NPTEL Week 13 Live Sessions

on Deep Learning (noc24_ee04)

A course offered by: Prof. Prabir Kumar Biswas, IIT Kharagpur

- **Exam preparation series**

Powered by:

By

Arka Roy

NPTEL PMRF TA

Prime Minister’s Research Fellow
Department of Electrical Engineering, IIT Patna
Web: https://sites.google.com/view/arka-roy/home

① Recall that McCulloch Pitts (MP) neuron aggregates the inputs and takes a decision based on this aggregation. If the sum of all inputs is greater than the threshold (θ), then the output of MP neuron is 1, otherwise the output is 0. We say that a MP neuron implements a boolean function if the output of the MP neuron is consistent with the truth table of the boolean function. (In other words, if for a given input configuration, the boolean function outputs 1 then the output of the neuron should also be 1. Similarly, if for a given input configuration, the boolean function outputs 0 then the output of the neuron should also be 0.)

3 [End sem, 2021, CS541, IIT Patna]

Consider the following boolean function:

①

$$f(x_1, x_2, x_3, x_4) = (x_1 \text{ AND } x_2) \text{ AND } (!x_3 \text{ AND } !x_4)$$

1

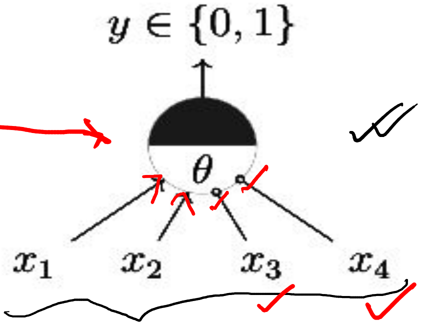
0

The MP neuron for the above boolean function is as follows:

$f(x_1, x_2 \ldots x_4)$

$= (x_1 \cdot x_2) \cdot (\bar{x}_3 \cdot \bar{x}_4)$

θ?



If any of the inhibitory input is = 1; then output = 0.

f

$\rightarrow x_3 = 1, f = 0$

$\rightarrow x_4 = 1, f = 0$

Recall that McCulloch Pitts (MP) neuron aggregates the inputs and takes a decision based on this aggregation. If the sum of all inputs is greater than the threshold (θ), then the output of MP neuron is 1, otherwise the output is 0. We say that a MP neuron implements a boolean function if the output of the MP neuron is consistent with the truth table of the boolean function. In other words, if for a given input configuration, the boolean function outputs 1 then the output of the neuron should also be 1. Similarly, if for a given input configuration, the boolean function outputs 0 then the output of the neuron should also be 0.

Consider the following boolean function:

$$f(x_1, x_2, x_3, x_4) = (x_1 \text{ AND } x_2) \text{ AND } (!x_3 \text{ AND } !x_4)$$

The MP neuron for the above boolean function is as follows:

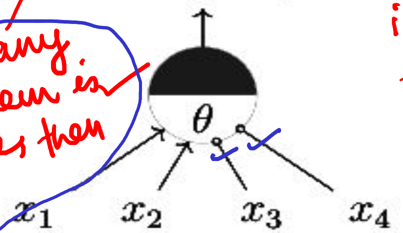
$$f(x_1, x_2, x_3, x_4) = (x_1 \cdot x_2) \cdot (\bar{x}_3 \cdot \bar{x}_4)$$

if any of x_1 or x_2 is equal to zero then also your final output $= 0$

$$y = \begin{cases} 1, & \sum_{\forall i} x_i \geq \theta \\ 0, & \text{elsewhere} \end{cases}$$

if either any of them is ones then

$f = 0$



Combination:-

x_1	x_2	x_3	x_4	
0	0	0	0	
0	0	0	1	1
0	0	1	0	2
0	0	1	1	3
0	1	0	0	
0	1	0	1	4
0	1	1	0	5
0	1	1	1	6
1	0	0	0	
1	0	0	1	7
1	0	1	0	8
1	0	1	1	9
1	1	0	0	
1	1	0	1	10
1	1	1	0	11

Recall that McCulloch Pitts (MP) neuron aggregates the inputs and takes a decision based on this aggregation. If the sum of all inputs is greater than the threshold (θ), then the output of MP neuron is 1, otherwise the output is 0. We say that a MP neuron implements a boolean function if the output of the MP neuron is consistent with the truth table of the boolean function. In other words, if for a given input configuration, the boolean function outputs 1 then the output of the neuron should also be 1. Similarly, if for a given input configuration, the boolean function outputs 0 then the output of the neuron should also be 0.

$y=1$

$\cancel{\theta}$

Consider the following boolean function:

$x_1 = 1, \; x_2 = 0, \; x_3 = 0, \; x_4 = 0 \rightarrow 0.$

$\sum x_i = 1 \not\geq 2.$

$1.$

$f(\cdot) = 0$

$y = 0$

$$f(x_1, x_2, x_3, x_4) = (x_1 \text{ AND } x_2) \text{ AND } (!x_3 \text{ AND } !x_4) = 0$$

The MP neuron for the above boolean function is as follows:

if your input combination $= 1, 1, 0, 0$ then your output will be equal to 1.

$x_1 = 1$

$x_2 = 2$

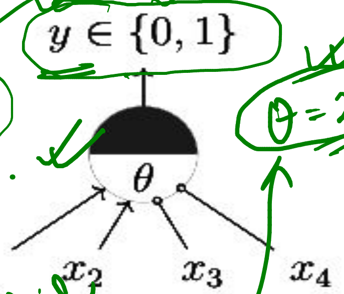
$x_3 = 0$

$x_4 = 0$

$f(x_1, x_2, x_3, x_4) = 1.$

$y = 1.$

$y = \begin{cases} 1; & \sum_i x_i \geq \theta \\ 0; & \text{elsewhere.} \end{cases}$



$\theta = 2$

min value of θ for which $y=1$,

$\theta \leq 2.$

$\sum x_i \geq \theta$

$1+1+0+0 = 2 \geq \theta$

Recall that McCulloch Pitts (MP) neuron aggregates the inputs and takes a decision based on this aggregation. If the sum of all inputs is greater than the threshold (θ), then the output of MP neuron is 1, otherwise the output is 0. We say that a MP neuron implements a boolean function if the output of the MP neuron is consistent with the truth table of the boolean function. In other words, if for a given input configuration, the boolean function outputs 1 then the output of the neuron should also be 1. Similarly, if for a given input configuration, the boolean function outputs 0 then the output of the neuron should also be 0.

Consider the following boolean function:

$0/1 \rightarrow 2^4 = 2\times2\times2\times2 = 16$

$\theta = 1$

A B

$2^2 = 4$

$$f(x_1, x_2, x_3, x_4) = (x_1 \text{ AND } x_2) \text{ AND } (!x_3 \text{ AND } !x_4)$$

$= 1.$

The MP neuron for the above boolean function is as follows:

$x_1 = 1$

$x_2 = 1$

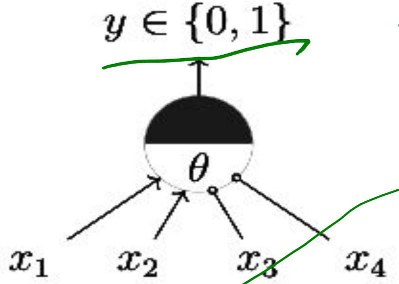
$x_3 = 0$

$x_4 = 0.$

$y = 1.$ is analogous to write

$f(\cdot) = 1.$

$y = f(\cdot) = 1.$



$\sum x_i \geq \theta.$

$2 \geq \theta.$

$\theta \leq 2$

θ value is min of

Lowest bound $= 2$

What should be the value of the threshold (θ) such that the MP neuron implements the above boolean function? (Note that the circle at the end of the input to the MP neuron indicates inhibitory input. If any inhibitory input is 1 the output will be 0.)

