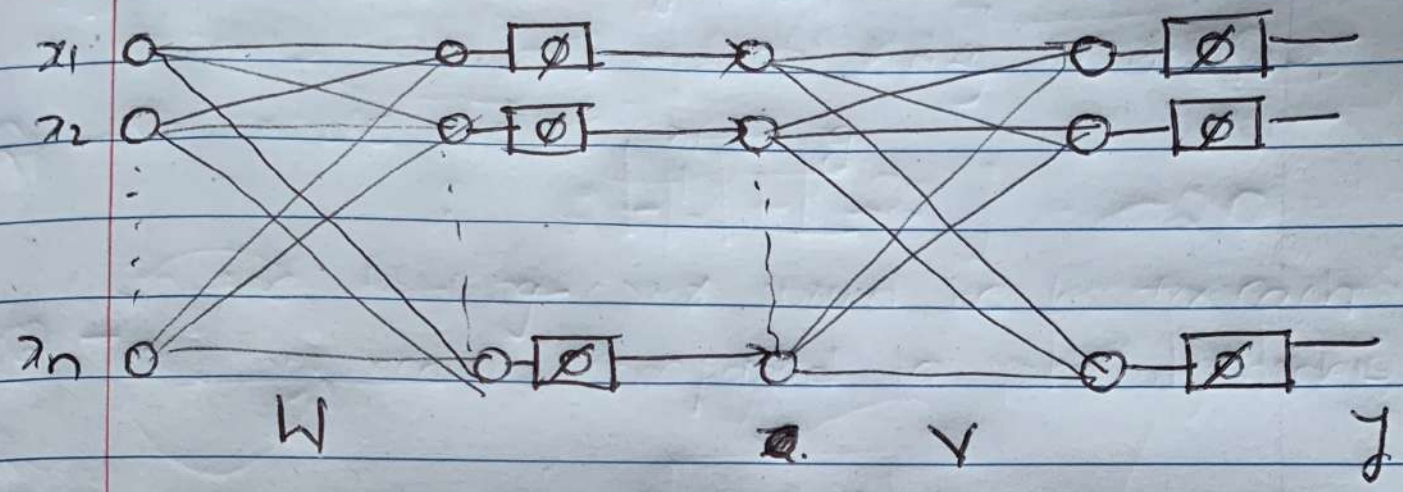
Name: Kalyan Kumar Paladugula

UIN: 679025059.

Neural Networks Midterm Sp2020

Q1) Given Network



$$y = \phi(V\{\phi(WX)\}) \quad ; \quad E = 1/2\,\|d-y\|^2.$$

$\eta = 1$

$$z = (d-y) \odot \phi'(V\phi(WX))$$

$$\frac{\partial E}{\partial V} = -\frac{1}{2} \cdot 2\,(d-y) \odot \phi'(V\phi(WX)) \cdot \phi(WX)$$

$$= -z \cdot \phi(WX)$$

$$\frac{\partial E}{\partial W} = \left[\left[-\frac{1}{2} \cdot 2 \cdot (d-y) \odot \phi'(V\phi(WX) \cdot V \cdot\right] \odot \phi'(WX)\right] \cdot X$$

$$= -\left[(z \cdot V) \odot \phi'(W \cdot X)\right] \cdot X.$$

$$V \leftarrow V + z \cdot \phi(WX)$$

$$W \leftarrow W + \left[(z \cdot V) \odot \phi'(W \cdot X)\right] \cdot X.$$

2) Single layer FF
n-neurons
Zero biases

$w_1, w_2 - - w_n$ are neuron weight vectors
$||w_1|| = ||w_2|| = - - - ||w_n||$

Let $x_1, x_1 - - x_m$ are input nodes.
X be the vector(input)

Consider the euclidean distance b/w input vector X and a weight vector w_i

$$d_i = \sqrt{||X - w_i||^2} = \sqrt{(X - w_i)^T \cdot (X - w_i)}$$

$$= \sqrt{X^T X - X^T w_i - w_i^T X + w_i^T w_i}$$

$$= \sqrt{||X||^2 + ||w_i||^2 - 2 X^T \cdot w_i}$$

Let $V_i = X^T \cdot w_i$ - Local field of neuron

$$d_i = \sqrt{||X||^2 + ||w_i||^2 - 2V_i}$$

since $||X||^2$ and $||w_i||^2$ are constant for all neuron's

For neuron 'i' to be the winner neuron V_i should be maximum

since $d_i = \sqrt{||X||^2 + ||w_i||^2 - 2V_i}$

$\Downarrow\Downarrow\Downarrow$

d_i should be minimum

In other words, winner neuron is the one whose weight vector is closest to an input in terms of Euclidean distance

(Q3) one input,
no bias
one output with step activation fn.

initial weight = 0

a) $w_{i+1} \leftarrow w_i + xy$

Initial weight $w_0 = 0$

i) First input $y_0 = \text{step}(w_0 \cdot x_0) = \text{step}(0) = 1$

New weight $w_1 = 1$

ii) second input $y_1 = \text{step}(w_1 \cdot x_1) = 1$

$w_2 = 1 + 1 \cdot 1 = 2.$

$w_3 = 3$

$w_4 = 4$

Limit of $w = \infty$

b) $w_1 \leftarrow \alpha w_0 + x_0 y_0$

$w_1 = x_0 y_0$

First Input $y_0 = \text{step}(w_0 \cdot x_0) = 1$

$w_1 = 1 \cdot 1 = 1$

Second Input $x_1 = 1$

$y_1 = \text{step}(w_1 \cdot x_1) = 1$

$w_2 = \alpha + 1$

$w_3 = 1 + \alpha + \alpha^2$

$1 + \alpha + \alpha^2 + \alpha^3 + \cdots + \infty$

Geometric progression

as $0 < \alpha < 1$, Limit of $W = \frac{1}{1-\alpha}$.

c) $w \leftarrow \alpha w + xy$.

