DS & AI

Machine Learning

SVM
1500+Series

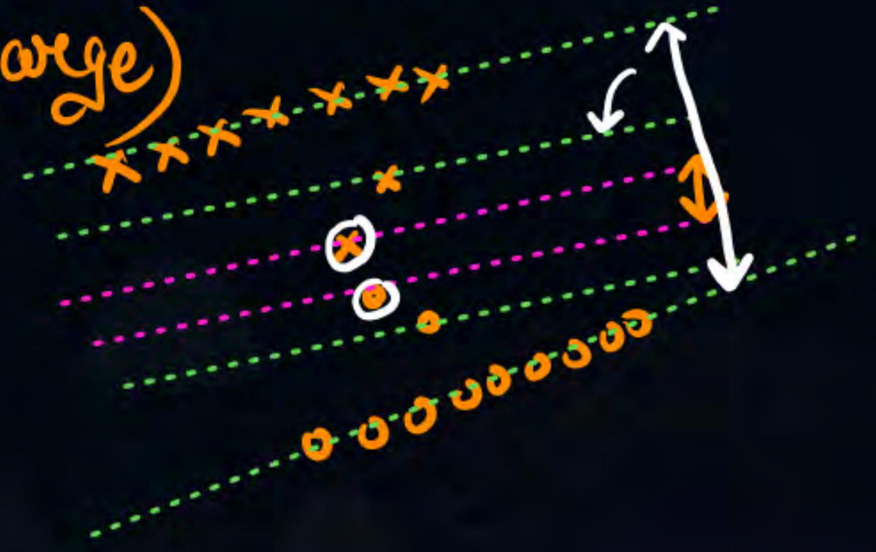
Lecture - [redacted]

By - Siddharth Sabharwal Sir

Topic : Support Vector Machine (SVM)

#Q. You are using a soft-margin SVM for binary classification. The regularization parameter (C) is set to 0.01. In the optimization problem, what is the effect of a small C value on the margin and misclassification tolerance?

(ξ can be large)



A Smaller margin and higher misclassification tolerance.

B Larger margin and lower misclassification tolerance.

C Smaller margin and lower misclassification tolerance.

D Larger margin and higher misclassification tolerance.

(d) ✓

Topic : Support Vector Machine (SVM)

#Q. You are using an RBF kernel in an SVM. The width parameter (gamma) is set to ~~0.1~~. What is the effect of increasing gamma on the SVM decision boundary? Small

Similar to 1NN

$e^{-\gamma \, |\text{distance b/w Point}|^2}$

A It results in a more flexible (complex) decision boundary.

B It results in a less flexible (simpler) decision boundary.

C It does not affect the decision boundary.

D The effect on the decision boundary depends on the value of C.

- γ inc, distance chota sa bhi $e^{-\gamma d^2} = 0$
- Closest points $e^{-\gamma |\text{distan}|^2} = 0$

Topic : Support Vector Machine (SVM)

#Q. You are training a support vector machine with a polynomial kernel. The kernel function is defined as $K(x, y) = \boxed{(x \cdot y + 1)^\wedge 2.}$ You want to calculate $K(3, 4)$. What is the value of $K(3, 4)$?

A 25

B 121

C 144

D $\boxed{169}$

Topic : Support Vector Machine (SVM)

#Q. You are using an RBF kernel in an SVM. The width parameter (gamma) is set to 0.1. What is the effect of increasing gamma on the SVM decision boundary?

done

A It results in a more flexible (complex) decision boundary.

B It results in a less flexible (simpler) decision boundary.

C It does not affect the decision boundary.

D The effect on the decision boundary depends on the value of C.

Topic : Support Vector Machine (SVM)

#Q. What is the primary objective of a Support Vector Machine (SVM)?

A. Minimize the number of support vectors.

B. Maximize the margin between classes.

C. Minimize the number of features.

D. Maximize the number of support vectors.

Topic : Support Vector Machine (SVM)

#Q. In SVM, what is the role of a kernel function?

A. It determines the regularization parameter.

B. It transforms data into a higher-dimensional space.

C. It computes the margin between classes.

D. It minimizes the number of support vectors.

Topic : Support Vector Machine (SVM)

#Q. If you have a linearly separable dataset with 100 data points, how many support vectors will an ideal SVM model have?

A 100

B 50

C 10

D 2

Topic : Support Vector Machine (SVM)

#Q. Which of the following is true regarding the regularization parameter (C) in SVM?

A. A higher C value leads to a larger margin. ✗

B. $\boxed{\text{A smaller C value allows for more misclassification.}}$ ✓

C. A higher C value allows for more misclassification. ✗

D. C has no effect on the SVM model. ✗

Topic : Support Vector Machine (SVM)

#Q. You have an imbalanced dataset with 90% of the data points in Class A and 10% in Class B. Which of the following is true about SVM performance in such cases?

A SVM is not suitable for imbalanced datasets.

B SVM tends to favor the minority class (Class B).

C SVM tends to favor the majority class (Class A).

D SVM equally balances the prediction performance between the classes.

Topic : Support Vector Machine (SVM)

#Q. Which type of SVM is most appropriate for handling classification tasks with noisy data?

A. Hard-margin SVM

B. $\boxed{\text{Soft-margin SVM}}$ ✓

C. Linear SVM

D. Kernel SVM

Topic : Support Vector Machine (SVM)

#Q. In an SVM with a polynomial kernel, if the degree of the polynomial is 2, what type of decision boundary is it likely to produce?

A. A linear decision boundary

B. $\boxed{\text{A quadratic decision boundary}}$

C. A higher-order polynomial decision boundary

D. An exponential decision boundary

Topic : Support Vector Machine (SVM)

#Q. In SVM, the term "support vectors" refers to:

A. Data points used to train the model.

B. Data points located far from the decision boundary.

C. Data points on the correct side of the decision boundary.

D. Data points closest to the decision boundary.

Topic : Support Vector Machine (SVM)

#Q. You have an SVM model with a linear kernel. The decision boundary is defined as $2x - 3y + 5 = 0$. You want to classify a point with coordinates (4, 7). What is the signed distance of the point to the decision boundary?

$$\frac{2\times4 - 3\times7 + 5}{\sqrt{2^2+3^2}} \Rightarrow \frac{-8}{\sqrt{13}} = -2.21$$

A 5

B -5

C 9

D -9

#Q. You are using an RBF kernel in an SVM. The width parameter (gamma) is set to 0.1. What is the effect of increasing gamma on the SVM decision boundary?

done ✓

A It results in a more flexible (complex) decision boundary

B It results in a less flexible (simpler) decision boundary.