A. $\theta = 1$

B. $\theta = 2$

C. $\theta = 3$

D. $\theta = 4$

Keeping the concept discussed in question 1 in mind, consider the following boolean function:

$$f(x_1, x_2, x_3, x_4) = (x_1 \text{ OR } x_2) \text{ AND } (!x_3 \text{ AND } !x_4) = (x_1 + x_2) \cdot (\overline{x_3} \cdot \overline{x_4})$$

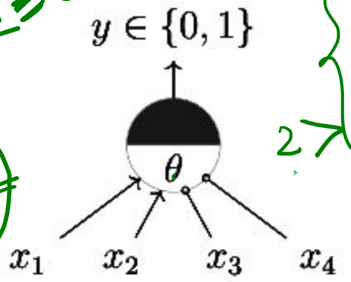
The MP neuron for the above boolean function is as follows:

$\sum_{\forall i} x_i \geq \theta.$

CASE 1 :→ $1 + 0 + 0 + 0 = 1 \geq \theta$

$y = 1.$

$\theta = 1.$



① $1 \ 0 \ 0 \ 0 \to f = 1;\ y = 1.$ $\quad 1 = 1(\theta)$

② $0 \ 1 \ 0 \ 0 \to f = 1.,\ y = 1$ $\quad 1 = 1(\theta)$

③ $1 \ 1 \ 0 \ 0 \to f = 1;\ y = 1$ $\quad 2 \geq 1(\theta)$

CASE 2 : $0 + 1 + 0 + 0 = 1 \geq \theta$

$\theta = 1.$

$\min\{\theta_1, \theta_2, \theta_3\}$

$= \min(1, 1, 2)$

$= 1.$

What should be the value of the threshold (θ) such that the MP neuron implements the above boolean function? (Note that the circle at the end of the input to the MP neuron indicates inhibitory input. If any inhibitory input is 1 the output will be 0.)

CASE 3 :– $1 + 1 + 0 + 0 = \underbrace{2 \geq \theta}_{\text{limiting value might be} = 2.}$

$\theta = 2$

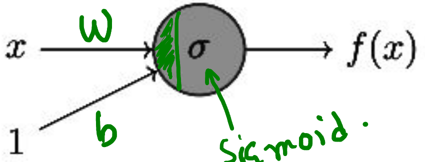
A. $\theta = 1$

B. $\theta = 2$

$\theta = 1$

③

$(y-f(x))f(x).(1-f(x))x$



$= \frac{1}{1+e^{-(wx+b)}}$

$\frac{\partial L}{\partial w}, \frac{\partial L}{\partial b} = ?$

$$f(x) = \frac{1}{1+e^{-(w \cdot x+b)}}$$

The value L is given by,

original output → Neural net. predicted output.

$$L = \frac{1}{2}(y - f(x))^2$$

→ squared error loss.

Here, x and y are constants and w and b are parameters that can be modified. In other words, L is a function of w and b.

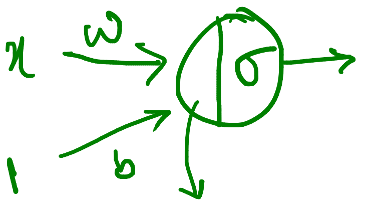
Derive the partial derivatives, $\frac{\partial L}{\partial w}$ and $\frac{\partial L}{\partial b}$ and choose the correct option.

A. $\frac{\partial L}{\partial w} = (y - f(x))f(x)(1 - f(x))$ and $\frac{\partial L}{\partial b} = (y - f(x))f(x)(1 - f(x))x$

B. $\frac{\partial L}{\partial w} = (y - f(x))(1 - f(x))x$ and $\frac{\partial L}{\partial b} = -(y - f(x))f(x)(1 - f(x))$

C. $\frac{\partial L}{\partial w} = -(y - f(x))f(x)(1 - f(x))x$ and $\frac{\partial L}{\partial b} = -(y - f(x))f(x)(1 - f(x))$

$z = wx+b.$



$f(z) = \sigma(z) = t$

$z = wx+b.$

$f(x) = \sigma(wx+b)$

$L = \frac{1}{2}(y - t)^2$

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial t}\,\frac{\partial t}{\partial z}\,\frac{\partial z}{\partial w}$$

sigmoid's derevative

$$= -(y-t) \times f(x)\,(1-f(x)) \times x$$

$$\frac{\partial L}{\partial t} = \frac{1}{\cancel{2}} \times \cancel{2} \times (y-t) \times \frac{d}{dt}(y-t)$$

$$= (y-t) \times (0-1)$$

$$= -(y-t)$$

$$\boxed{\frac{\partial L}{\partial w} = -(y-f(x)) \times f(x)\,(1-f(x)) \times x.}$$

$$\frac{\partial L}{\partial b} = \underbrace{\frac{\partial L}{\partial t}\,\frac{\partial t}{\partial z}\,\frac{\partial z}{\partial b}}$$

$$= -(y-f(x))\, f(x)\,(1-f(x)) \cdot 1$$

Consider the function E as given below,

tanh.

$\hat{z} = c(ax+by)+dz.$

$E = f(\hat{z}) = \tanh(\hat{z})$

$$E = g(x, y, z) = f(c(ax + by) + dz)$$

$\frac{\partial E}{\partial a} = ?$

Represented as a graph, we have

$E = f(\hat{z})$ $\quad \hat{z} = c(ax+by)+dz.$

$\frac{\partial E}{\partial \hat{z}} =$

$\frac{\partial E}{\partial a} = \frac{\partial E}{\partial \hat{z}} \frac{\partial \hat{z}}{\partial a}$

Here x, y, z are inputs (constants) and a, b, c, d are parameters (variables). m is an intermediate computation and f is some differentiable function. Specifically, let us consider f to be the *tanh* function.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

deravative of tanh function

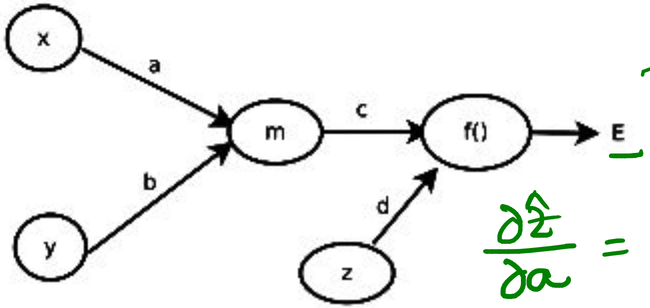
$\tanh(\hat{z}) = \frac{e^{\hat{z}} - e^{-\hat{z}}}{e^{\hat{z}} + e^{-\hat{z}}}$

Note that here E is a function of a, b, c, d. Compute the following partial derivatives of E with respect to a i.e $\frac{\partial E}{\partial a}$, and choose the correct option.