$0 < \alpha < 1$

$w_0 = 0$

training sequence: $0, 1, 0, 1, 0, 1 \ldots$

First Input:

$x_0 = 0 \quad w_0 = 0$

$y_0 = \text{step}(0 \cdot 0) = 1$

$w_1 = \alpha \cdot 0 + 0 \cdot 1 = 0$

Second input:

$x_1 = 1, \quad w_1 = 0$

$y_1 = 1$

$w_2 = \alpha \cdot w_1 + 1 \cdot 1$

$= 1$

Third input:

$x_2 = 0, \quad w_2 = 1$

$y_2 = 1$

$w_3 = \alpha \cdot w_2 + 0$

$= \alpha$

Fourth Input:

$x_3 = 1, \quad w_3 = \alpha.$

$y_3 = \text{step}(\alpha \cdot 1) = 1$

$w_4 = \alpha^2 + 1.$

Fifth Input

$x_4 = 0, \quad w_4 = \alpha^2 + 1.$

$y_4 = 1$

$w_5 = \alpha(\alpha^2 + 1) + 0$

$= \alpha^3 + \alpha.$

Sixth Input:

$x_5 = 1, \quad w_5 = \alpha^3 + \alpha.$

$y_5 = 1$

$w_6 = \alpha(\alpha^3 + \alpha) + 1.$

$= \alpha 1 + \alpha^2 + \alpha^4$

w_1 odd 1.

w_2 even α.

w_3 odd α^2+1

w_4 even $\alpha^3+\alpha$.

$\alpha^4+\alpha^2+1$.

$\alpha^5+\alpha^3+\alpha$.

~~$1+\alpha$ α^2+1 $\alpha(\alpha^2+1)$~~

i) <u>when n is even:</u>

$$\alpha+\alpha^3+\alpha^5 \ldots = \alpha(1+\alpha^2+\alpha^4+\ldots)$$

common ratio $= \alpha^2$

Geometric progression

as $0<\alpha<1$ Limit of $w = \frac{\alpha}{1-\alpha^2}$.

ii) <u>when n is odd</u>

$1+\alpha^2+\alpha^4 \ldots$

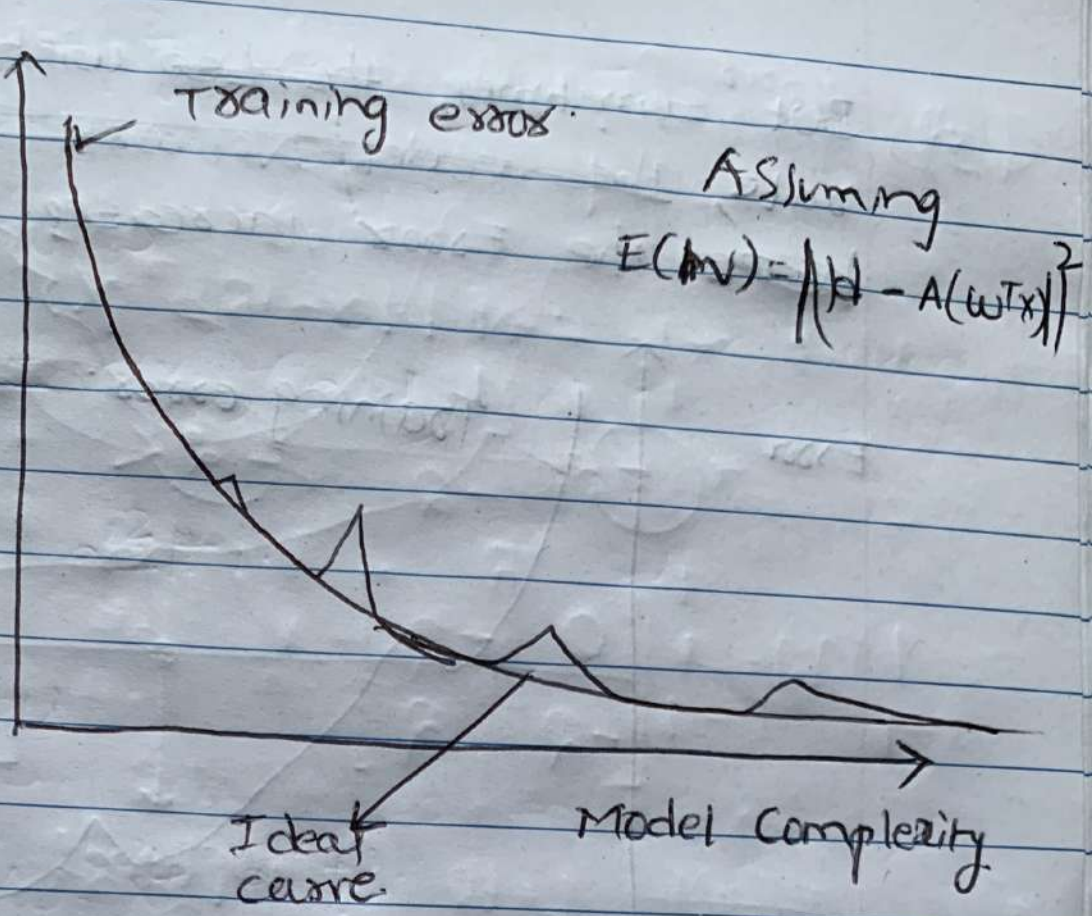
common ratio $= \alpha^2$

<u>G.P</u>

as $0<\alpha<1$ $\lim_{w\to\infty} = \frac{1}{1-\alpha^2}$.

