0 + \left(-\frac{4}{m} (y^i - w_1 x^i - w_0)^3 \right)$

$w_1 = w_1 + \left(-\frac{4}{m} x_i (y^i - w_1 x^i - w_0)^3 \right)$

End for

End loop

*where y^i is the actual class of ith dataset.

Q3] a]

	Bias	Variance
L2 regularization in LR	Increases	Decreases
Increasing K in KNN	Increases	Decreases
Pruning Decision Tree	Increases	Decreases
Adding hidden layer to NN	Decreases	Increases

> KNN is different from other models because increasing value of K increases stability & reducing variance.

b] • No - Because model is verly unlikely to overfit & the testing error might still be decreasing.

• Yes - Because then α_m would be ∞ & weights would be 0 & so it learns nothing. Also, if the current weak classifier has 0 error it might fit overall data & we can simply use it to classify data leaving out the other models.

(since it $\alpha_m = \infty$ & others have a constant α_m)

$\hookrightarrow G(x_i)$

False

c] No. A neural need can represent any CNF with one hidden layer only if it can have exponential nodes, we cannot limit it to $O(nk)$

d] False. We might need exponential nodes to represent a CNF with one hidden layer. So, a polynomial is not enough

→ For c & d suppose we have a boolean function which outputs 0, 1 for all combination of inputs. Truth table has 2^n combinations

So, In worst case we may want to have n^n layer NN.
↳ nodes in

$2^n > n^k$

$2^n > nk$

Asymptotically,

Q4] a) Circle - - - -

E1 E2 E3 E4 E5 E6

b)

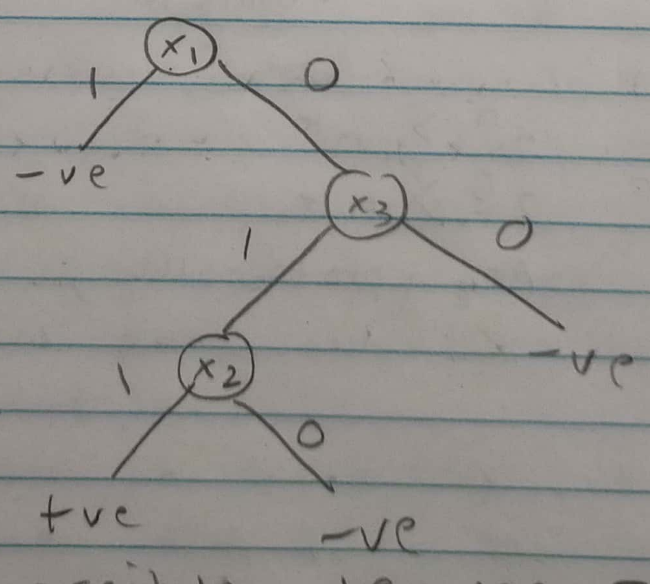
On the branch $x_3 = 1$
E2, E4

Entropy $x_2 = \frac{1}{2}(1 \log 1) + \frac{1}{2}(1 \log 1)$
$= 0$

Entropy $x_4 = \frac{1}{2}\left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) = 1$

Entropy $x_5 = 1$

Use x_2,



c] Max possible leaves =
min $(2^n, d)$
$n \rightarrow$ number of features

Fall 2023

Q1] $Y \in \{A, B\}$, Features $\in \{1, 2\}$

a) Log Likelihood,

$$\Rightarrow \sum_{K=1}^{m} \log(P(Y = c_i)) + \sum_{K=1}^{m} \sum_{j=1}^{n} \log P(X = x_j \mid Y = c_i)$$

$$\log(P(Y=c_i)) = \log\left[\theta^{\{x=A\}} (1-\theta)^{\{x=B\}}\right]$$

$$\log(P(X=x_j \mid Y=c_i)) = \log\left[(\lambda_{Aj}(1-\lambda_{Aj}))^{x_j\{y=A\}} \times \lambda_{Bj}(1-\lambda_{Bj})^{x_j\{y=B\}}\right]$$

b] MLE for λ_{Aj} , $j \in \{1, 2\}$

$$\Rightarrow \frac{\partial \log P(X = x_j \mid Y = c_i)}{\partial \lambda_{Aj}} = 0$$

$$\Rightarrow \sum_{K=1}^{m} \{y=A\} \frac{\partial}{\partial \lambda_{Aj}}\left[\log(\lambda_{Aj}) + x_j^K \log(1-\lambda_{Aj})\right] = 0$$

$$\Rightarrow \sum_{K=1}^{m} \{y=A\}\left[\frac{1}{\lambda_{Aj}} - \frac{x_j^K}{1-\lambda_{Aj}}\right] = 0$$

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

$$\Rightarrow \frac{d_A}{\lambda_{Aj}} - \frac{\sum_{K=1}^{m} x_j^K \{y=A\}}{1-\lambda_{Aj}} = 0$$

$$\Rightarrow d_A - d_A \lambda_{Aj} - \lambda_{Aj} \sum_{K=1}^{m} x_j^K \{y=A\} = 0$$

$$\Rightarrow \boxed{\lambda_{Aj} = \frac{d_A}{d_A + \sum_{K=1}^{m} x_j^K \{y=A\}}}$$

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

Similarly,

$$\boxed{\lambda_{Bj} = \frac{d-d_A}{d-d_A + \sum_{k=1}^{m} x_j^k \{y=B\}}}$$

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

c] $\theta = \frac{3}{7}$, $1-\theta = \frac{4}{7}$

$$\lambda_{A1} = \frac{3}{3+(9+2+6)} = \frac{3}{20}$$

$$\lambda_{A2} = \frac{3}{3+(2+6+4)} = \frac{3}{15} = \frac{1}{5}$$

$$\lambda_{B1} = \frac{4}{4+(4+2+2+5)} = \frac{4}{17}$$

$$\lambda_{B2} = \frac{4}{4+(2+7+1+1)} = \frac{4}{15}$$