A. $\frac{\partial E}{\partial a} = (1 - f(c(ax + by) + dz)^2)cx$

B. $\frac{\partial E}{\partial a} = c(1 - f(c(ax + by) + dz)^2)$

C. $\frac{\partial E}{\partial a} = (1 - f(c(ax + by) - dz)^2)cx$



$\frac{\partial \hat{z}}{\partial a} = cx$

$$E = \tanh(\hat{z}) = \left(\frac{e^{\hat{z}} - e^{-\hat{z}}}{e^{\hat{z}} + e^{-\hat{z}}}\right)$$

$$\frac{\partial E}{\partial z} = \frac{\partial}{\partial \hat{z}}\left(\frac{e^{\hat{z}} - e^{-\hat{z}}}{e^{\hat{z}} + e^{-\hat{z}}}\right)$$

$$= \frac{(e^{\hat{z}} + e^{-\hat{z}})(e^{\hat{z}} + e^{-\hat{z}}) - (e^{\hat{z}} - e^{-\hat{z}})(e^{\hat{z}} - e^{-\hat{z}})}{(e^{\hat{z}} + e^{-\hat{z}})^2}$$

$$\frac{d}{dz} e^{-\hat{z}} = e^{-\hat{z}} \cdot (-1) = -e^{-\hat{z}}$$

$$e^{-\hat{z}} \xrightarrow{d/d\hat{z}} e^{-\hat{z}} \times (-1)$$

$$= \frac{(e^{\hat{z}} + e^{-\hat{z}})^2 - (e^{\hat{z}} - e^{-\hat{z}})^2}{(e^{\hat{z}} + e^{-\hat{z}})^2} = 1 - \left(\frac{e^{\hat{z}} - e^{-\hat{z}}}{e^{\hat{z}} + e^{-\hat{z}}}\right)^2$$

tanh function

$$= \frac{(e^{\hat{z}} + e^{-\hat{z}})^2}{(e^{\hat{z}} + e^{-\hat{z}})^2} - \frac{(e^{\hat{z}} - e^{-\hat{z}})^2}{(e^{\hat{z}} + e^{-\hat{z}})^2} = 1 - [\tanh(z)]^2$$

$$= 1 - f^2(z)$$

$$\hat{z} = c(ax + by) + dz.$$

$$\frac{\partial E}{\partial a} = \frac{\partial E}{\partial \hat{z}} \frac{\partial \hat{z}}{\partial a}$$

$$= (1 - f^2(\hat{z})) \cdot cx$$

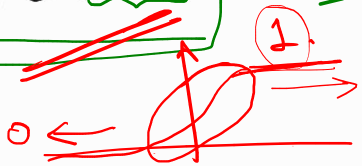
$$= [1 - f^2(c(ax+by)+dz)] \cdot cx.$$

3. The logistic function is defined as follows,

$\frac{y_2-y_1}{x_2-x_1} = \frac{\Delta y}{\Delta x} = \text{slope} \rightarrow \frac{dy}{dx}.$

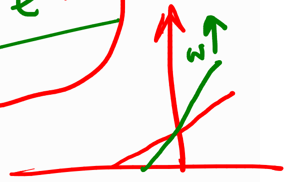
$$\boxed{f(x) = \frac{1}{1+e^{-(wx+b)}}} \rightarrow \text{sigmoid}$$

$w\uparrow$



$\frac{1}{1+e^{-x}}$

where w and b are parameters.
What would happen if w increases?

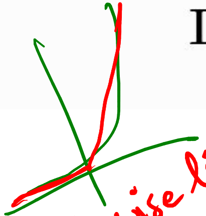


A. The slope of the logistic function increases

B. The slope of the logistic function decreases

C. The centre point of the logistic function moves to the left

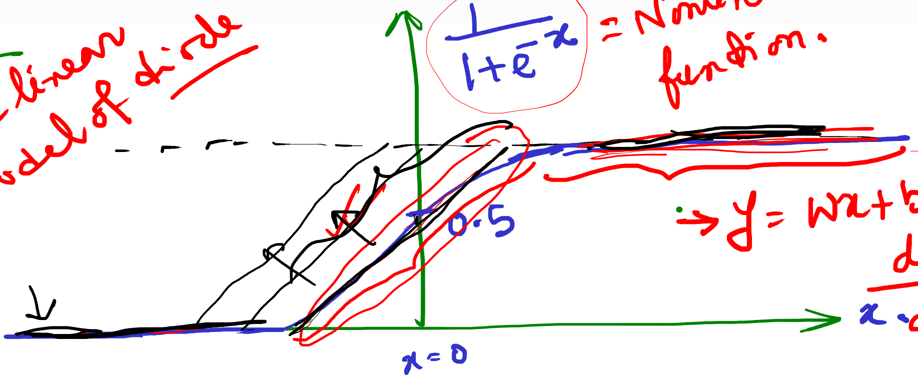
D. The centre point of the logistic function moves to the right



piecewise linear model of circle

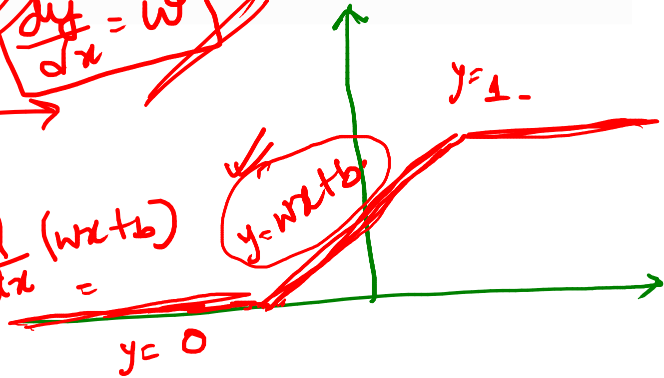
$\frac{1}{1+e^{-x}}$ = Nonlinear function.

$\boxed{\frac{dy}{dx} = w}$



$\rightarrow y = wx+b.$

$\frac{dy}{dx} = \frac{d}{dx}(wx+b)$



3. The logistic function is defined as follows,

$$f(x) = \frac{1}{1+e^{-(wx+b)}}$$

where w and b are parameters.

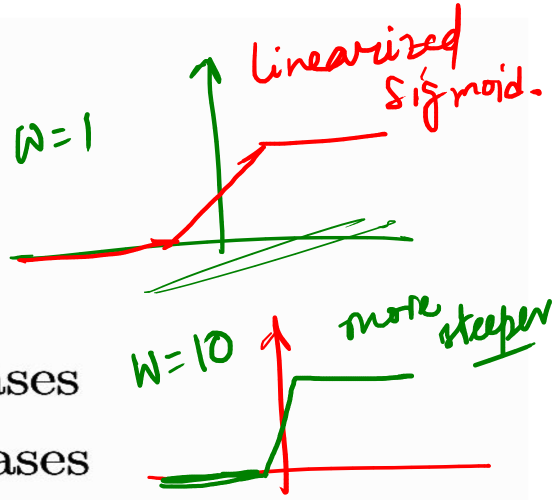
What would happen if w increases?

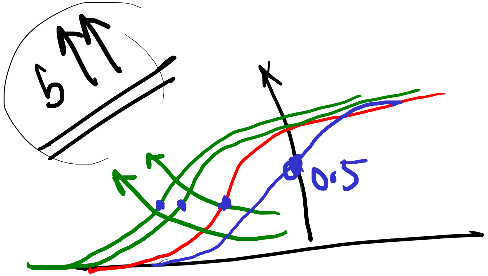
A. The slope of the logistic function increases

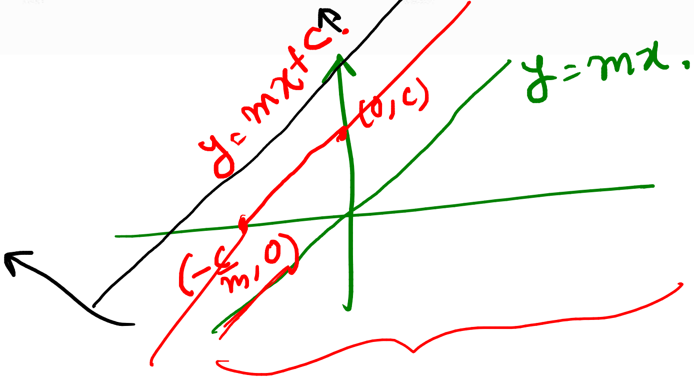
B. The slope of the logistic function decreases