4)

a)

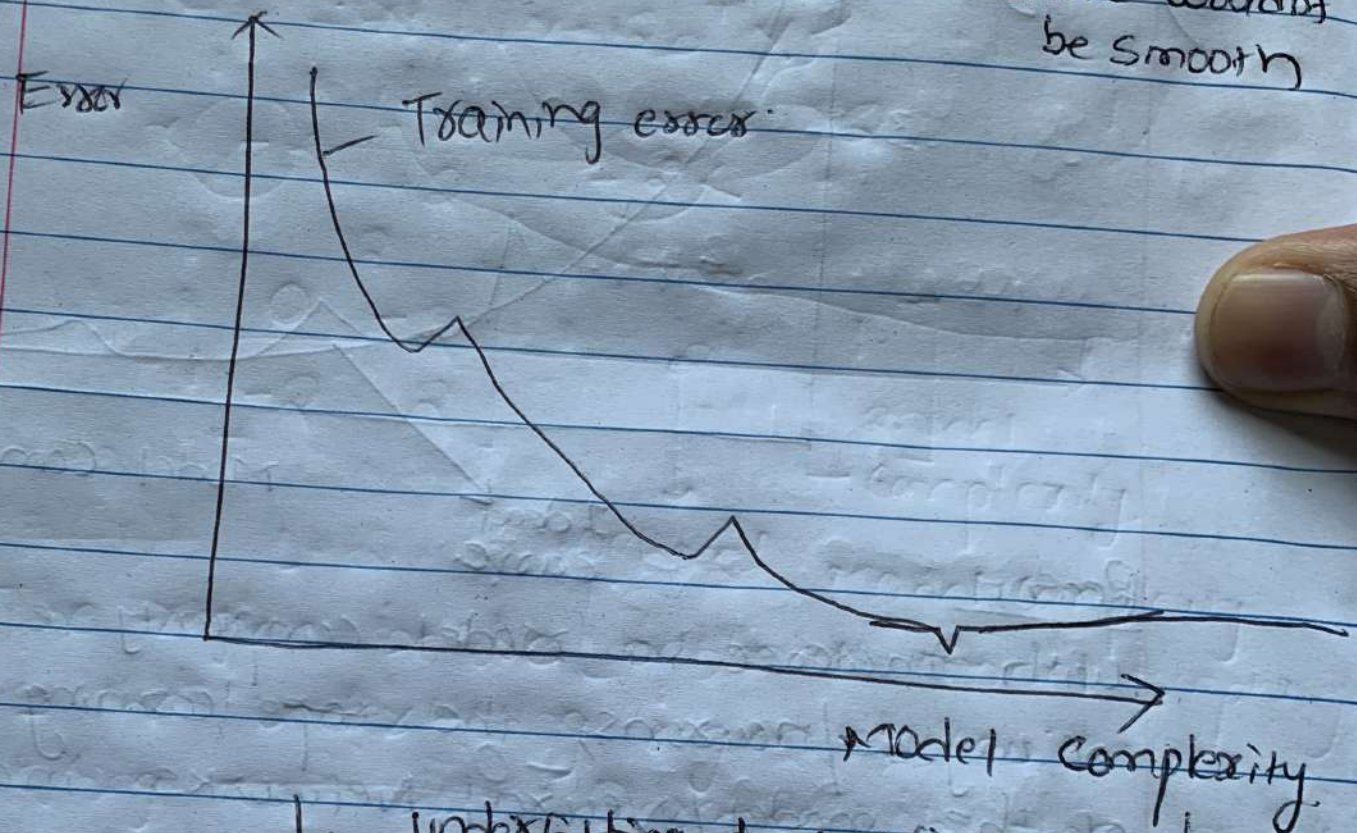


Reason:

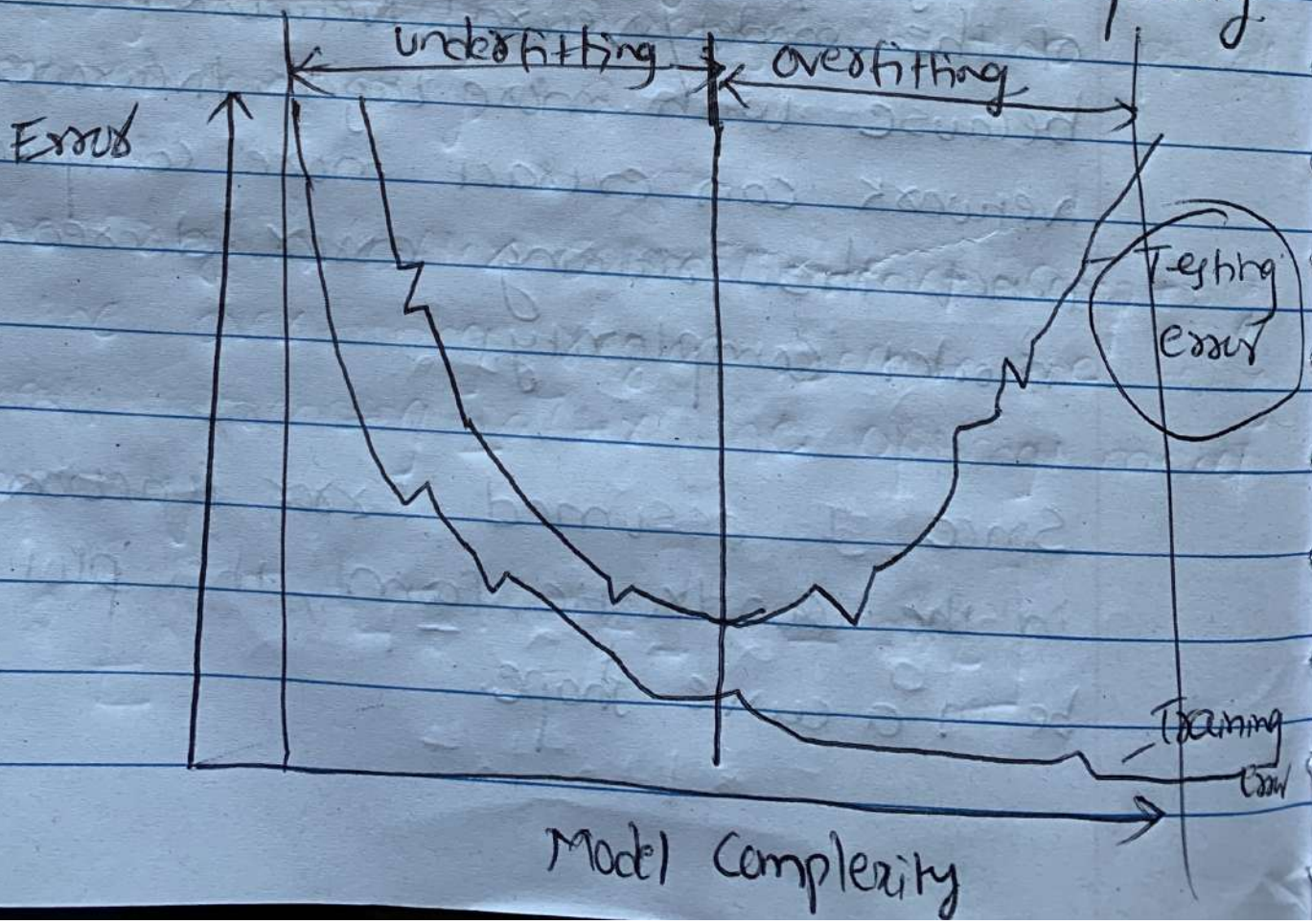
With Increase in Model complexity ie number of Layers/neurons, the ~~cap~~ learning capacity of the model/Neural Networks increases because with more Layers/Neurons, the Network can extract more complex features Hence, the training error decreases with model complexity.

Since I assumed Error function to be in the Quadratic form, the plot would be in a curve shape.

But Sometimes, ~~the loss~~ with high learning rate, ~~the model/Network skips the minima~~ ~~here~~ the Error increases and curve would not be smooth

