Powered by:

NPTEL Week 4 Live Sessions

on Deep Learning (noc24_ee04)

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

- Python coding for SVM, KNN
- Week 3 quiz solution
- Week 4 practice questions

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

Prime Minister's Research Fellows
Ministry of Education
Government of India

Content of the live session

1. Python coding for week-3 content ← → KNN → SVM.
2. Quiz solution of week 3
3. Solving numerical problems from week 4

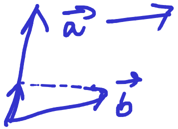
MNIST

CIFAR-10

IRIS

1. Find the scalar projection of vector b = <-3, 2> onto vector a = <1,1>?

   a. 0
   b. $\frac{1}{\sqrt{2}}$
   c. $\frac{-1}{\sqrt{2}}$
   d. $\frac{-1}{2}$



$\left(\frac{\vec{a}}{\|\vec{a}\|}\right) \cdot \vec{b}$ = scalar projection.

$= \frac{-3 \times 1 + 2 \times 1}{\sqrt{1^2 + 1^2}} = \frac{-1}{\sqrt{2}}$

3. Suppose there is a feature vector represented as [1, 4, 3]. What is the distance of this feature vector from the separating plane $x_1 + 2x_2 - 2x_3 + 3 = 0$. Choose the correct option.

   a. 1
   b. 5
   c. 3
   d. 2 ✓

$W^t x + b = 0$

$\vec{v} = [1, 4, 3]$

$\frac{\vec{w}}{\|\vec{w}\|} \cdot \vec{v} + b$ 1×3

Decision surface.

plane eq: $x_1 + 2x_2 - 2x_3 + 3 = 0$.

$W = [1 \ \ 2 \ \ -2]$ 1×3 $\quad b = 3$.

$$\text{Distance} = \frac{1 + (2 \times 4) - (2 \times 3) + 3}{\sqrt{1^2 + 2^2 + 2^2}} = \frac{1 + 8 - 6 + 3}{\sqrt{4 + 4 + 1}} = \frac{6}{3} = 2.$$

10 Find the distance of the 3D point, $\vec{P} = (-2, 4, 1)$ from the plane defined by $2x + 3y + 6z + 7 = 0$?

plane.

a. 3
b. 4
c. 0
d. ∞ (infinity)

$$\frac{\vec{w}\cdot\vec{p} + \text{bias}}{\|\vec{w}\|}$$

$\vec{w} = [2 \quad 3 \quad 6]$.

Bias $= 7$.

$$= \frac{(2\times -2) + (3\times 4) + (6\times 1) + 7}{\sqrt{2^2 + 3^2 + 6^2}}$$

$$= \frac{-4 + 12 + 6 + 7}{\sqrt{49}} = \frac{21}{7} = 3.$$

~~4 + 9 + 36~~

~~=49~~

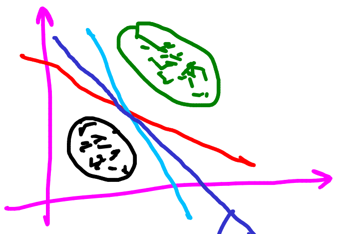
If we employ SVM to realize two input logic gates, then which of the following will be true? → Maximize the classification with maximized Margin value.

a. The weight vector for AND gate and OR gate will be same.

b. The margin for AND gate and OR gate will be same.

c. Both the margin and weight vector will be same for AND gate and OR gate.

d. None of the weight vector and margin will be same for AND gate and OR gate.



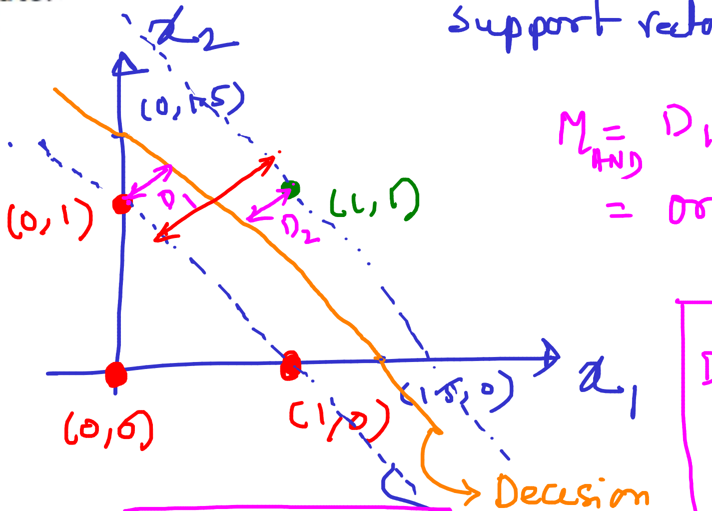
support vectors are = (1,1), (0,1), (1,0)

AND Gate :→

Feature x_1	Feature x_2	Label/class. y
0	0	0
0	1	0
1	0	0
1	1	1

(rows 1–3: 0; row 4: 1)

$W_{AND} = [1, 1, -1.5]$.