C. The centre point of the logistic function moves to the left

D. The centre point of the logistic function moves to the right







$y = mx + c.$

$mx - y = -c.$

$\Rightarrow \frac{x}{\left(-\frac{c}{m}\right)} + \frac{y}{c} = 1.$

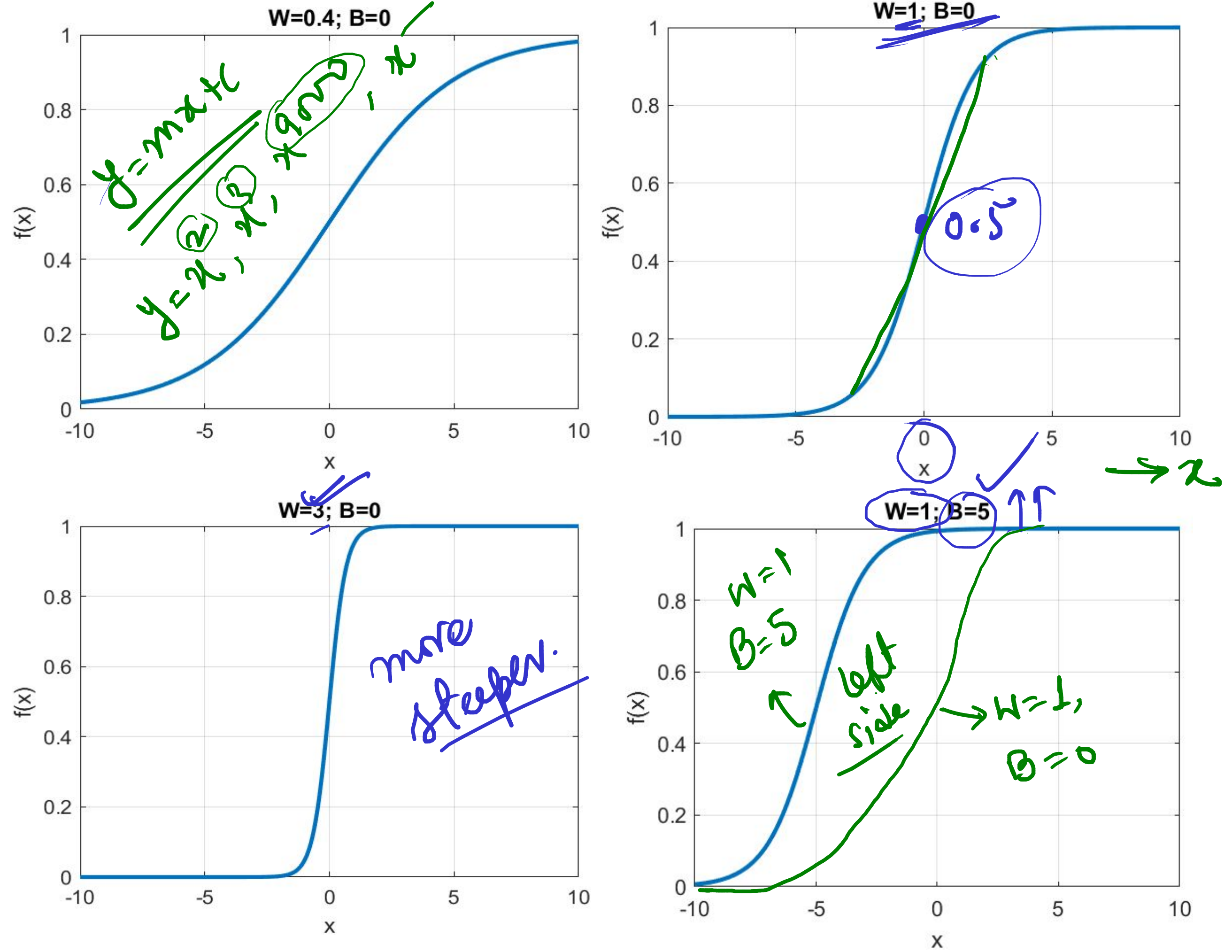
4. Keeping in mind the logistic function defined in question 3, what would happen if b increases?

A. The slope of the logistic function increases

B. The slope of the logistic function decreases

C. The centre point of the logistic function moves to the left

D. The centre point of the logistic function moves to the right



$$\frac{1}{1+e^{-(wx+b)}}$$

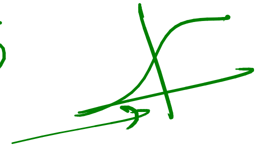
$w=1;\ b=0$

$$\frac{1}{1+e^{-x}}$$

$x \in (-\infty, \infty)$

W=1
B=1

$$\frac{1}{1+e^{-(\ )}}$$



$$\frac{1}{1+e^{-(w}} = f($$

Consider the variable x and functions $h_{11}(x)$, $h_{12}(x)$ and $h_{21}(x)$ such that

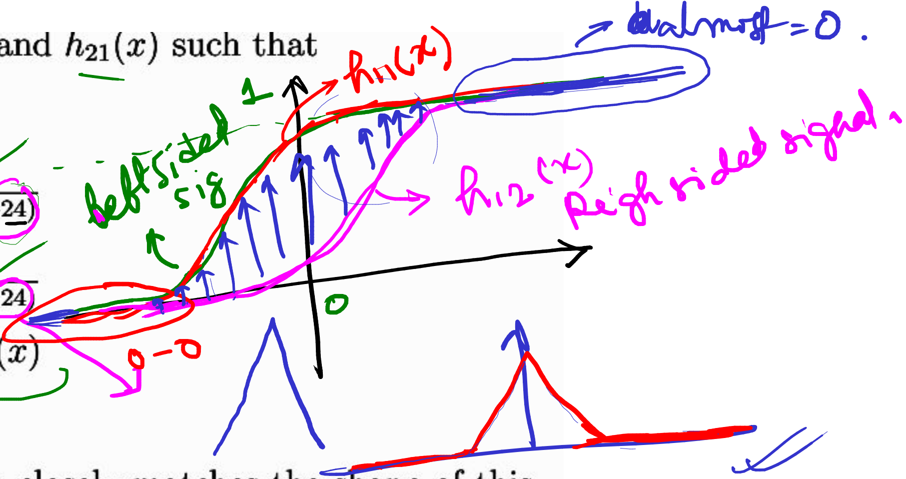
$y = \frac{1}{1+e^{-x}}$ $\left\{ \begin{array}{l} x \in (-\infty, \infty) \\ y \in [0,1] \end{array} \right.$

$y = \frac{1}{1+e^{-(wx+b)}}$

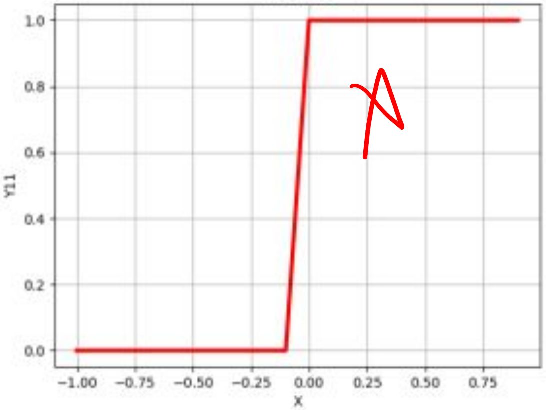
$$h_{11}(x) = \frac{1}{1+e^{-(400x+24)}}$$

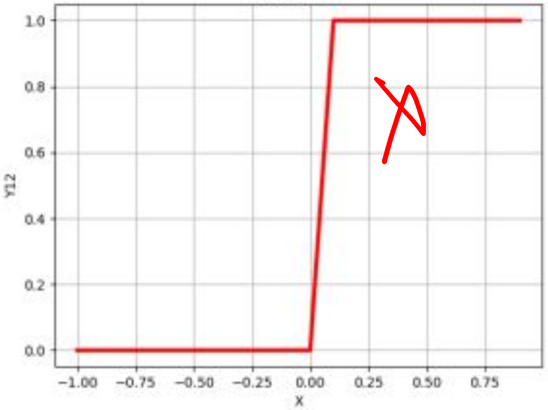
$$h_{12}(x) = \frac{1}{1+e^{-(400x-24)}}$$