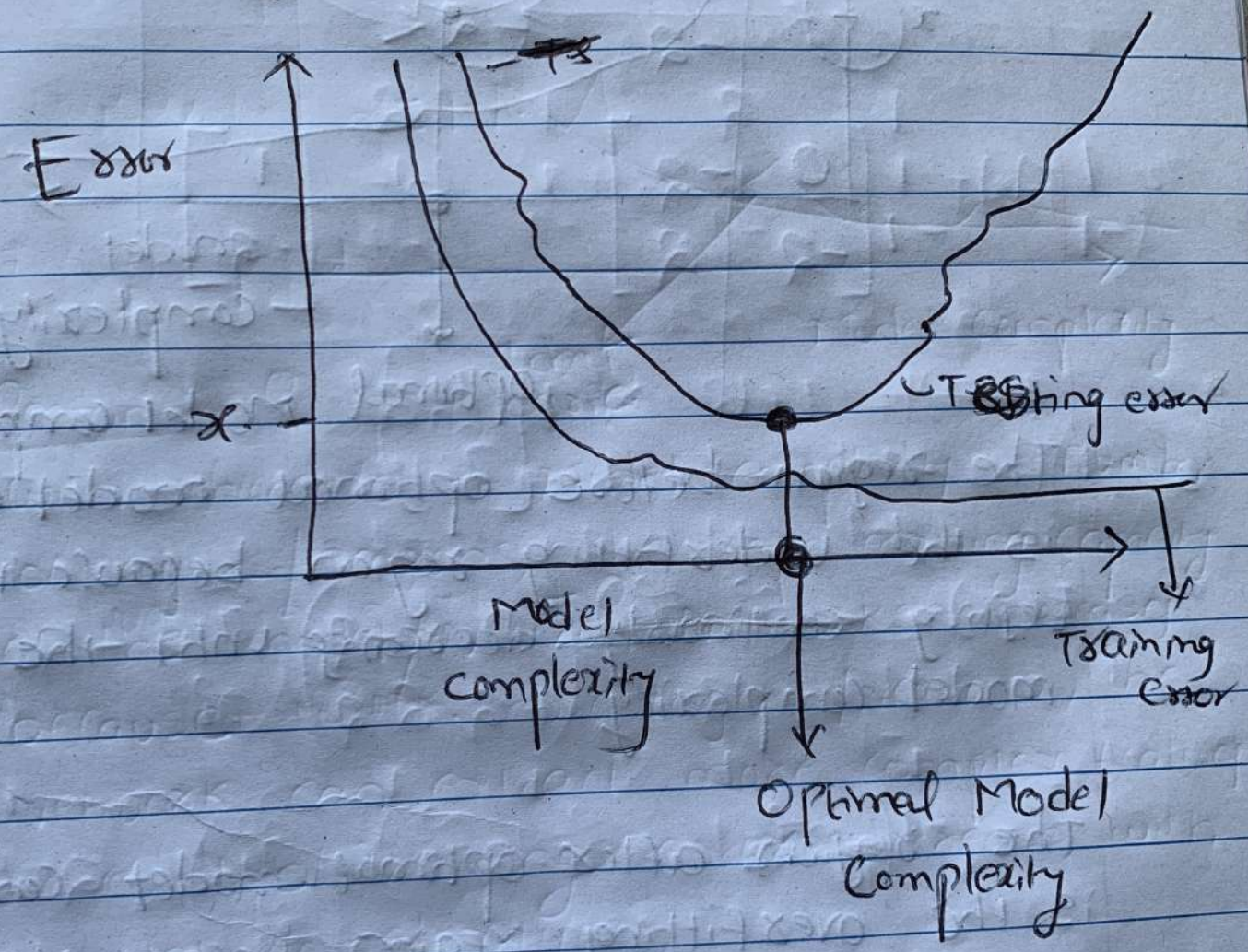


b)



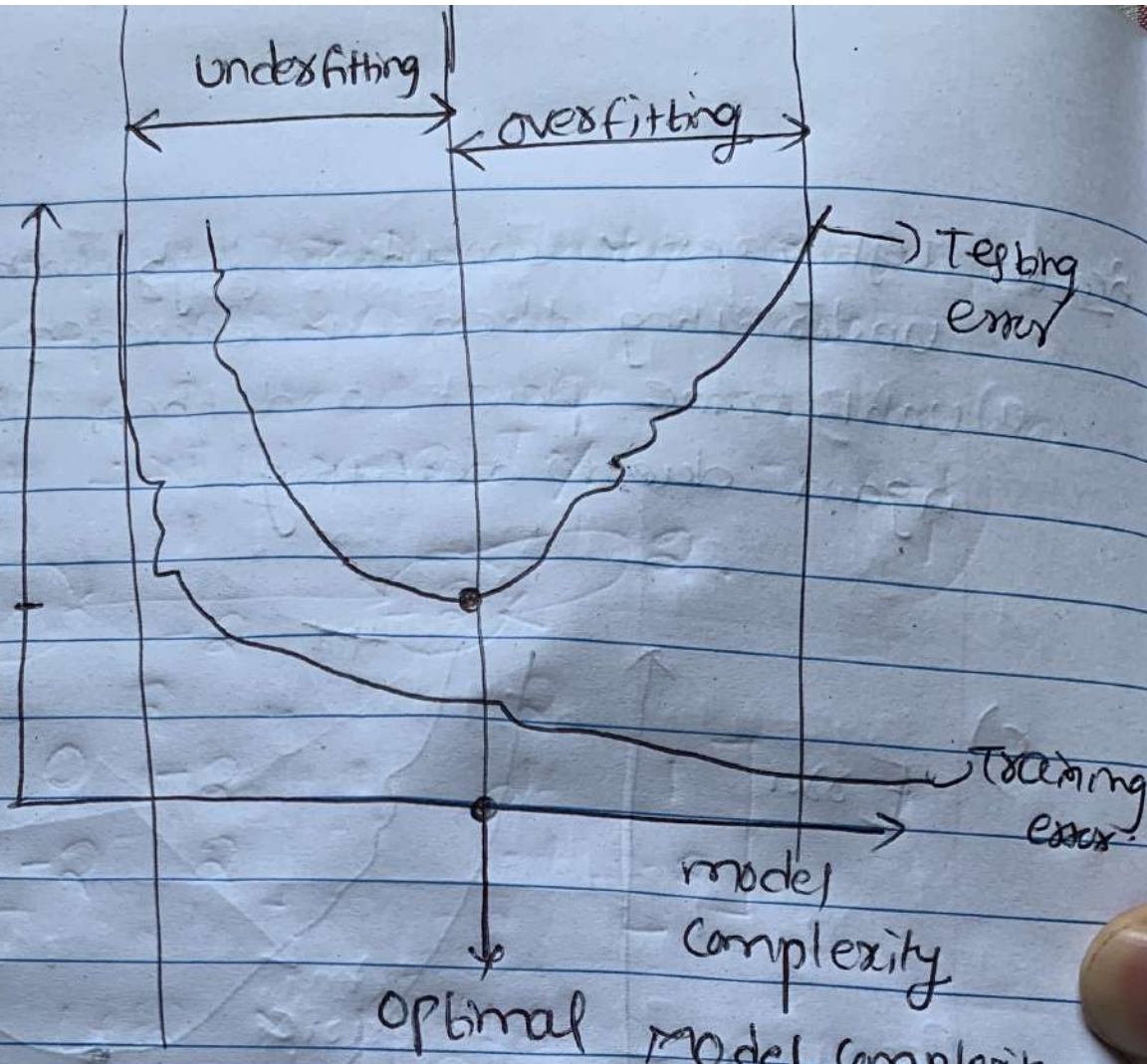
Reason: Testing error ~~would~~ is high initially due to underfitting then it decreases gradually until some point and then it increases again due to overfitting

c)



Our model should perform well on unseen data i.e testing data. Hence, optimal model complexity would be the one where testing error is minimum

d)



The region before optimal model complexity is the underfitting region because the testing ~~continuously~~ decreases until the optimal model complexity

The region after optimal model complexity is the overfitting region because the testing starts increasing after the optimal model complexity