Decision Boundary

$\Rightarrow \boxed{x_1 + x_2 = 1.5}$

$\Rightarrow \frac{x_1}{1.5} + \frac{x_2}{1.5} = 1$

$x + y = 1$

$M_{AND} = D_1 + D_2$

= orthogonal dist (0,1) to Decision Surface + (1,1) to Decision surface

$$D_1 = \frac{|0+1-1.5|}{\sqrt{1^2+1^2}} = \frac{1}{2\sqrt{2}}$$

$$D_2 = \frac{|1+1-1.5|}{\sqrt{1^2+1^2}} = \frac{1}{2\sqrt{2}}$$

$$M = \frac{1}{\sqrt{2}}$$

If we employ SVM to realize two input logic gates, then which of the following will be true?

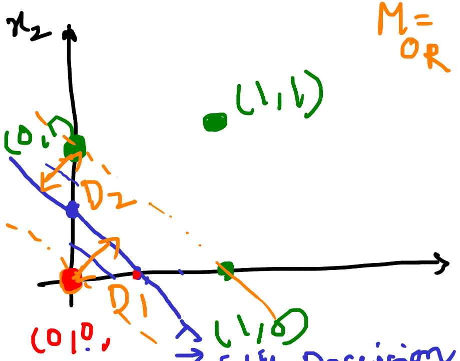
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

a. ~~The weight vector for AND gate and OR gate will be same.~~

b. ~~The margin for AND gate and OR gate will be same.~~

c. ~~Both the margin and weight vector will be same for AND gate and OR gate.~~

d. ~~None of the weight vector and margin will be same for AND gate and OR gate.~~



→ SVM Decision Boundary.

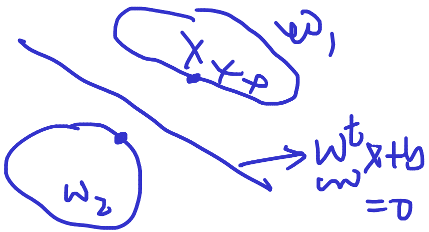
→ $x_1 + x_2 = 0.5$

$W_{OR} = [1, 1, -0.5]$

$$M_{OR} = D_1 + D_2 = \frac{|(0\times1) + (0\times1) - 0.5|}{\sqrt{1^2+1^2}} + \frac{|(0\times1) + (1\times1) - 0.5|}{\sqrt{1^2+1^2}}$$

$$= \frac{1}{2\sqrt{2}} + \frac{1}{2\sqrt{2}} = \frac{1}{\sqrt{2}}.$$

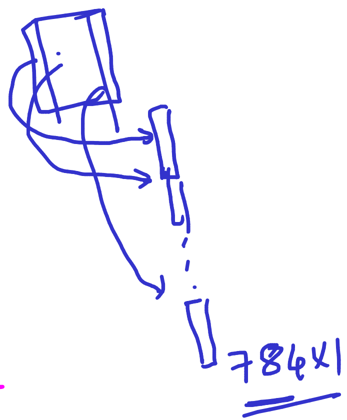
$\boxed{M_{AND} = M_{OR}}$ $\boxed{W_{OR} \neq W_{AND}}$



$\text{Margin} = \frac{2}{\|w\|}$

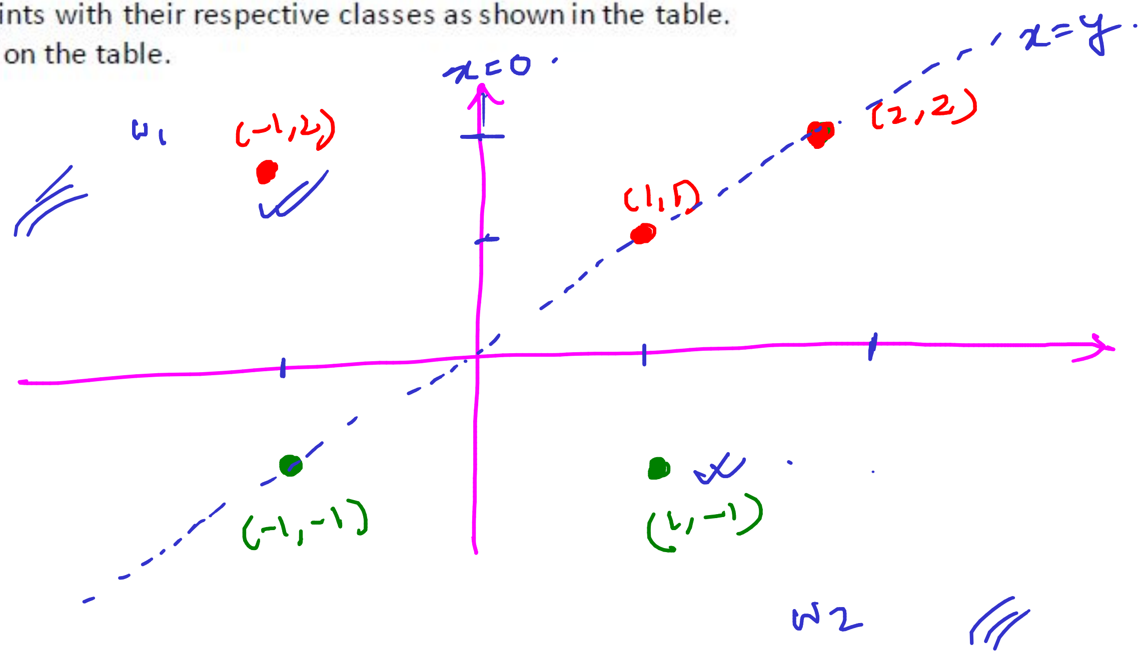
$y_i(w^t x_i + b) \geq 1$

28x28



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.