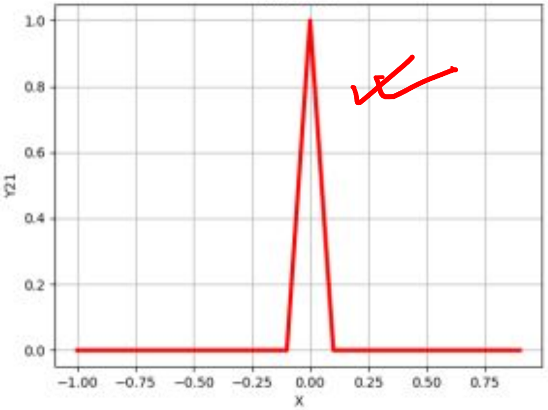
$$\underbrace{h_{21}(x) = h_{11}(x) - h_{12}(x)}$$



Plot the function $h_{21}(x)$ and choose the option which closely matches the shape of this function for $x \in (-1, 1)$







Consider the function $f(\theta) = f(x, y, z) = x^2 + y^2 + z^2 - 8$. Suppose you start with $\theta_0 = \{1, -1, 1\}$ and run one step of gradient descent with the learning rate $\eta = 1$. What will be the updated value of θ?

A. [1, -1, 1]

B. [1, -1, -1]

C. [-1, -1, 1]

D. [-1, 1, -1]

$f(\theta) = f(x, y, z) = x^2 + y^2 + z^2 - 8.$

$x_1 \leftarrow x_0 - \eta \frac{\partial f}{\partial x}.$

$= 1 - 1 \times (2x)\Big|_{x_0 = 1}$

$= 1 - 2$

$= -1$

$(x_0, y_0, z_0) = \{1, -1, 1\}.$

$y_1 \leftarrow y_0 - \eta \frac{\partial f}{\partial y}$

$z_1 \leftarrow z_0 - \eta \frac{\partial f}{\partial z}$

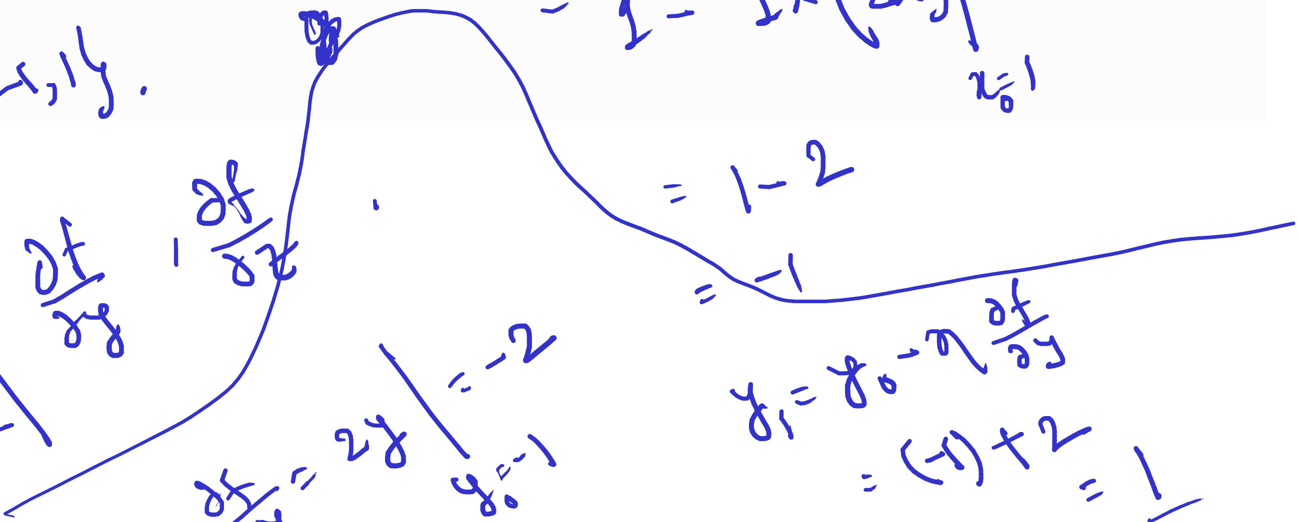
$\frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$

$\frac{\partial f}{\partial y} = 2y\Big|_{y_0 = -1} = -2$

$y_1 = y_0 - \eta \frac{\partial f}{\partial y}$

$= (-1) + 2$

$= 1$



Consider a box which contains 100 balls of which 30 are red, 50 are green and 20 are blue. Your friend peeps into the box and estimates the number of red, green and blue balls as 50, 25, 25. What is the cross entropy between the true distribution over the colors in the box and the distribution predicted by your friend?

A. 1.5

B. 1.0

C. 1.7

D. 2.3

$$\text{CCE loss} = -\sum_{\forall i} P_{true}^{(i)} \log P_{pred}^{(i)}$$

$$= -\left[0.3 \log_{10} 0.5 + 0.5 \log 0.25 + 0.2 \log 0.25\right].$$

$$= 1.7$$

Consider a vector $a \in \mathbb{R}^n$ and let $b \in \mathbb{R}^n$ be the output of the softmax function applied to this vector. The i-th entry of b is given by:

$\frac{1}{1+e^{-(w_1x_1+w_2x_2)}}$ softmax function

$$b_i = \frac{e^{a_i}}{\sum_{j=1}^{n} e^{a_j}}$$

$\sum b_i = 1 \to$ Probability.

$b_i > 0$.

$b_i \to$ Probability measure

Now suppose we introduce a parameter k such that

$ka_i \to \hat{a}_i$

$$\hat{b}_i = \frac{e^{\hat{a}_i}}{\sum e^{\hat{a}_j}}$$

$$b_i = \frac{e^{k*a_i}}{\sum_{j=1}^{n} e^{k*a_j}}$$

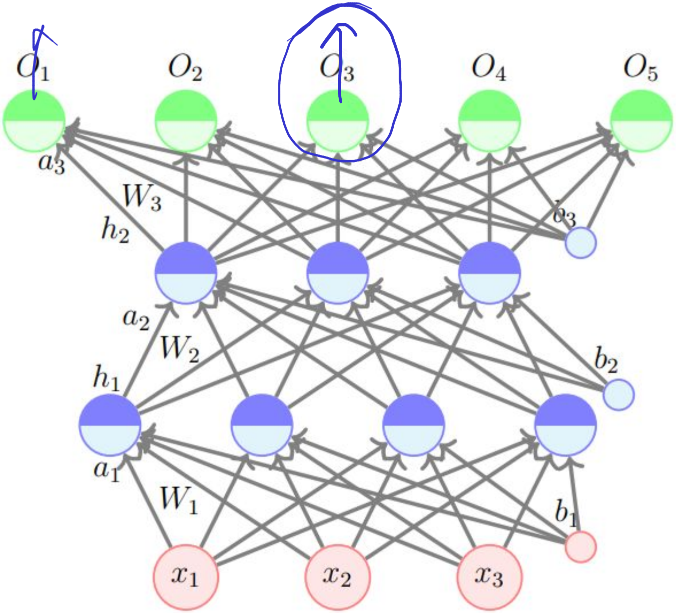
The updated function will also represent same nature as softmax

Can b still represent a probability distribution ?