C It does not affect the decision boundary

D The effect on the decision boundary depends on the value of C.

#Q. In SVM, what is the role of a kernel function?

A It determines the regularization parameter.

B It transforms data into a higher-dimensional space. ✓

C It computes the margin between classes.

D It minimizes the number of support vectors.

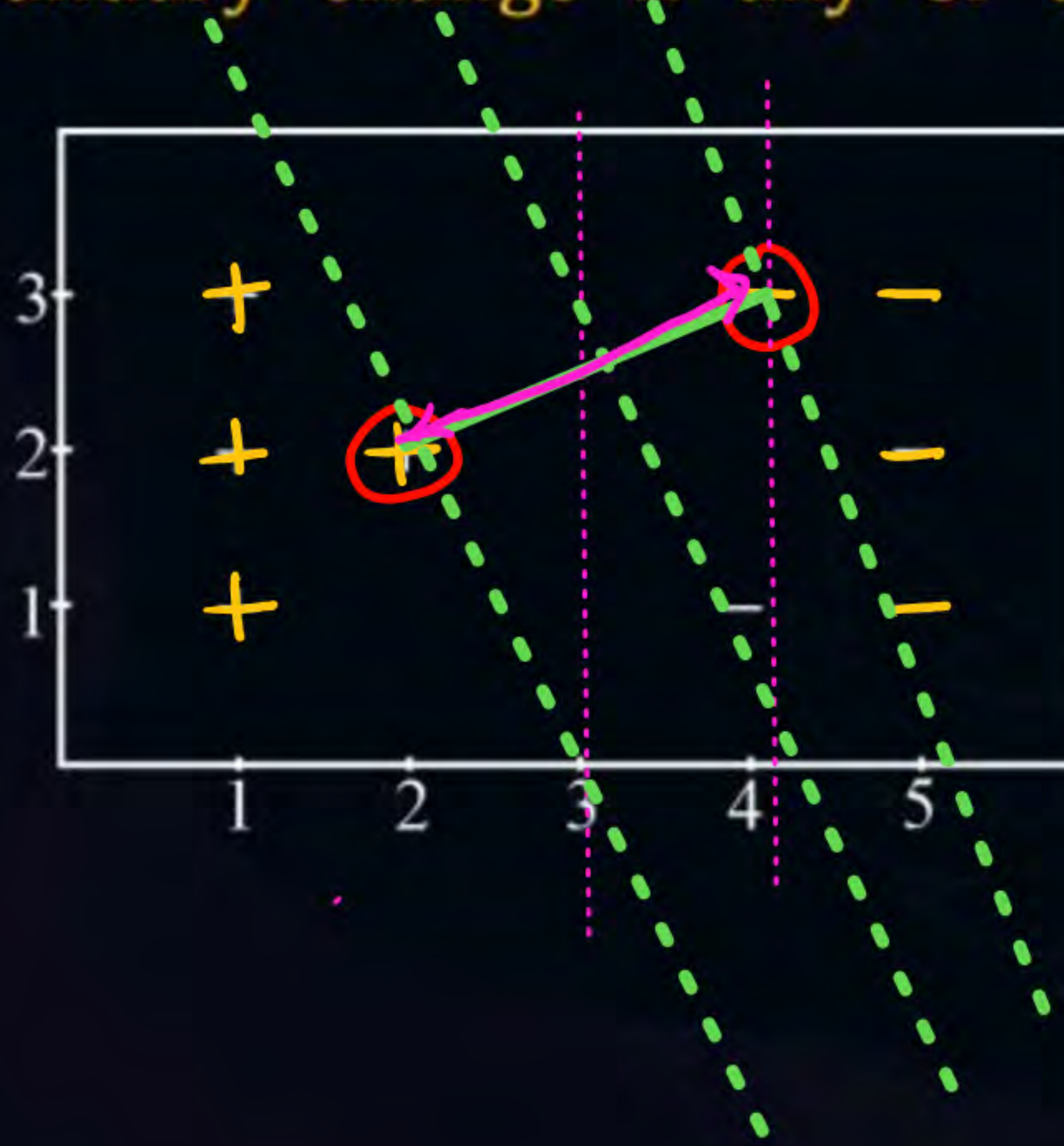
#Q. Suppose you are using a Linear SVM classifier with 2 class classification problem. Consider the following data in which the points circled red represent support vectors.

Will the decision boundary change if any of the red points are removed?

A Yes

B No

#Q. Suppose you are using a Linear SVM classifier with 2 class classification problem. Consider the following data in which the points circled red represent support vectors.
Will the decision boundary change if any of the red points are removed?



A Yes

B No

#Q. In the linearly non-separable case, what effect does the C parameter have on the SVM mode

A it determines how many data points lie within the margin

B it is a count of the number of data points which do not lie on their respective side of the hyperplane

C it allows us to trade-off the number of misclassified points in the training data and the size of the margin C ✓

D it counts the support vectors

#Q. Suppose that we use a RBF kernel with appropriate parameters to perform classification on a particular two class data set where the data is not linearly separable. In this scenario

A. the decision boundary in the transformed feature space is non-linear

B. $\boxed{\text{the decision boundary in the transformed feature space is linear}}$ ✓

C. the decision boundary in the original feature space is linear

D. $\boxed{\text{the decision boundary in the original feature space is non-linear}}$ ✓

$W^Tx+b=0$

#Q. In a hard margin SVM ~~$W^T + X + b$~~ $= 0$, suppose X_js are the support vectors and α_j s the corresponding Lagrange multipliers, then which of the following statements are correct:

A $W = \Sigma a_1 y_1 X_j$

B $\Sigma a_1 y_1 = 0$

C Either A or B

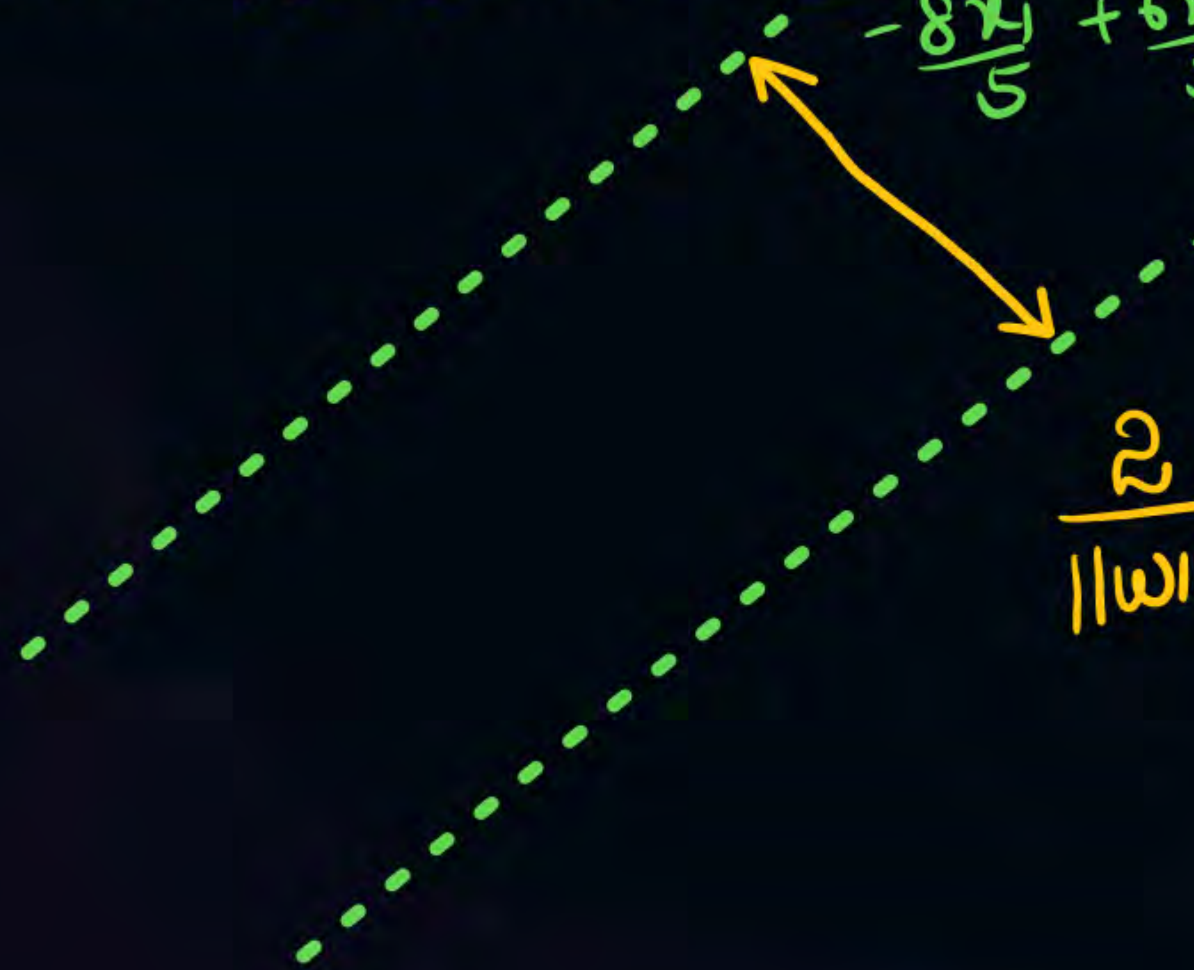
D Both A and B

#Q. In a support vector machine (SVM) for classification of the points x, let the hyperplanes be given as

$$-8x_1 + 6x_2 + 3 \geq 5$$

$$-8x_1 + 6x_2 + 3 \leq -5$$

The distance between the hyperplanes is given as



$$-\frac{8x_1}{5} + \frac{6x_2}{5} + 3/5 = 1$$

$$-\frac{8x_1}{5} + \frac{6x_2}{5} + 3/5 = -1$$

$$\frac{2}{\|w\|} = \frac{2}{\sqrt{\frac{8^2}{5^2} + \frac{6^2}{5^2}}}$$

$$= \frac{2}{\sqrt{100/25}} \Rightarrow 1$$

A. 2/10

B. 5

C. 1

D. 10

#Q. Suppose we have the below set of points with their respective classes as shown in the table. Answer the following question based on the table.

What will happen to maximum margin if we remove the point (–1, 0) from the training set?

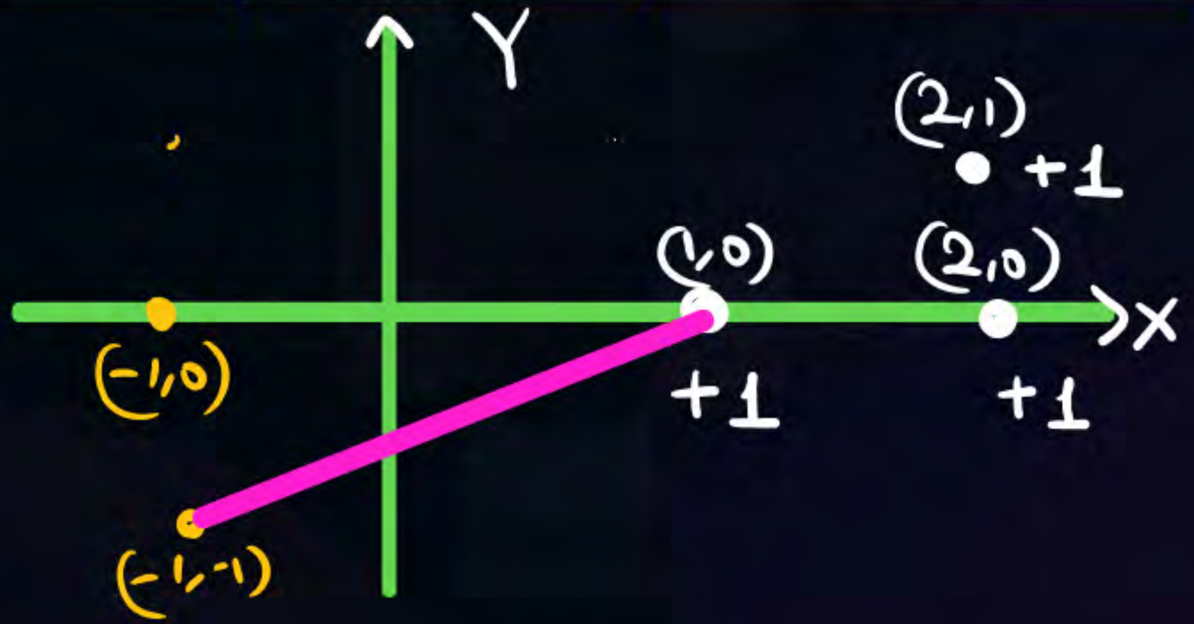
X	Y	Class Label
1	0	+1
–1	0	–1
2	1	+1
–1	–1	–1
2	0	+1