X	Y	Class Label
1	1	+1
-1	-1	-1
2	2	+1
-1	2	+1
1	-1	-1



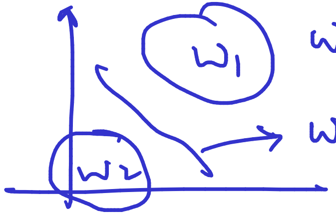
What can be a possible decision boundary of the SVM for the given points?

a. $y = 0$

b. $x = 0$

c. $x = y$

d. $x + y = 1$



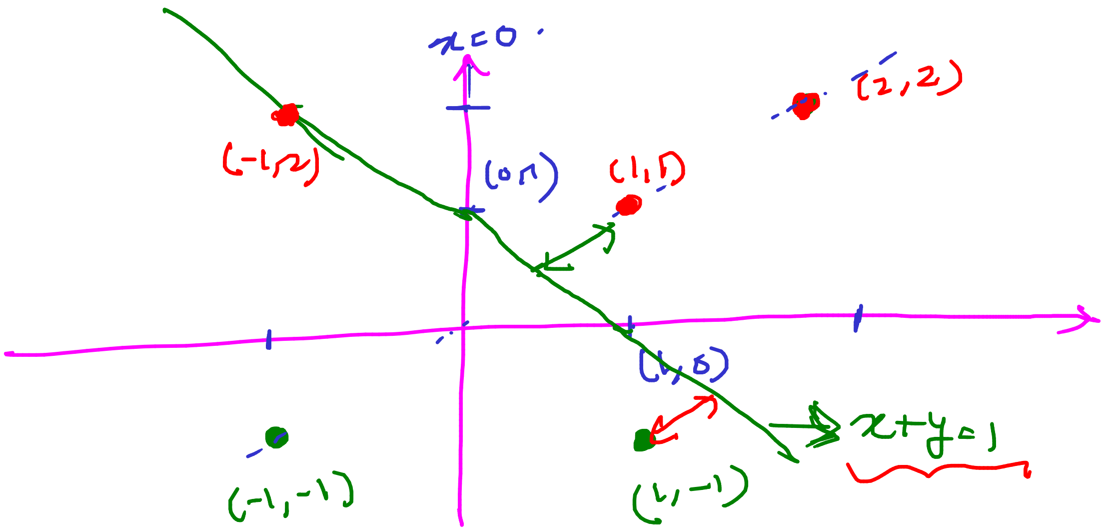
$w^t x + b > 0$

$w^t x + b = 0$.

$w^t x + b < 0$ → Can't be decided.

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.

X	Y	Class Label
1	1	+1
-1	-1	-1
2	2	+1
-1	2	+1
1	-1	-1



$$M = \frac{|x| + |x| - 1}{\sqrt{1^2+1^2}} + \frac{|1-1-1|}{\sqrt{1^2+1^2}}$$

$$= \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = \frac{2}{\sqrt{2}} = \sqrt{2}$$

$$= 1.414$$

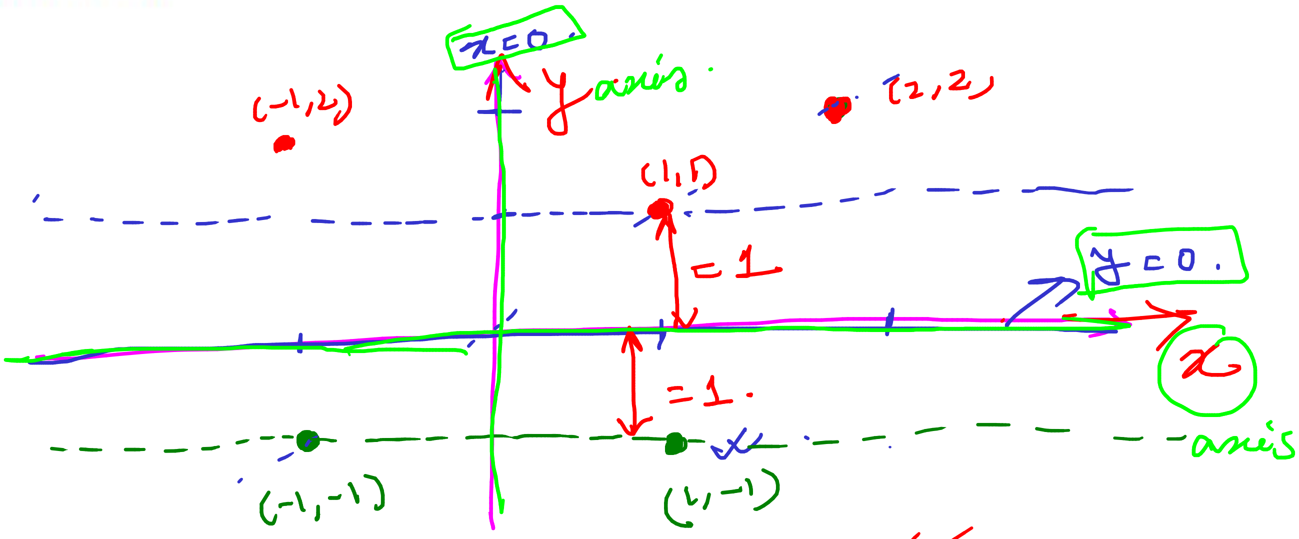
What can be a possible decision boundary of the SVM for the given points?