A. True

B. False

Hence $\hat{b}_i$ will also have same prob. properties



$x = [1.5, 2.5, 3]$

$b_1 = [0.1, 0.2, 0.3, 0.4]$

$b_2 = [5.2, 3.2, 4.3]$

$b_3 = [0.2, 0.45, 0.75, 0.55, 0.95]$

all the weights in layer 1 are 0.05,

all the weights in layer 2 are 0.025.

all the weights in layer 3 are 1.0,

what will be the value of O_3

A. 0.132

B. 0.189

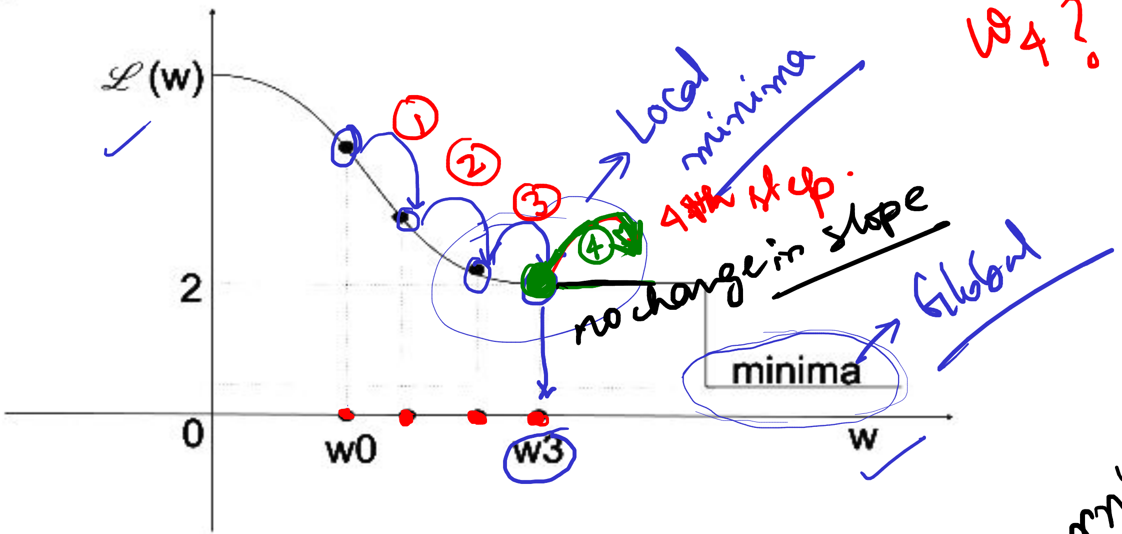
C. 0.229

D. 0.753

Consider the loss function $\mathscr{L}(w)$ as shown in the figure below. You are interested in finding the minima of this function *i.e.*, the value(s) of w for which the function will take its lowest value. To do so you run gradient descent starting with a random value w_0 (the leftmost red dot in the figure). After running, three steps of gradient descent you have the updated value of w as w_3. The red dots in the figure show the value of w at each step and the blue dots show the corresponding value of the loss function $\mathscr{L}(w)$. Now, what will happen if you run the 4th step of gradient descent, *i.e.*, if you try to update the value of w using the gradient descent update rule. Assume that the learning rate is 1.

$$w_4 = w_3 - \eta \frac{\partial \mathscr{L}(w)}{\partial w} \Big|_{w=w_3}$$

w_4 ?



$$\frac{\partial L}{\partial w} = 0$$

$w_4 = w_3$

(Trapped in local minima)

A. the value of w will increase (*i.e.*, $w_4 > w_3$)

B. the value of w will remain the same (*i.e.*, $w_4 = w_3$) → Learning will be stopped

C. the value of w will decrease (*i.e.*, $w_4 < w_3$)

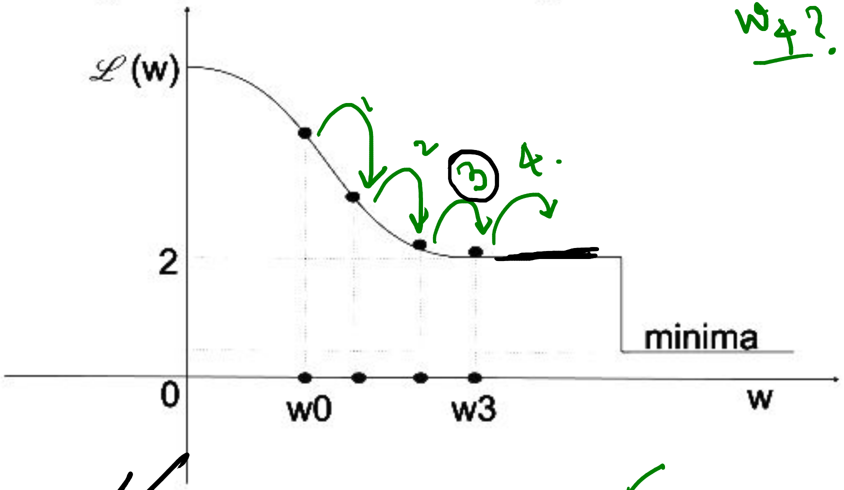
Continuing the previous question and referring to the same figure again, suppose instead of gradient descent you ran 3 iterations of momentum based gradient descent resulting in the value w_3 as shown in the figure. Note that the update rule of momentum based gradient descent is:

→ optimizers.

$$update_t = \gamma \cdot update_{t-1} + \eta \nabla w_t$$
$$w_{t+1} = w_t + update_t$$

Assume that the learning rate is 1 and the momentum parameter $\gamma > 0$. Now, what will happen if you run the 4th step of gradient descent, *i.e.*, if you try to update the value of w using the update rule of momentum based gradient descent.

w_4?



A. the value of w will increase ($i.e.$, $w_4 > w_3$)

B. the value of w will remain the same ($i.e.$, $w_4 = w_3$)

C. the value of w will decrease ($i.e.$, $w_4 < w_3$)

$update_t = \gamma \cdot update_{t-1} + \eta \nabla w_t$.

$\underbrace{w_{t+1}}_{4} = w_3 + update_t$.

$= w_3 + \gamma \cdot update_{t-1}$

$\gamma > 0$.

$w_4 > w_3$

Suppose we choose a model $f(x) = \sigma(wx + b)$ which has two parameters w, b. Further, assume that we are trying to learn the parameters of this model using 200 training points. If we use mini-batch gradient descent with a batch size of 10 then how many times will each parameter get updated in one epoch.

A. 10

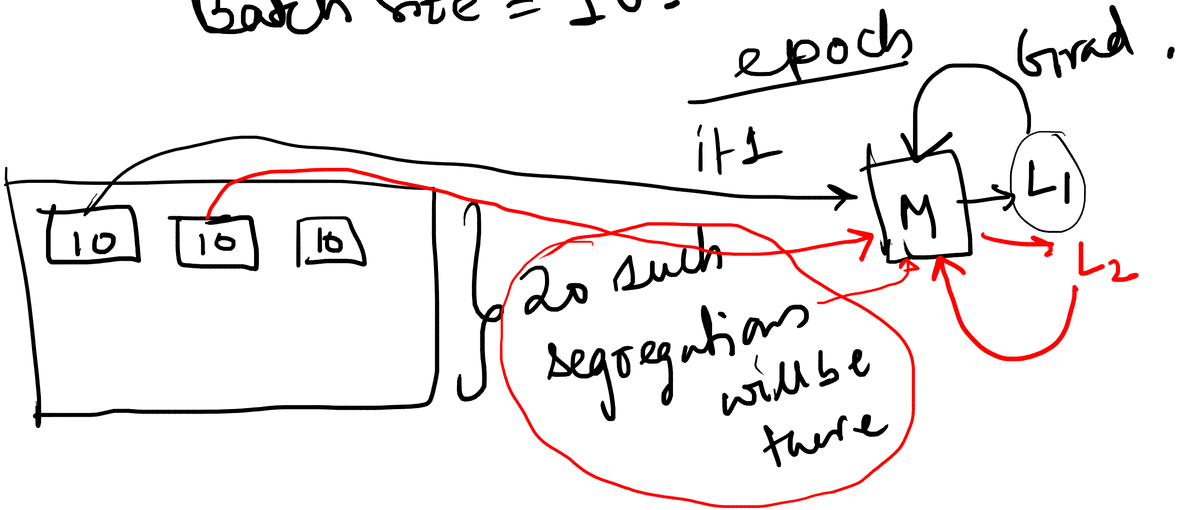
B. 20

C. 100

D. 200

No of Training Images = 200.