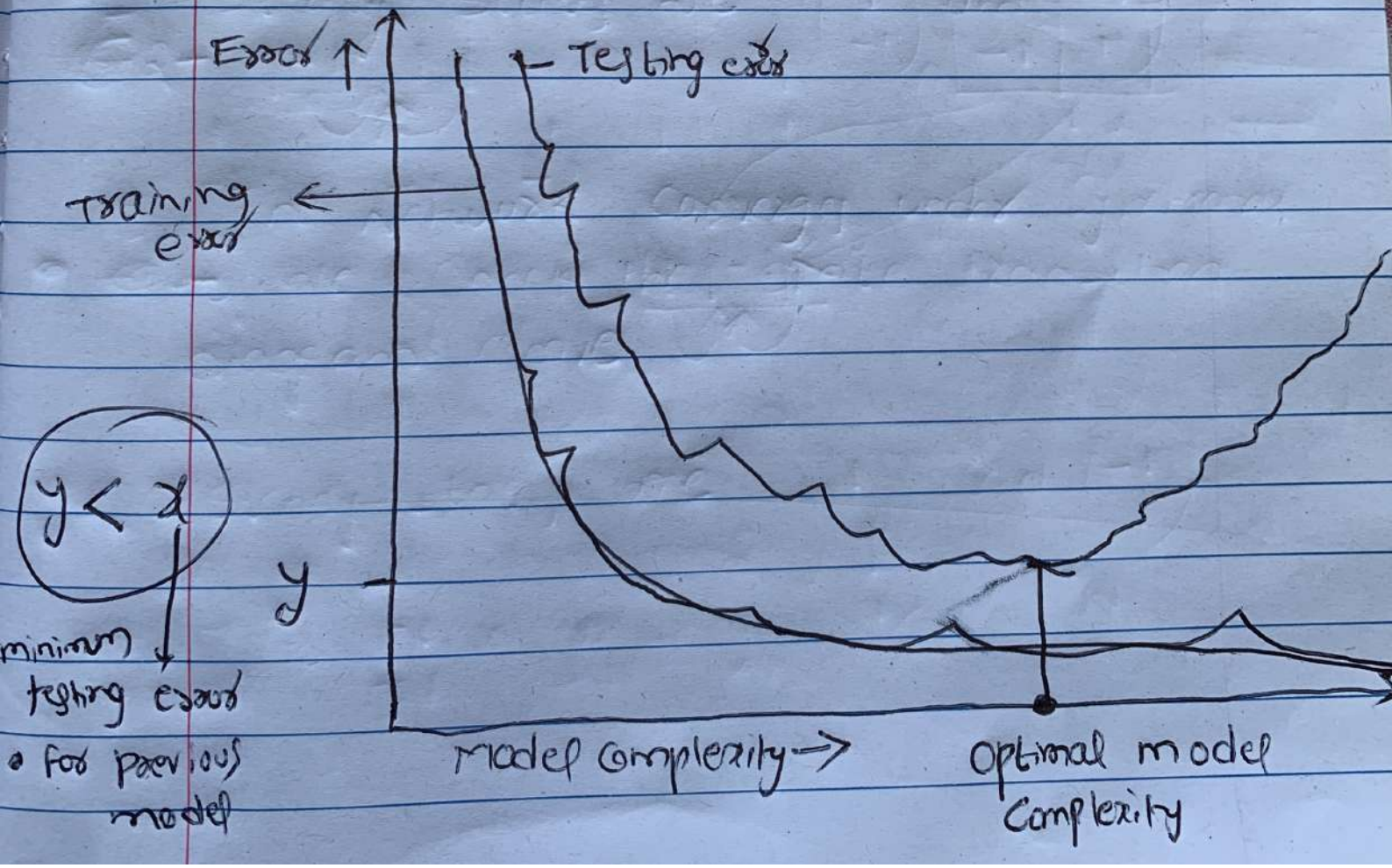
4)

e) With more training data, the variances of the models decreases which make them less vulnerable to overfitting.

So, the testing error would be less for each of the models compared to previous models.

Hence, even the high complex model would not overfit.

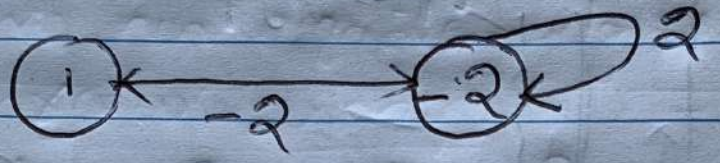
→ we would get broad error (testing curve).



Sometimes, ~~we would~~ the optimal model complexity increases due to decrease in variance and testing error.

(Q5) $\phi(x) = 1$ if $x \geq 0$
-1 if $x < 0$

$E(x) = \frac{1}{2} x^T W x - b^T x$



$$W = \begin{bmatrix} 0 & -2 \\ -2 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

~~$E(x)$~~ Let $x = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}$

$$E(x) = -(\tilde{x}_1 \ \tilde{x}_2)\begin{pmatrix} 0 & -2 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} - 2(1 \ -2)\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix}$$

$$= (-2\tilde{x}_2 \quad 2\tilde{x}_1 - 2\tilde{x}_2)\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} - 2(\tilde{x}_1 - 2\tilde{x}_2)$$

$$= 2\left[\tilde{x}_1\tilde{x}_2 + (\tilde{x}_2)^2 - \tilde{x}_1\tilde{x}_2\right]$$

$$= 2\left[2\tilde{x}_1\tilde{x}_2 - 2(\tilde{x}_2)^2 - \tilde{x}_1 + 2\tilde{x}_2\right]$$

$$E(x) = 2\left(2\tilde{x}_1\tilde{x}_1 - 2(\tilde{x}_2)^2 - \tilde{x}_1 + 2\tilde{x}_2\right)$$

a) $x_1 = \begin{pmatrix}1\\1\end{pmatrix}$, $x_2 = \begin{pmatrix}1\\-1\end{pmatrix}$, $x_3 = \begin{pmatrix}-1\\1\end{pmatrix}$, $x_4 = \begin{pmatrix}-1\\-1\end{pmatrix}$