a. $y = 0$
b. $x = 0$
c. $x = y$
d. $x + y = 1$ ✓

$-1+2=1$

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.

X	Y	Class Label
1	1	+1
-1	-1	-1
2	2	+1
-1	2	+1
1	-1	-1



$M = 1 + 1 = 2.$

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

a. $y = 0$

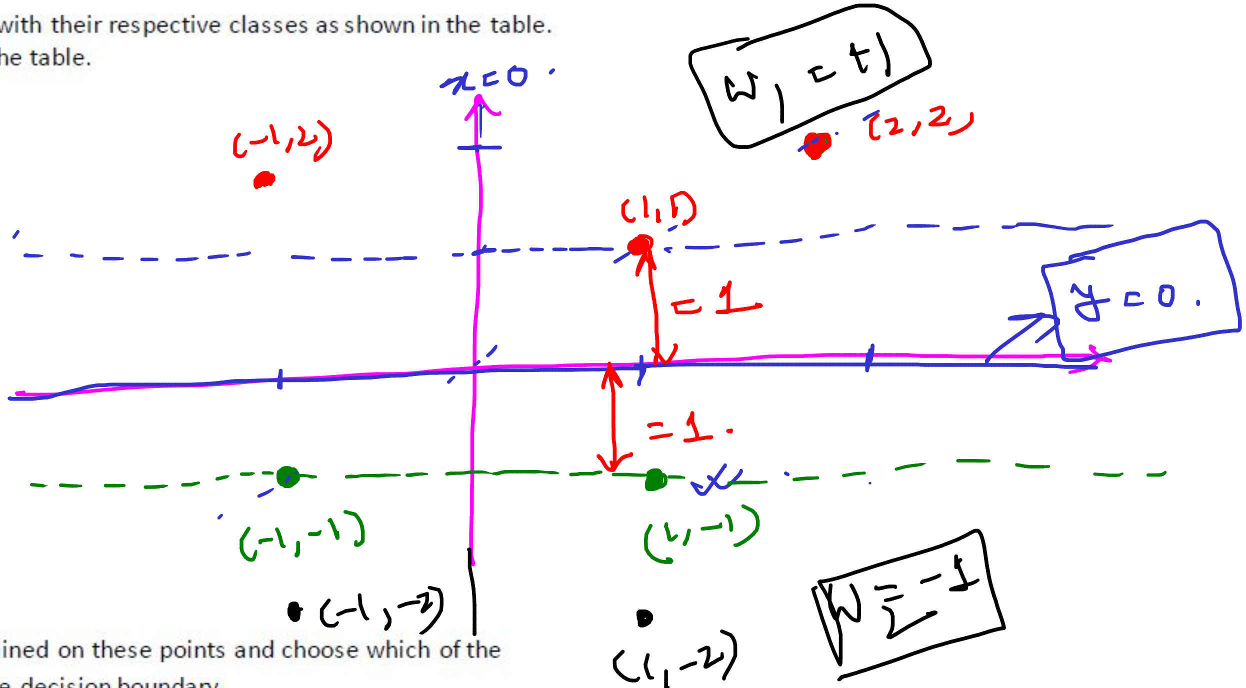
b. $x = 0$

c. $x = y$

d. $x + y = 1$

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.

X	Y	Class Label
1	1	+1
-1	-1	-1
2	2	+1
-1	2	+1
1	-1	-1



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.

i) The point (-1,-2) is classified as -1

ii) The point (-1,-2) is classified as +1

iii) The point (1,-2) is classified as -1

iv) The point (1,-2) is classified as +1

a. Only statement ii is true
b. Both statements i and iii are true
c. Both statements i and iv are true
d. Both statements ii and iii are true

Suppose we have one feature $x \in R$ and binary class y. The dataset consists of 3 points: p1: (x1, y1) = (−1, −1), p2: (x2, y2) = (1, 1), p3: (x3, y3) = (3, 1). Which of the following true with respect to SVM?