Batch size = 10.



Consider the following matrix:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}_{3\times 3}$$

$|A - \lambda I| = 0$ order 3 charact

$\lambda_1, \lambda_2, \lambda_3$ (eigen values).

then you get the eigen vectors.

Which of the following vectors is not an eigenvector of this matrix ?

A. $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $x_1 \to EV$ $Ax = \lambda x$

B. $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ x_2

C. $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ x_3

D. $\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ x_4

$$\begin{bmatrix} & & \end{bmatrix}_{3\times 3} \cdot \begin{bmatrix} \end{bmatrix}_{3\times 1} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3\times 1}$$

$Ax_1 = \lambda_1 x_1$

$Ax_2 = \lambda_2 x_2$

(A) $$\underbrace{\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{A \cdot x} = \begin{bmatrix} 1+1+2 \\ 1+2+1 \\ 2+1+1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

$= 4 \cdot x.$

$= \lambda \cdot x$

Consider the following matrix:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

Which of the following vectors is not an eigenvector of this matrix ?

A. $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ → ev.

B. $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ → ev

C. $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ → ev.

D. $\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$

$Ax = \lambda x$

scalar value

d

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1+1+0 \\ -1+2+0 \\ -2+1+0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

$\cancel{= \lambda x}$

not valid $\cancel{Ax = \lambda x}$ for option D.

$Ax = \lambda x$

$$\lambda = 1 \qquad 1 \cdot \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = (-1) \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \lambda . x .$$

Which of the following sets of vectors **does not** form a valid basis in $\mathbb{R}^3$

A. $\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}-1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\-2\\1\end{bmatrix}$

B. $\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}3\\2\\1\end{bmatrix}, \begin{bmatrix}1\\4\\5\end{bmatrix}$

C. $\begin{bmatrix}1\\3\\1\end{bmatrix}, \begin{bmatrix}4\\-1\\2\end{bmatrix}, \begin{bmatrix}6\\5\\4\end{bmatrix}$

D. $\begin{bmatrix}3\\5\\7\end{bmatrix}, \begin{bmatrix}1\\2\\2\end{bmatrix}, \begin{bmatrix}4\\7\\4\end{bmatrix}$

Basis → Vector quantites which are linearly independent with each other.

→ Linear Algebra !!

$\underbrace{\vec{v}_1, \vec{v}_2, \vec{v}_3}$

$a_1\vec{v}_1 + a_2\vec{v}_2 + a_3\vec{v}_3 = 0.$

only $a_1 = a_2 = a_3 = 0$ are zero in each.

→ Then you can not represent ~~write~~ them in any sort of linear combination.

$\begin{array}{c|cc} & A & B \\ \hline A & (A,A) & (A,B) = 0 \\ B & (B,A) & (B,B) \end{array}$

(A,B) → ~~statistically~~ independent

A	B
Gender	profession.

$v_1 + v_2 = v_3$

$v_2 + v_3 = 0$

→ $\begin{bmatrix}1\\0\\0\end{bmatrix}^A, \begin{bmatrix}0\\1\\0\end{bmatrix}^B, \begin{bmatrix}0\\0\\1\end{bmatrix}^C$

→ linearly independent

Which of the following sets of vectors **does not** form a valid basis in $\mathbb{R}^3$

A. $\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}-1\\0\\1\end{bmatrix}, \begin{bmatrix}1\\-2\\1\end{bmatrix}$

B. $\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}3\\2\\1\end{bmatrix}, \begin{bmatrix}1\\4\\5\end{bmatrix}$

C. $\begin{bmatrix}1\\3\\1\end{bmatrix}, \begin{bmatrix}4\\-1\\2\end{bmatrix}, \begin{bmatrix}6\\5\\4\end{bmatrix}$

D. $\begin{bmatrix}3\\5\\7\end{bmatrix}, \begin{bmatrix}1\\2\\2\end{bmatrix}, \begin{bmatrix}4\\7\\4\end{bmatrix}$

$$2\begin{bmatrix}1\\3\\1\end{bmatrix} + 1\begin{bmatrix}4\\-1\\2\end{bmatrix} - 1\begin{bmatrix}6\\5\\4\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix} = 0$$

$a_1 = 2$

$a_2 = 1$

$a_3 = -1$

$\{a_1, a_2, a_3\} \neq 0.$

$a_1\vec{v}_1 + a_2\vec{v}_2 + a\vec{v}_3 = 0$

$a_1 = a_2 = a_3 = 0$

linear independe

A [ ] B [ ]

$C = ?$

C [ ]

Consider a square matrix $A \in \mathbb{R}^{3\times 3}$ such that $A^T = A$. My friend told me that the following three vectors are the eigenvectors of this matrix A:

$$x = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, z = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

Is my friend telling the truth ?

A. Yes

B. No

C. Can't say without knowing all the elements of A

D. Yes, only if all the diagonal elements of A are 1

Midsem, IIT Patna

Symmetric Matrix.

$\underbrace{x, y, z}$ Mutually orthogonal with each.

Covariance Matrix? → symmetric Matrix

PCA → Principle compents.

A, B, C

$A \perp B$.

$A \perp B$, $B \perp C$, $C \perp A$

$A = $ symmetric.

$\{x, y, z\} \to$ eigen vectors.

$x \perp y \Rightarrow \vec{x} \cdot \vec{y} = 0.$

$\Rightarrow -1 + 1 + 1 \neq 0$

A model having a high bias with very few parameters will

Preconceived notion

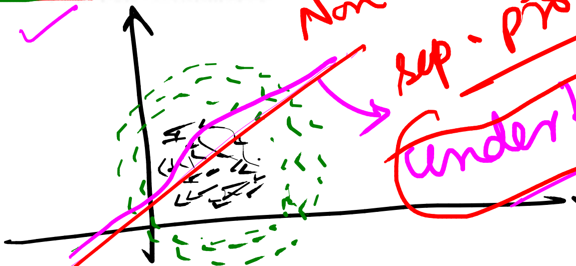
dumb

A. underfit the training data

B. overfit the training data

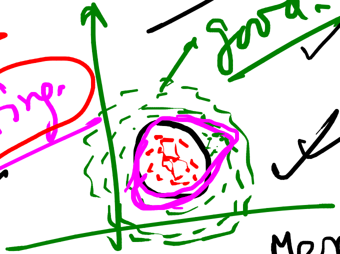
Non Linearly sep. prob.

underfitting



Generalize

good



Memorization

100%

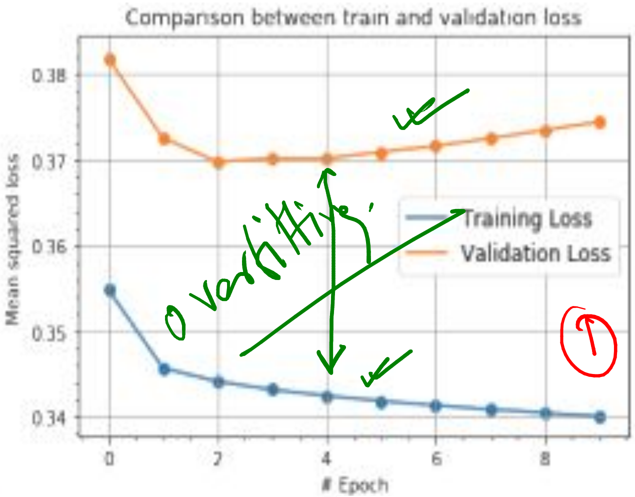
overfitting 30%

Is the following statements true: A complex model will be more sensitive to changes in training data than a simple model.