$y_1 = \phi(-2\tilde{x}_2 + 1)$, $y_2 = \phi(-2\tilde{x}_1 + 2\tilde{x}_2 - 2)$

Asynchronous update:

i) updating second neuron first

$x_1 \Rightarrow \begin{pmatrix}1\\1\end{pmatrix} \Rightarrow \begin{matrix}2 & 1 & 2 & 1\\ 1 & 1 & 1 & 1\\ -1 & -1 & -1 & -1\end{matrix} = \begin{pmatrix}1\\-1\end{pmatrix}$

$x_2 \Rightarrow \begin{pmatrix}1\\-1\end{pmatrix} \Rightarrow \begin{matrix}2 & 1 & 2 & 1\\ 1 & 1 & 1 & 1\\ -1 & -1 & -1 & -1\end{matrix} = \begin{pmatrix}1\\-1\end{pmatrix}$

$x_3 \Rightarrow \begin{pmatrix}-1\\1\end{pmatrix} \Rightarrow \begin{matrix}2 & 1 & 2 & 1\\ -1 & -1 & -1 & -1\\ 1 & 1 & 1 & 1\end{matrix} = \begin{pmatrix}-1\\1\end{pmatrix}$

$x_4 \Rightarrow \begin{pmatrix}-1\\-1\end{pmatrix} \Rightarrow \begin{matrix}2 & 1 & 2 & 1\\ -1 & 1 & 1 & 1\\ -1 & -1 & -1 & -1\end{matrix} = \begin{pmatrix}1\\-1\end{pmatrix}$

ii) updating First Neuron First

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \begin{matrix} 1 & 2 & 1 & 2 \\ -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 \end{matrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$x_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \Rightarrow \begin{matrix} 1 & 2 & 1 & 2 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \end{matrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$x_3 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \Rightarrow \begin{matrix} 1 & 2 & 1 & 2 \\ -1 & -1 & -1 & -1 \\ 1 & +1 & 1 & 1 \end{matrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$x_4 = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \Rightarrow \begin{matrix} 1 & 2 & 1 & 2 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \end{matrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$E(x) = 2\left[2\tilde{x}_1\tilde{x}_2 - 2(\tilde{x}_2)^2 - \tilde{x}_1 + 2\tilde{x}_2\right]$$

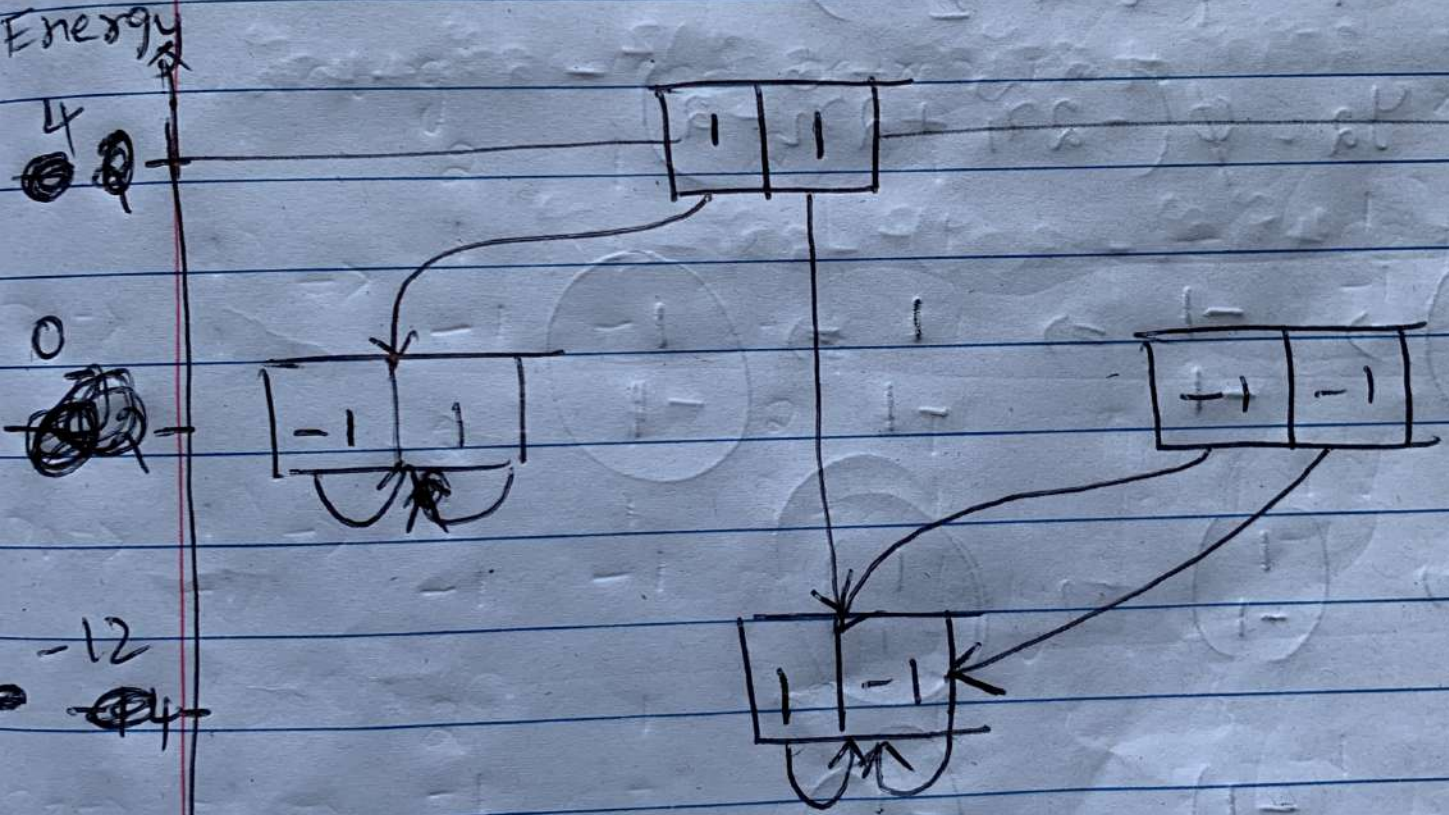
$$E(x_1) = 2\left[2 - 2 - 1 + 2\right] = 4$$