1

$x \in \mathbb{R}$

$y \in [-1, 1]$

a. Maximum margin will increase if we remove the point p2 from the training set. $M=4$

b. Maximum margin will increase if we remove the point p3 from the training set. $M=2$

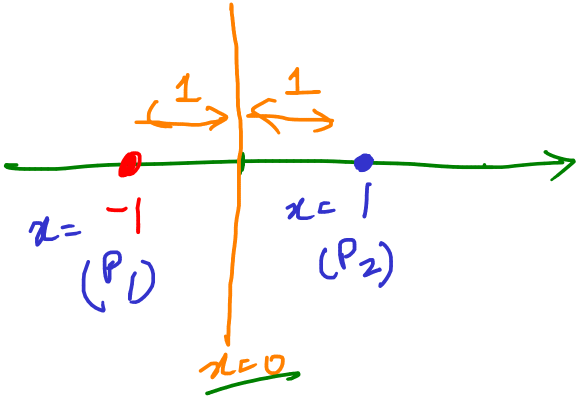
c. Maximum margin will remain same if we remove the point p2 from the training set.

d. None of the above.

Feature

Class.

Point	x	y	
P_1	−1	−1	→ Red
P_2	1	1	
P_3	3	1	→ Blue

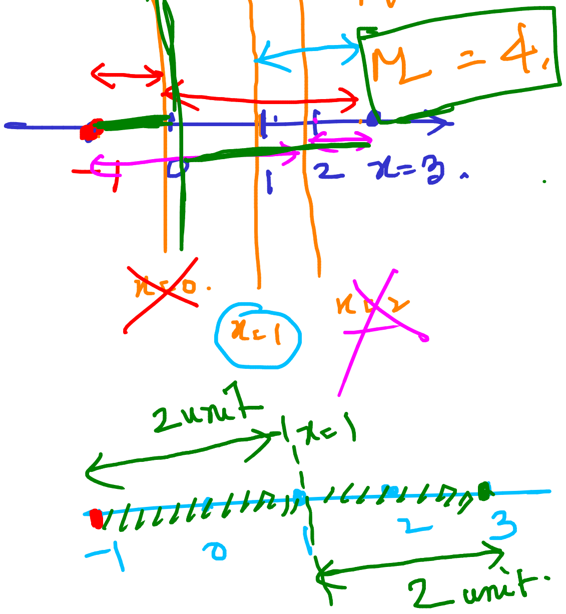


$M = 1 + 1 = 2.$

$\boxed{M = 2}$

A

$M = 2 + 2$



Choose the correct option regarding classification using SVM for two classes

Statement i : While designing an SVM for two classes, the equation $y_i(a^t x_i + b) \geq 0$ is used to choose the separating plane using the training vectors.

Statement ii : During inference, for an unknown vector x_j , if $y_j(a^t x_j + b) \cancel{\geq} 0$, then the vector can be assigned class 1.

Statement iii : During inference, for an unknown vector x_j , if $(a^t x_j + b) > 0$, then the vector can be assigned class 1.

Statement iv : While designing an SVM for two classes, the equation $y_i(a^t x_i + b) \geq 1$ is used to choose the separating plane using the training vectors.

a. Only Statement i is true
b. Both Statements ii and iii are true
c. Both Statements i and ii are true
d. Both Statements iii and iv are true

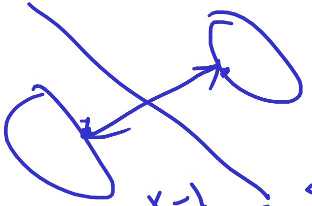
Traing.
Test.
Inference

Testing.

Training:-
$y_i(a^t x_i + b) \geq 1$

Testing:-
$a^t x + b > 0$ $\in \omega_1$
$a^t x + b < 0$ ω_2.

$y_i \in \{1, -1\}$
$y_i(a^t x + b) \geq 1$ $\in \omega_1$
$a^t y + b < 1 \in \omega_2$



$a^t x + b = -2$ $\times -1$ $= 2 \geq 1$
ω_2

$y_i(a^t x_i + b) \geq 1$

The shape of the loss landscape during optimization of SVM resembles to which structure?

a. Linear
b. Ellipsoidal
c. Non-convex with multiple possible local minimum
d. Paraboloid

Margin = $\frac{2}{\|w\|}$

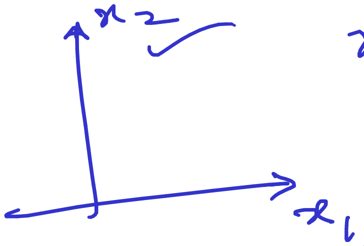
Constraint:- $y_i(w^T x_i + b) \geq 1$

Lagrangian:- $\mathcal{L}(\alpha)$

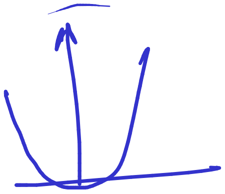
$\frac{\partial \mathcal{L}}{\partial b}$, $\frac{\partial \mathcal{L}}{\partial w}$.