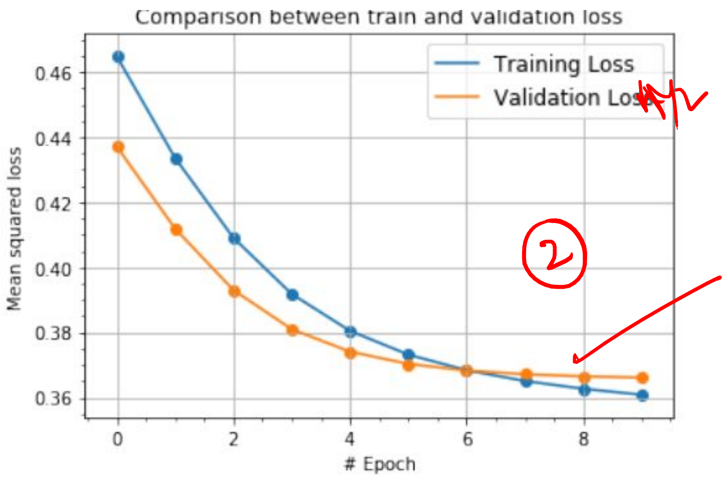
A. True

B. False

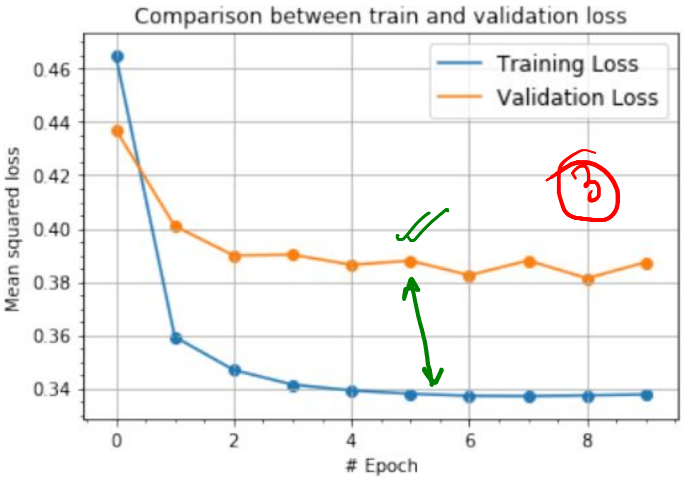
that your data is linearly separable


Comparison between train and validation loss
Mean squared loss
Epoch
Training Loss
Validation Loss
Overfitting
1

M1


Comparison between train and validation loss
Mean squared loss
Epoch
Training Loss
Validation Loss
2

M2


Comparison between train and validation loss
Mean squared loss
Epoch
Training Loss
Validation Loss
3

M3.

The maxout activation function is given by $\underbrace{f(x) = max(w_1^T x + b_1, w_2^T x + b_2)}$. You can obtain the ReLU function from this by setting:

$f(x) = max(w_1^t x + b_1, w_2^t x + b_2)$

A. $b_1 = b_2 = 0$

B. $w_1 = b_1 = b_2 = 0$, $w_2 = 1$

C. $w_1 = w_2 = 0$

$f(x) = max(0, x)$

$max(w_1^t x, w_2^t x)$

$w_1 = 0$

$b_1 = b_2 = 0$

$max(0, w_2^t x)$

$max(0, x)$

The batch normalization layer does not introduce any new parameters.

A. True

B. False

Scaling and shifting. (β, γ)

We can use backpropogation to train a deep neural network only if all the hidden layers in the network have the same activation function.

A. True

B. False

ReLU

gelu

tanh

$w_2 = \lambda$

$max(w_1^t x, w_2^t x)$

$x = 0$ $max(0, 0) = 0$

$max(-w_1^t, -w_2^t)$

What is the output dimension of the resulting image when a 7×7 kernel is applied to a 9×9 image?

$P=0, S=1.$

A. 3×3

B. 5×5

C. 2×2

$$\frac{9-7+(2\times 0)}{1}+1$$

$$= 2+1 = 3$$

3×3

Given a $128 \times 128 \times 3$ image and 6 filters of size $9 \times 9 \times 3$, what will be the dimension of the output volume when a stride of 1 and a padding of 2 is considered?

A. $124 \times 124 \times 6$

B. $119 \times 119 \times 6$

C. $121 \times 121 \times 6$

$P=2$

$S=1$

$I=128$

$K=9$

$f=6.$

$$O = \frac{128-9+(2\times 2)}{1}+1$$

$$= 128-4$$

$$= 124.$$

$124\times 124\times 6.$

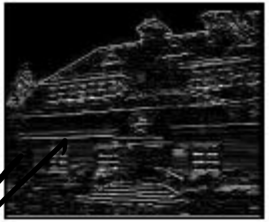
Consider the image shown below:

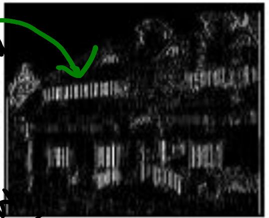
Figure 1: House Image

What will be the result of applying the following kernel to the above image ?

$$K = \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}$$

Horizontal edge detector.

A. 

B. 

Vertical edge detector

C. 

D. 

$$\begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}$$

Vertical edge

Image high frequ. info is embedded in Edges/Boundaries.

$-1 \quad 2 \quad -1$

high frequency information.

high frequ

brow

green

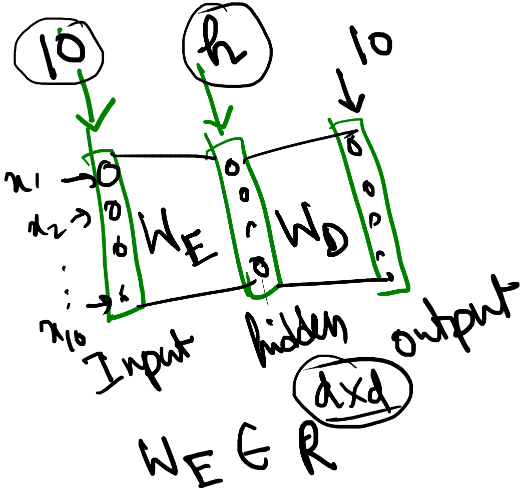
Consider an autoencoder which has one input layer, one hidden layer and one output layer. There is a weight connecting every neuron in the input layer to every neuron in the hidden layer and similarly there is a weight connecting every neuron in the hidden layer to every neuron in the output layer. There are no bias parameters in the network. Now, if the input to the network is $x \in \mathbb{R}^{10}$ and the total number of weights in the network is 160, then what kind of autoencoder is this ?

A. Overcomplete autoencoder

B. Undercomplete autoencoder

C. Can't say because the size of the hidden layer in not known

$x \in \mathbb{R}^{10 \times 1} \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ \vdots \\ x_{10} \end{bmatrix}$

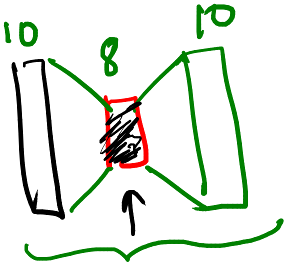


Total no of weight = 160.

$\Rightarrow (10 \times h) + (h \times 10) = 160.$

$\Rightarrow 20h = 160.$

$h = 8.$



Thank You