$E(x_2) = 2[-2 - 1 - 1 - 2] = -12$

$E(x_3) = 2[-2 - 1 + 1 + 2] = 0$

$E(x_4) = 2[2 - 1 + 1 - 2] = 0$

State transition diagram with Energy levels



Steady states are $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

b) Yes. The Network always converge under the asynchronous update. Check the state transition diagram and also since the weight matrix is symmetric and has non-negative diagonal elements

c) synchronous update

$$y_1 = \phi(-2\tilde{x}_2 + 1)$$

$$y_2 = \phi(-2\tilde{x}_1 + 2\tilde{x}_2 - 2)$$

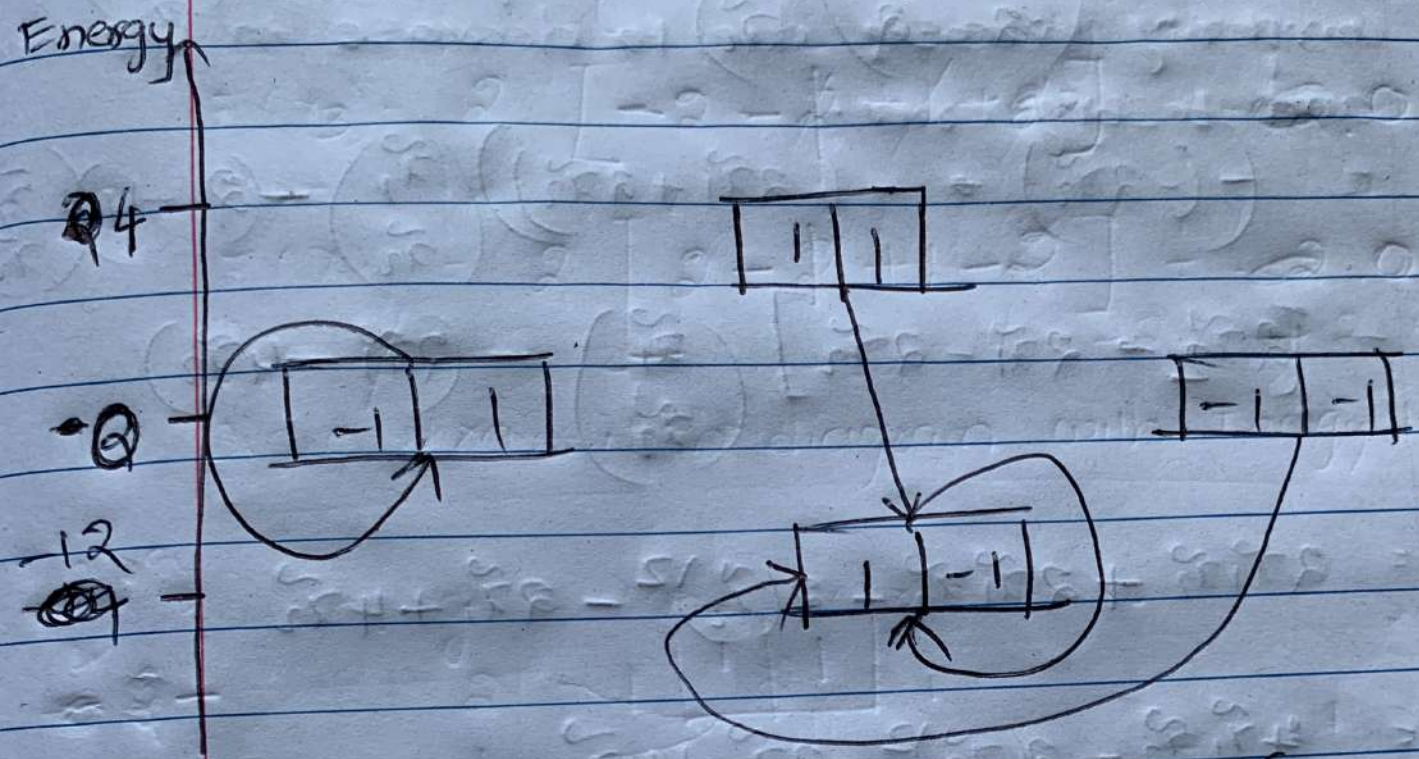
$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \Rightarrow \begin{matrix} -1 \\ -1 \end{matrix} \Rightarrow \begin{matrix} 1 \\ -1 \end{matrix} \Rightarrow \begin{matrix} 1 \\ -1 \end{matrix}$

$x_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$x_3 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} -1 \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$x_4 = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

State transition diagram



Steady states are $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$

The network converges under Synchronous update. Check the state transition diagram above