$\mathcal{L}_{(\alpha)} = \sum\sum \alpha_i \alpha_j$

α^2



$x_1 + x_2 = 1.5$



Parabola

$w = [1 \quad 1]$

$b = 1.5$

784

$y = x^2$

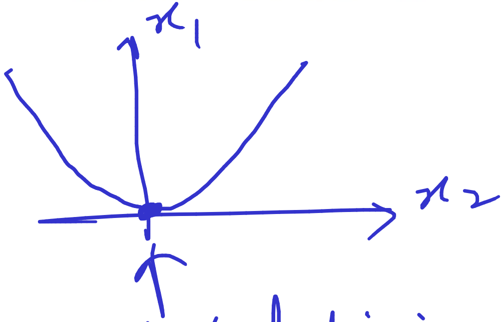
How many local minimum can be encountered while solving the optimization for maximizing margin for SVM?

a. 2

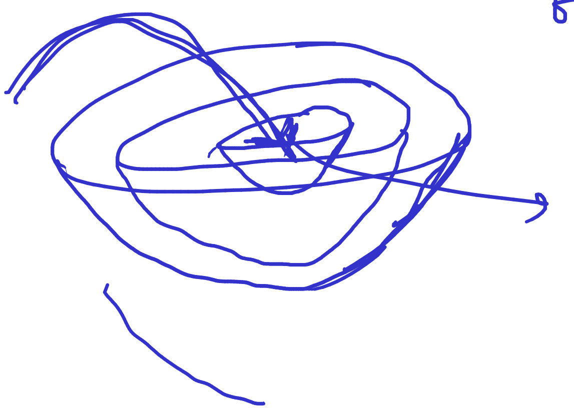
b. 1

c. ∞ (infinite)

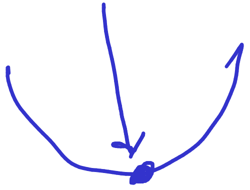
d. 0



Global Minima = local minima



minima

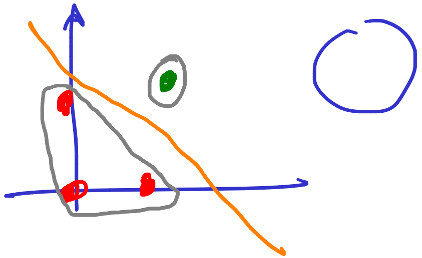


Which of the following cannot be realized with single layer perceptron (only input and output layer)?

a. AND
b. OR
c. NAND
d. XOR

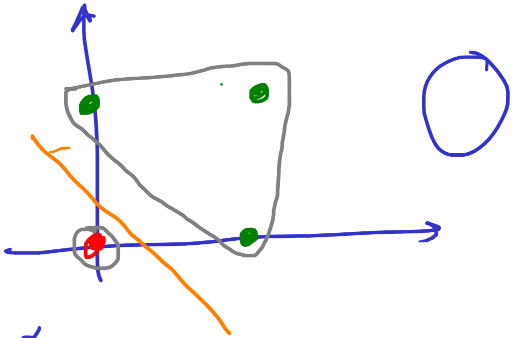
Linearly separable case → Single Neurone.

AND



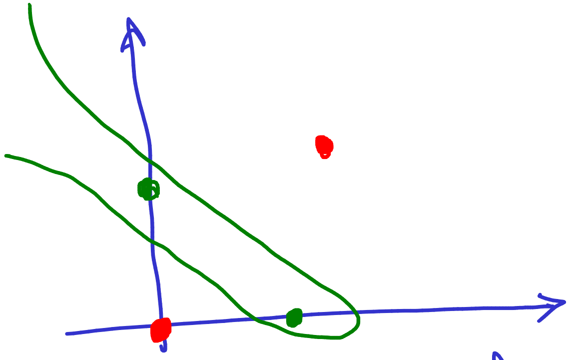
Nonlinear → More than one Neurone

OR



XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0



line → Nonlinear Decision Boundary.

For a function f (θ_0, θ_1), if θ_0 and θ_1 are initialized at a local minimum, then what should be the values of θ_0 and θ_1 after a single iteration of gradient descent:

a. θ_0 and θ_1 will update as per gradient descent rule
b. θ_0 and θ_1 will remain same
c. Depends on the values of θ_0 and θ_1
d. Depends on the learning rate