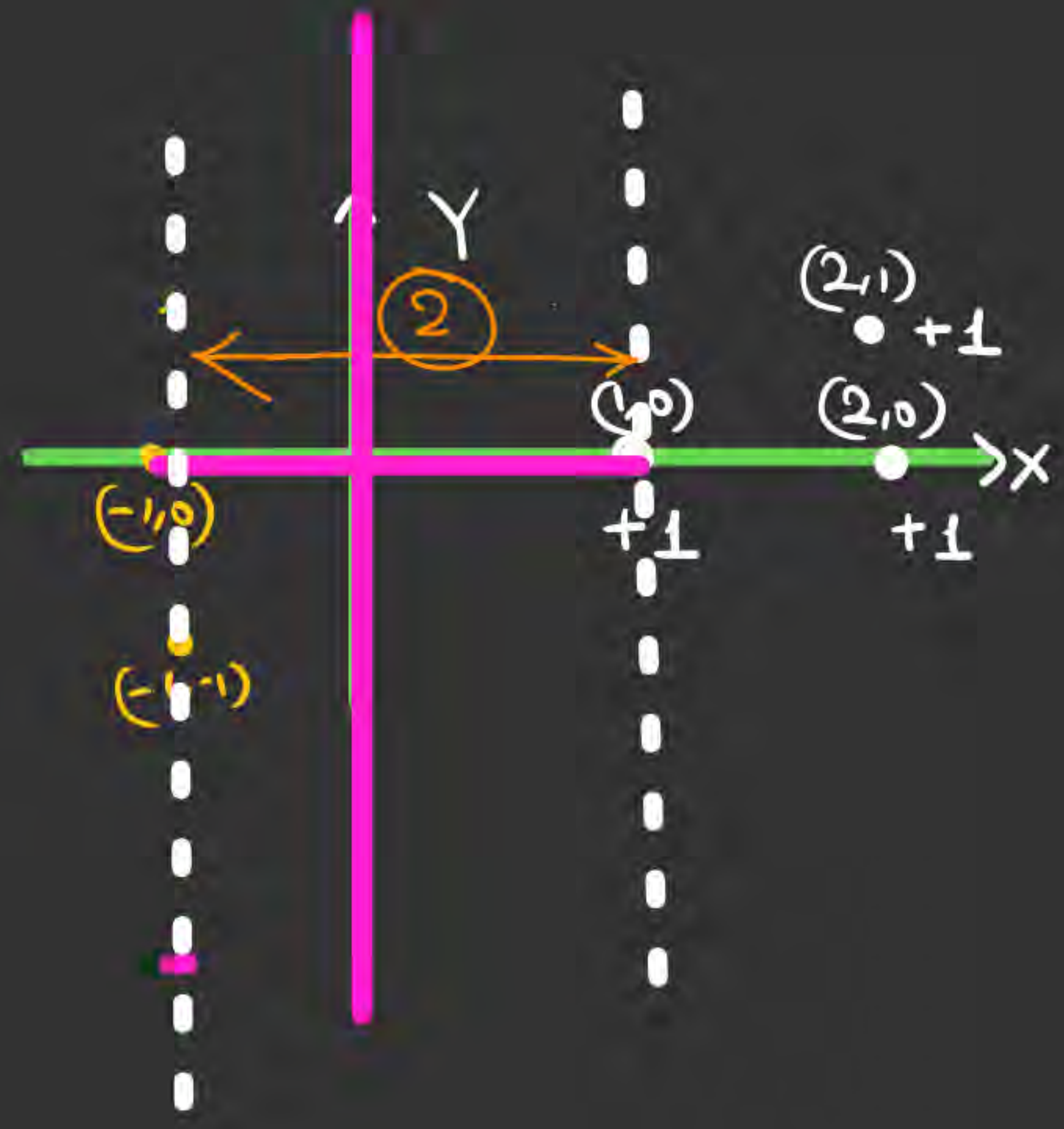
A Maximum margin will decrease

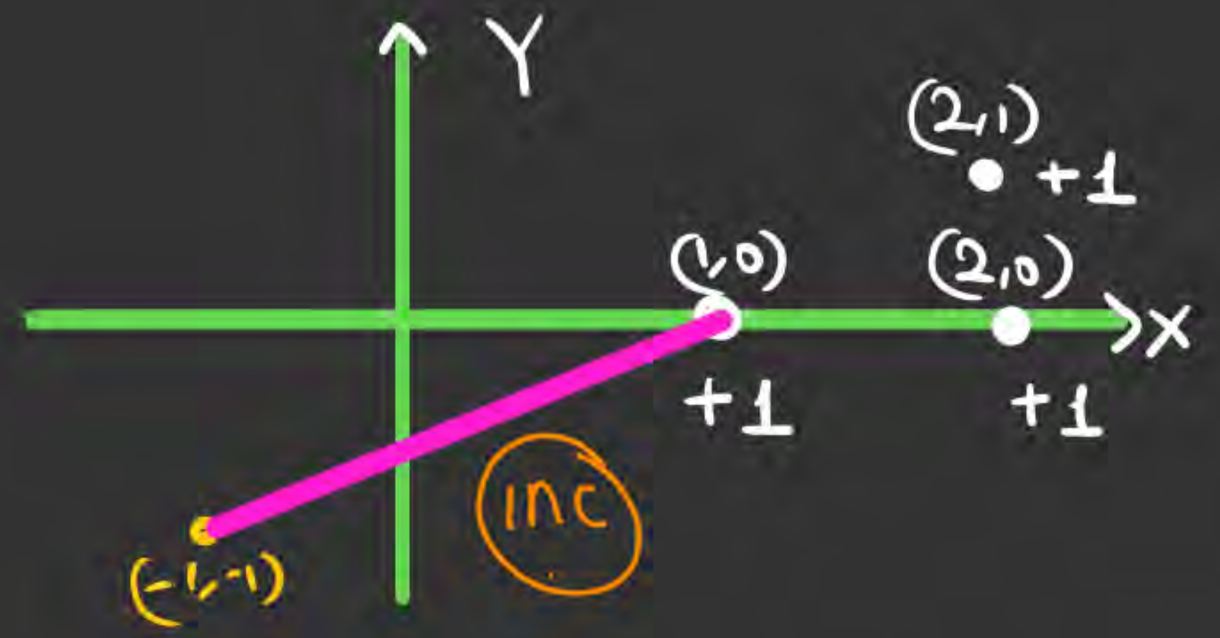
B $\boxed{\text{Maximum margin will increase}}$ ✓