$f(\theta_1, \theta_0)$

learning rate

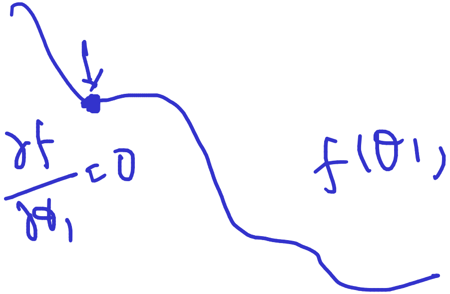
$$\theta(k+1) \Leftarrow \theta(k) - \eta \cdot \frac{\partial f}{\partial \theta}.$$

$$\theta_1(k+1) = \theta_1(k) - \eta \cancel{\frac{\partial f}{\partial \theta_1}}^{\,0}$$

$$\underbrace{\theta_1(k+1) = \theta_1(k)}_{\text{same}}.$$

$$\theta_0(k+1) = \theta_0(k) - \eta \cancel{\frac{\partial f}{\partial \theta_0}}^{\,0}.$$

$$\underbrace{\theta_0(k+1) = \theta(k)}_{\text{same}}$$



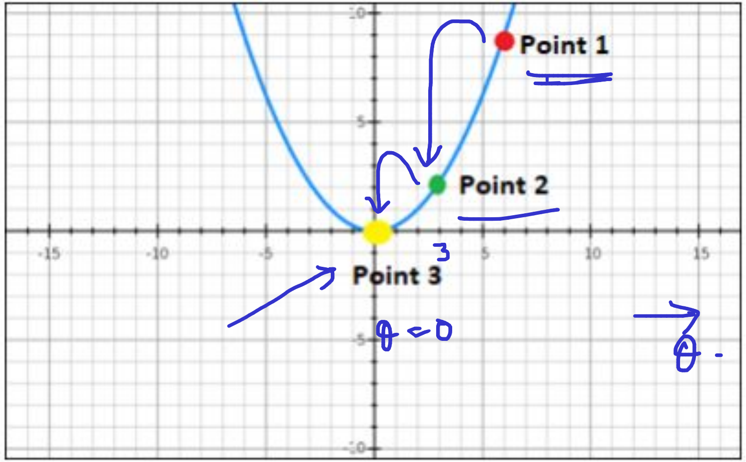
$\frac{\partial f}{\partial \theta_1} = 0$

$f(\theta_1, \theta_2)$

Suppose for a cost function $J(\theta) = 0.25\theta^2$ as shown in graph below, refer to this graph and choose the correct option regarding the Statements given below. θ is plotted along horizontal axis.

$$J(\theta) = 0.25\,\theta^2$$

$$\theta(k+1) = \theta(k) - \eta \frac{\partial J(\theta)}{\partial \theta}.$$



$$\frac{\partial J(\theta)}{\partial \theta} = 0.25 \times 2\theta.$$

$$\boxed{\frac{\partial J(\theta)}{\partial \theta} = 0.5\theta.}$$

$$\frac{\partial J}{\partial \theta} \propto \theta.$$

Weight update $\propto \theta$.

Statement i: The magnitude of weight update at the green point is higher than the magnitude of weight update at yellow point.

Statement ii: The magnitude of weight update at the green point is higher than the magnitude of weight update at red point.

a. Only Statement i is true
b. Only Statement ii is true
c. Both Statement i and ii are true
d. None of them are true

θ point 1 $>$ θ point 2 $>$ θ point 3.

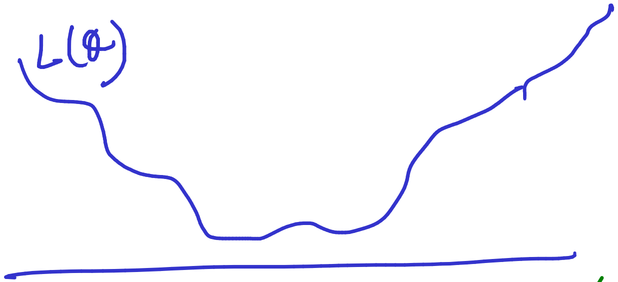
Weight update point $(1 > 2 > 3)$.

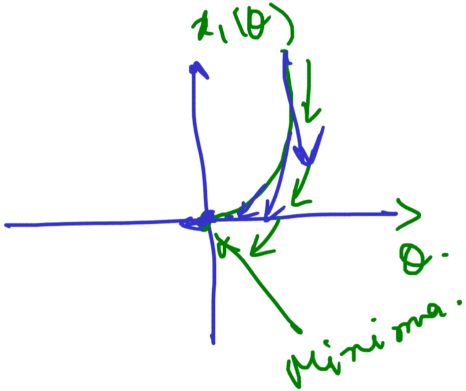
Choose the correct option. Gradient of a continuous and differentiable function is:

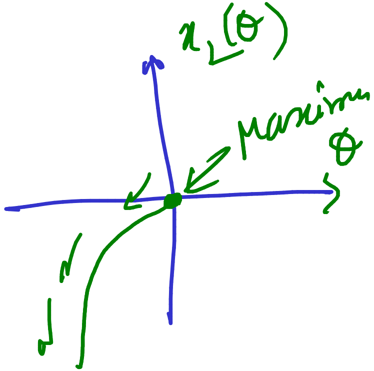
i) is zero at a minimum
ii) is non-zero at a maximum
iii) is zero at a saddle point
iv) magnitude decreases as you get closer to the minimum

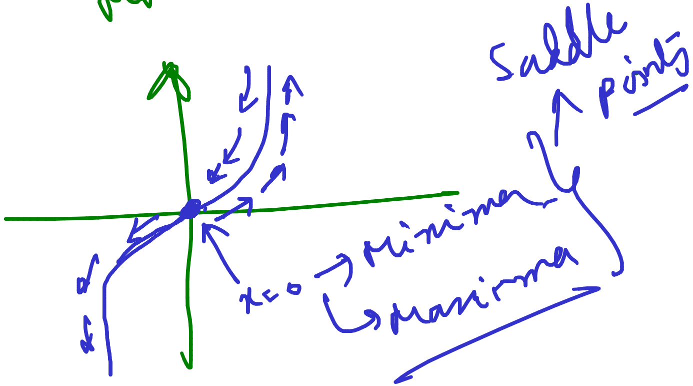
a. Only option (i) is corerct
b. Options (i), (iii) and (iv) are correct
c. Options (i) and (iv) are correct
d. Only option (ii) is correct

$$\frac{\partial L(\theta)}{\partial \theta} = 0$$ | minima | maxima









Input to SoftMax activation function is [3,1,2]. What will be the output?

a. [0.58,0.11, 0.31]
b. [0.43,0.24, 0.33]
c. [0.60,0.10,0.30]
d. [0.67, 0.09,0.24]

$$V = [3, 1, 2].$$

$$\text{Softmax}(v)$$

$$v = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}_{3\times 1} \xrightarrow{\sigma} \begin{bmatrix} o_1 \\ o_2 \\ o_3 \end{bmatrix}_{3\times 1} = O$$

$$\boxed{O_i = \frac{e^{v_i}}{\sum_{\forall i} e^{v_i}}}$$

$$O_1 = \frac{e^{v_1}}{\sum_{i=1}^{3} e^{v_i}} = \frac{e^{v_1}}{e^{v_1} + e^{v_2} + e^{v_3}} = \frac{e^3}{\underbrace{e^3 + e^2 + e^1}} = 0.665 \approx 0.67$$

$$O_2 = , O_3 =$$

Which of the following options is true?

a. In Stochastic Gradient Descent, a small batch of sample is selected randomly instead of the whole data set for each iteration. Too large update of weight values leading to faster convergence
b. In Stochastic Gradient Descent, the whole data set is processed together for update in each iteration.
c. Stochastic Gradient Descent considers only one sample for updates and has noisier updates.
d. Stochastic Gradient Descent is a non-iterative process

SG → for each sample.

FBG → updating for the whole dataset/full batch.

MBG → Mini batch weight update

What are the steps for using a gradient descent algorithm?

1. Calculate error between the actual value and the predicted value
2. Re-iterate until you find the best weights of network
3. Pass an input through the network and get values from output layer
4. Initialize random weight and bias
5. Go to each neurons which contributes to the error and change its respective values to reduce the error

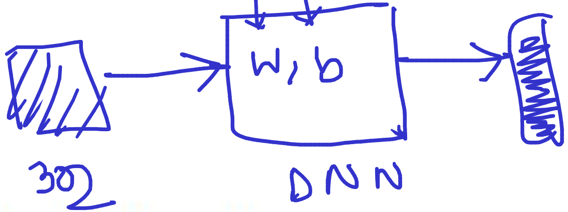
a. 1, 2, 3, 4, 5

b. 5, 4, 3, 2, 1

c. 3, 2, 1, 5, 4

d. 4, 3, 1, 5, 2

① Initialize.



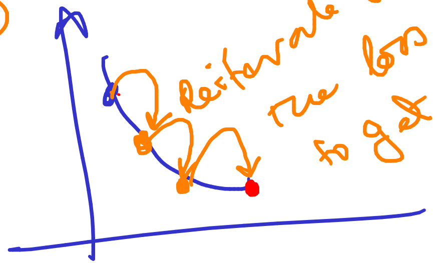
② Pass input and get o/p

③ Loss value calculate.

④ Gradient, Gradient descent rule, update the weights of the neurons → Reduce the loss.

⑤

4 → 3 → 1 → 5 → 2.



Reiterate over the loss curve to get the best result.

$J(\theta) = 2\theta^2 - 2\theta + 2$ is a given cost function? Find the correct weight update rule for gradient descent optimization at step t+1? Consider, α=0.01 to be the learning rate.

a. $\theta_{t+1} = \theta_t - 0.01(2\theta - 1)$
b. $\theta_{t+1} = \theta_t + 0.01(2\theta - 1)$
c. $\theta_{t+1} = \theta_t - (2\theta - 1)$
d. $\theta_{t+1} = \theta_t - 0.02(2\theta - 1)$

$J(\theta) = 2\theta^2 - 2\theta + 2.$

$\alpha = 0.01$

$(t+1)^{th}$ iteration →

$\theta(t+1) = \theta(t) - \alpha . \frac{\partial J(\theta)}{\partial \theta}$

$\frac{\partial J}{\partial \theta} = 4\theta - 2$
$= 2(2\theta - 1)$

$\theta(t+1) = \theta(t) - 0.01 \times 2 \times (2\theta - 1)$

$\boxed{\theta(t+1) = \theta(t) - 0.02(2\theta - 1)}$