C Maximum margin will remain same

D Can not decide




Y
2
(2,1)
+1
(1,0)
(2,0)
X
(-1,0)
+1
+1
(-1,-1)


Y
(2,1)
+1
(1,0)
(2,0)
X
+1
+1
inc
(-1,-1)

#Q. Suppose we have the below set of points with their respective classes as shown in the table. Answer the following question based on the table.

What can be a possible decision boundary of the SVM for the given points?

X	Y	Class Label
1	0	+1
+1	0	−1
2	1	+1
+1	−1	−1
2	0	+1

A. $y = 0$

B. $\boxed{x = 0}$

C. $x = y$

D. $x + y = 1$

#Q. Suppose we have the below set of points with their respective classes as shown in the table. Answer the following question based on the table.

Find the decision boundary of the SVM trained on these points and choose which of the following statements are true based on the decision boundary.

A, D

X	Y	Class Label
1	0	+1
–1	0	–1
2	1	+1
–1	–1	–1
2	0	+1

Class 0 ← | → class1

x=0

- **A** The point (–1, –2) is classified as –1
- **B** The point (1, –2) is classified as –1
- **C** The point (–1, –2) is classified as +1
- **D** The point (1, –2) is classified as +1

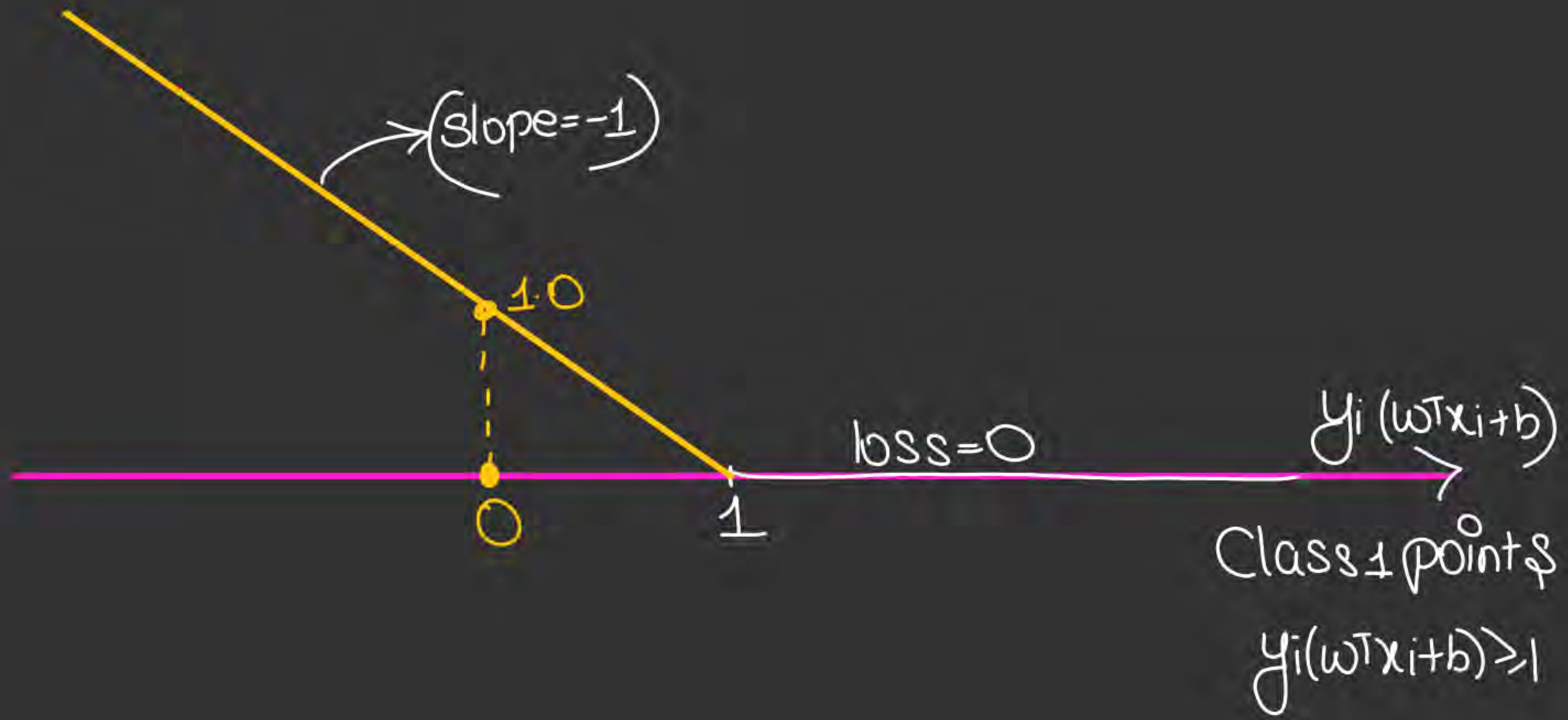
#Q. For a binary classification task, the decision function of an SVM is $f(x) = w \cdot x + b$, where $w = [2, -1]$ and $b = -1$. If a data point $x = [1, 1]$ belongs to the positive class ($y = +1$), what is the hinge loss?

- **A** 0
- **B** 0.5
- **C** 1.0
- **D** 1.5

$\underbrace{\text{Hingeloss function}}$

$(2x-y-1)$ classifier

$\underbrace{wx+b} \Rightarrow (2-1-1)=0$ @ classifier



(slope=-1)
1.0
0
1
loss=0
$y_i(w^T x_i+b)$
Class 1 points
$y_i(w^T x_i+b) \geq 1$

#Q. A soft-margin SVM is trained with a regularization parameter $C = 10$ If $\xi_1 = 2$, $\xi_2 = 1.5$ and $\xi_3 = 0$ for three misclassified points, what is the total penalty due to slack variables?

H.W

$\left(C\sum \xi_i\right)$

A. 20.0

B. 15.0

C. 3.5

D. 0.0

#Q. Given the kernel function $\boxed{K(x, z) = (1 + x \cdot z)^2}$ which of the following is the equivalent transformation in feature space?

Point x, z are of 2 dimensions.

A $[x_1^2, \sqrt{2}x_1x_2, x_2^2]$

B $[1, x_1^2, x_1x_2, x_2^2]$

C $\boxed{[1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2]}$

D $[1, x_1, x_2, x_1x_2]$

$K(x, z) = (1 + xz)^2$

- $K(x, z) = (1 + x_1z_1 + x_2z_2)^2$
- $K(x, z) = (1 + (x_1z_1)^2 + (x_2z_2)^2 + 2x_1z_1 + 2x_2z_2 + 2x_1z_1x_2z_2)$
- $\phi(x)\phi(z)^T$

$$[1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, \sqrt{2}x_1x_2, x_2^2]\begin{bmatrix}1\\ \sqrt{2}z_1\\ \sqrt{2}z_2\\ z_1^2\\ \sqrt{2}z_1z_2\\ z_2^2\end{bmatrix}$$

$$(1 + 2z_1x_1 + 2x_2z_2 + x_1^2z_1^2 + 2x_1x_2z_1z_2 + x_2^2z_2^2)$$

#Q. In a dual formulation of SVM, the optimization problem is:

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

Subject to $\sum_{i=1}^{N} \alpha_i y_i = 0$ and $0 \le \alpha_i \le C$.

If $C = 1$, $y = [1, -1]$ and the kernel matrix is

$$K = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix},$$

what are the optimal α_1 and α_2?

A $\alpha_1 = 0.5, \alpha_2 = 0.5$

B $\alpha_1 = 1.0, \alpha_2 = 0.0$

C $\alpha_1 = 0.0, \alpha_2 = 1.0$

D $\alpha_1 = 1.0, \alpha_2 = 1.0$ ✓

$$\alpha_1 + \alpha_2 - \frac{1}{2}(\alpha_1^2 + \alpha_2^2) + \frac{\alpha_1 \alpha_2}{2}$$

$$1 + 1 - \frac{1}{2}(2) + \frac{1}{2} \Rightarrow 3/2$$

$$\max_{\alpha} \left(\alpha_1 + \alpha_2 - \frac{1}{2} \sum_{i=1}^{2} \sum_{j=1}^{2} \alpha_i \alpha_j y_i y_j K(x_i, x_j)\right)$$

$$(\alpha_1 + \alpha_2) - \frac{1}{2}\left[\alpha_1^2 y_1^2 K(x_1, x_1) + \alpha_2^2 y_2^2 K(x_2, x_2) + \alpha_1 \alpha_2 y_1 y_2 K(x_2, x_1) + \alpha_2 \alpha_1 y_2 y_1 K(x_1 x_2)\right]$$

$$(\alpha_1 + \alpha_2) - \frac{1}{2}\left[\alpha_1^2 (1)(1) + \alpha^2 + 2\alpha_1 \alpha_2 (-1)(.5)\right]$$

$$\alpha_1 + \alpha_2 - \frac{1}{2}(\alpha_1^2 + \alpha_2^2 - \alpha_1 \alpha_2)$$

Topic : Unsupervised Learning & MLP

#Q. What is the function of the hidden layers in a multilayer perceptron (MLP)?

A. Hidden layers perform feature extraction and transformation

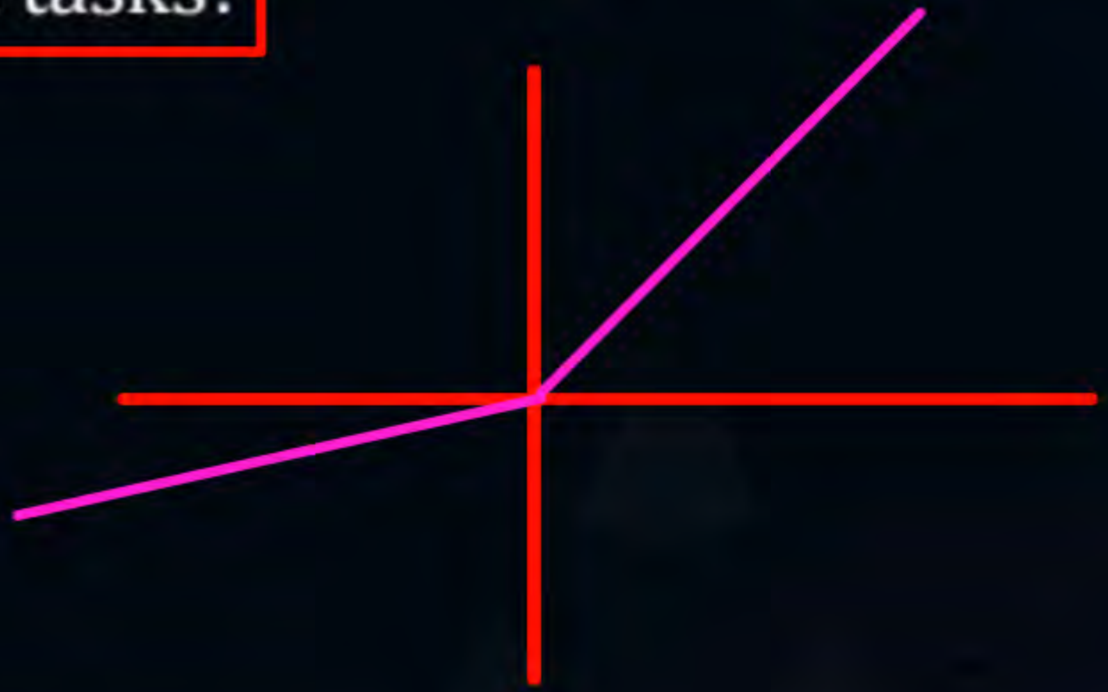
B. Hidden layers provide direct access to the input data

C. Hidden layers perform the final classification or regression task

D. Hidden layers are not necessary in an MLP

Topic : Unsupervised Learning & MLP

#Q. Which of the following activation function is commonly used in feed-forward neural networks for classification tasks?



A ~~Linear activation function~~

B Sigmoid activation function

C ReLU (Rectified Linear Unit) activation function

D Softmax activation function

Topic : Unsupervised Learning & MLP

#Q. How does a feedforward neural network differ from other types of neural networks?

Skip

NO

A Feedforward neural networks allow feedback connections between neurons.

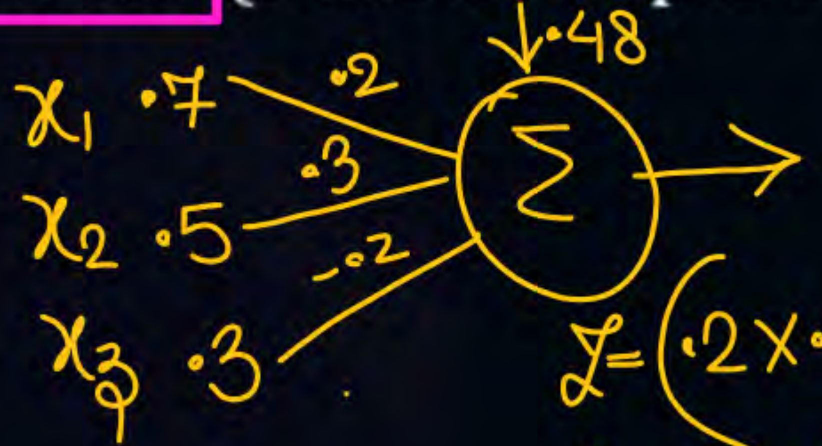
B Feedforward neural networks can only have one layer of neurons.

C Feedforward neural networks do not contain any hidden layers.

D In feedforward neural networks, information flows in one direction, from input to output, without cycles or loops.

Topic : Unsupervised Learning & MLP

#Q. Given an input layer with three nodes x_1 x_2 and x_3 having value 0.7,0.5 and 0.3 respectively, associated with weights 0.2,0.3 and -0.2. If the bias is 0.48 and a binary sigmoid activation function is used, find the output of the neural network? (Rounded upto 2 decimal)



$$Z = (.2 \times .7 + .3 \times .5 - .3 \times .2 + .48) = .71$$

$$f(z) = \left(\frac{1}{1+e^{-z}}\right) = \underline{.670}$$

Topic : Unsupervised Learning & MLP

#Q. Which of the following statements are true? B, C

A The chances of overfitting decreases with increasing the number of hidden nodes and increasing the number of hidden layers.

B A neural network with one hidden layer can represent any Boolean function given sufficient number of hidden units and appropriate activation functions.

C Two hidden layer neural networks can represent any continuous functions (within a tolerance) as long as the number of hidden units is sufficient and appropriate activation functions used.

D All of the above

#Q. Consider a simple neural network with one hidden layer having three neurons and an output layer with one neuron. The activation function used is the sigmoid function. The network is trained using back propagation. The input features are $x_1 = 0.8$, $x_2 = 0.6$ and the target output is $Y_{target} = 0.9$. The weights and biases are initialized as follows:

- Hidden layer weights: $w_{11} = 0.4$, $w_{21} = 0.2$, $w_{31} = -0.5$, $(w_{12} = w_{11}, w_{22} = w_{21}, w_{31} = w_{32})$
- Output layer weights: $w_{12} = 0.7$, $w_{22} = -0.3$, $w_{32} = 0.6$
- Hidden layer biases: $b_1 = 0.3$, $b_2 = -0.1$, $b_3 = 0.2$
- Output layer bias: $b_2 = -0.2$

Now, calculate the error term (delta) for the output neuron using the back propagation formula:

$$\delta_{output} = (y_{output} - y_{target}) \cdot \sigma'(z_{output})$$

Topic : Unsupervised Learning & MLP

#Q. Consider a simple neural network with one hidden layer having three neurons and an output layer with one neuron. The activation function used is the sigmoid function. The network is trained using back propagation. The input features are $x_1 = 0.8$, $x_2 = 0.6$ and the target output is $Y_{target} = 0.9$. The weights and biases are initialized as follows:

- Hidden layer weights: $w_{11} = 0.4$, $w_{21} = 0.2$, $w_{31} = -0.5$
- Output layer weights: $w_{12} = 0.7$, $w_{22} = -0.3$, $w_{32} = 0.6$
- Hidden layer biases: $b_1 = 0.3$, $b_2 = -0.1$, $b_3 = 0.2$
- Output layer bias: $b_2 = -0.2$

Now, calculate the error term (delta) for the output neuron using the back propagation formula:

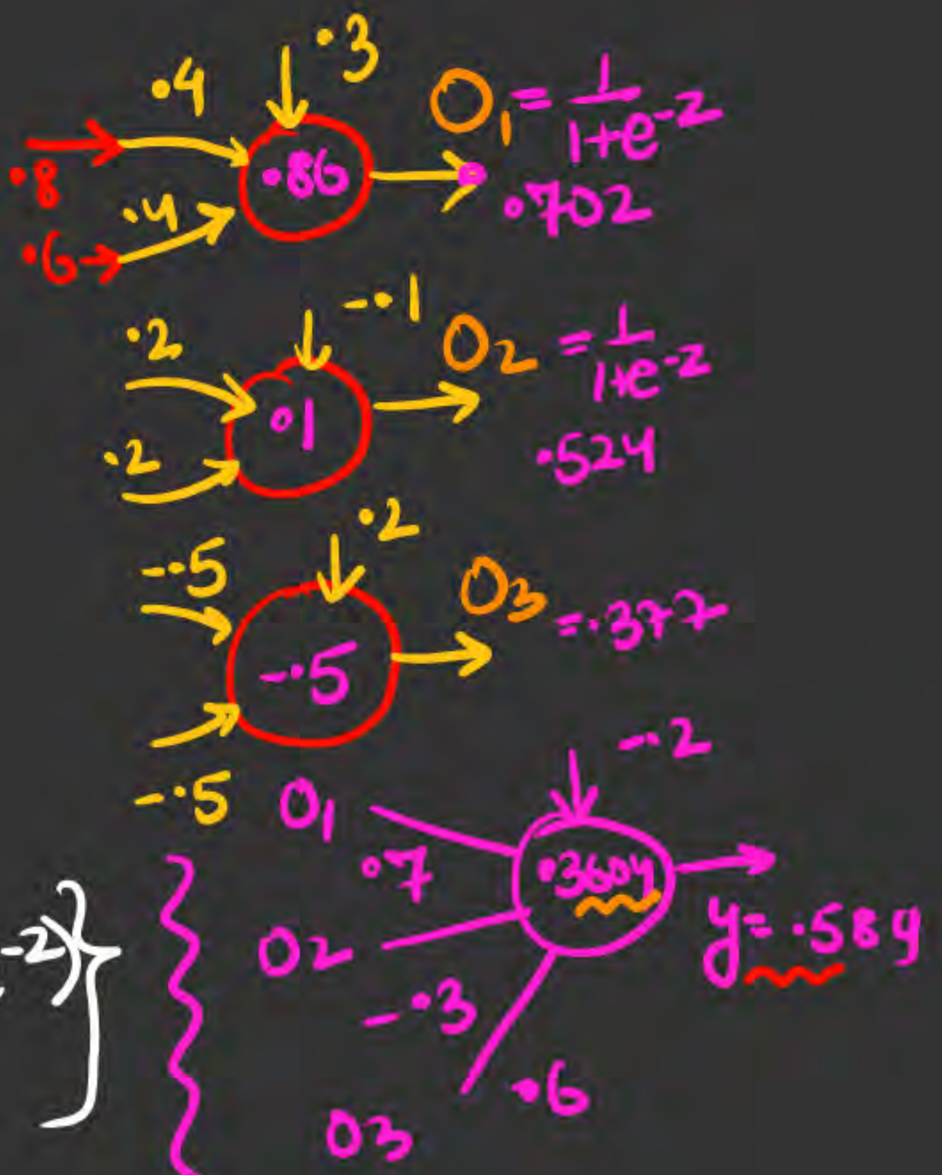
$$\boxed{\delta_{output} = (y_{output} - y_{target}) \cdot \sigma'(z_{output})}$$

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

$$\frac{\partial y_{out}}{\partial z} = \left\{\frac{-1}{(1+e^{-z})^2} \cdot (-)(e^{-z})\right\}$$

$$\Rightarrow \frac{e^{-z}}{(1+e^{-z})^2}$$

$$\Rightarrow .242$$



#Q. Consider a neural network with a single neuron in the input layer and a single neuron in the output layer. The weights connecting these neurons are given as follows:

$$w_1 = -2 \; w_2 = -5 \text{ and } w_3 = -4$$

Given an input vector X = [2, 3, 1] perform the forward propagation to calculate the output of the neural network.

What will be the weighted sum of input?

A. –23

B. –46

C. –20

D. +21

#Q. Consider a single perceptron with sign activation function. The perceptron is represented by weight vector $[0.4 - 0.3\ 0.1]^t$ and a bias $0 = 0$. If the input vector to the perceptron is $X = [0.2\ 0.6\ 0.5]$ then the output of the perceptron is :

A. 1

B. 0

C. –0.05

D. –1

Sign Function

#Q. Let W_{ij} represents weight between node i at layer k and node j at layer (k -1) of a given multilayer perceptron. The weight updation using gradient descent method is given by
Where a and E represents learning rate and Error in the output respectively.

A $W_{ij}(t+1) = W_{ij}(t) + \alpha \frac{\partial E}{\partial W_{ij}}, 0 \le \alpha \le 1$

B $W_{ij}(t+1) = W_{ij}(t) - \alpha \frac{\partial E}{\partial W_{ij}}, 0 \le \alpha \le 1$

C $W_{ij}(t+1) = \alpha \frac{\partial E}{\partial W_{ij}}, 0 \le \alpha \le 1$

D $W_{ij}(t+1) = -\alpha \frac{\partial E}{\partial W_{ij}}, 0 \le \alpha \le 1$

#Q. Consider a single perceptron with sign activation function. The perceptron is represented by weight vector $[0.4\ -0.3\ 0.1]^t$ and a bias $\theta = 0$. If the input vector to the perceptron is $X = [0.2\ \ 0.6\ \ 0.5]$ then the output of the perceptron is:

A. 1

B. 0

C. –0.05

D. –1

Topic : Unsupervised Learning & MLP

#Q. What is the function of the hidden layers in a multilayer perceptron (MLP)?

A. Hidden layers perform feature extraction and transformation

B. Hidden layers provide direct access to the input data

C. Hidden layers perform the final classification or regression task

D. Hidden layers are not necessary in an MLP

#Q. If a neural network with 4 hidden layers each having 50 neurons is trained with an input size of 20 and an output size of 10, how many parameters are there in the first hidden layer?

A. 1050

B. 1050 + 50

C. 1050 + 50

D. 1050 + 50 + 50

#Q. What is the number of parameters in a neural network with 10 input neurons, 10 neurons in the hidden layer, and 2 output neurons, assuming the hidden layer uses a bias term?

A 120

B 132

C 140

D 150

Topic : Unsupervised Learning & MLP

#Q. For a neural network with 5 layers, if each layer has 100 neurons, and each neuron has connections to all neurons in the previous layer, how many parameters are there in the network?

A) 40400

B) 50500

C) 51000

D) 50500 + 100

#Q. Given an input layer with three nodes x_1, x_2 and x_3 having value 0.7, 0.5 and 0.3 respectively, associated with weights 0.2, 0.3 and –0.2. If the bias is 0.48 and a binary sigmoid activation function is used, find the output of the neural network? (Rounded up to 2 decimal).

THANK - YOU