University of Asia Pacific

Admit Card

Final-Term Examination of Spring, 2020

Financial Clearance	PAID

Registration No : 17101007

Student Name : Mahnaz Rafia Islam

Program : Bachelor of Science in Computer Science and Engineering

SI.NO.	COURSE CODE	COURSE TITLE	CR.HR.	EXAM. SCHEDULE
1	CSE 400	Project / Thesis	3.00	
2	CSE 401	Mathematics for computer Science	3.00	
3	CSE 403	Artificial Intelligence and Expert Systems	3.00	
4	CSE 404	Artificial Intelligence and Expert Systems Lab	1.50	
5	CSE 405	Operating Systems	3.00	
6	CSE 406	Operating Systems Lab	1.50	
7	CSE 407	ICTLaw, Policy and Ethics	2.00	
8	CSE 410	Software Development	1.50	
9	CSE 427	Topics of Current Interest	3.00	

Total Credit: 21.50

1. Examinees are not allowed to enter the examination hall after 30 minutes of commencement of examination for mid semester examinations and 60 minutes for semester final examinations.

2. No examinees shall be allowed to submit their answer scripts before 50% of the allocated time of examination has elapsed.

3. No examinees would be allowed to go to washroom within the first 60 minutes of final examinations.

4. No student will be allowed to carry any books, bags, extra paper or cellular phone or objectionable items/incriminating paper in the examination hall.
Violators will be subjects to disciplinary action.

This is a system generated Admit Card. No signature is required.

Admit Card Generation Time: 26-Oct-2020 05:55 PM

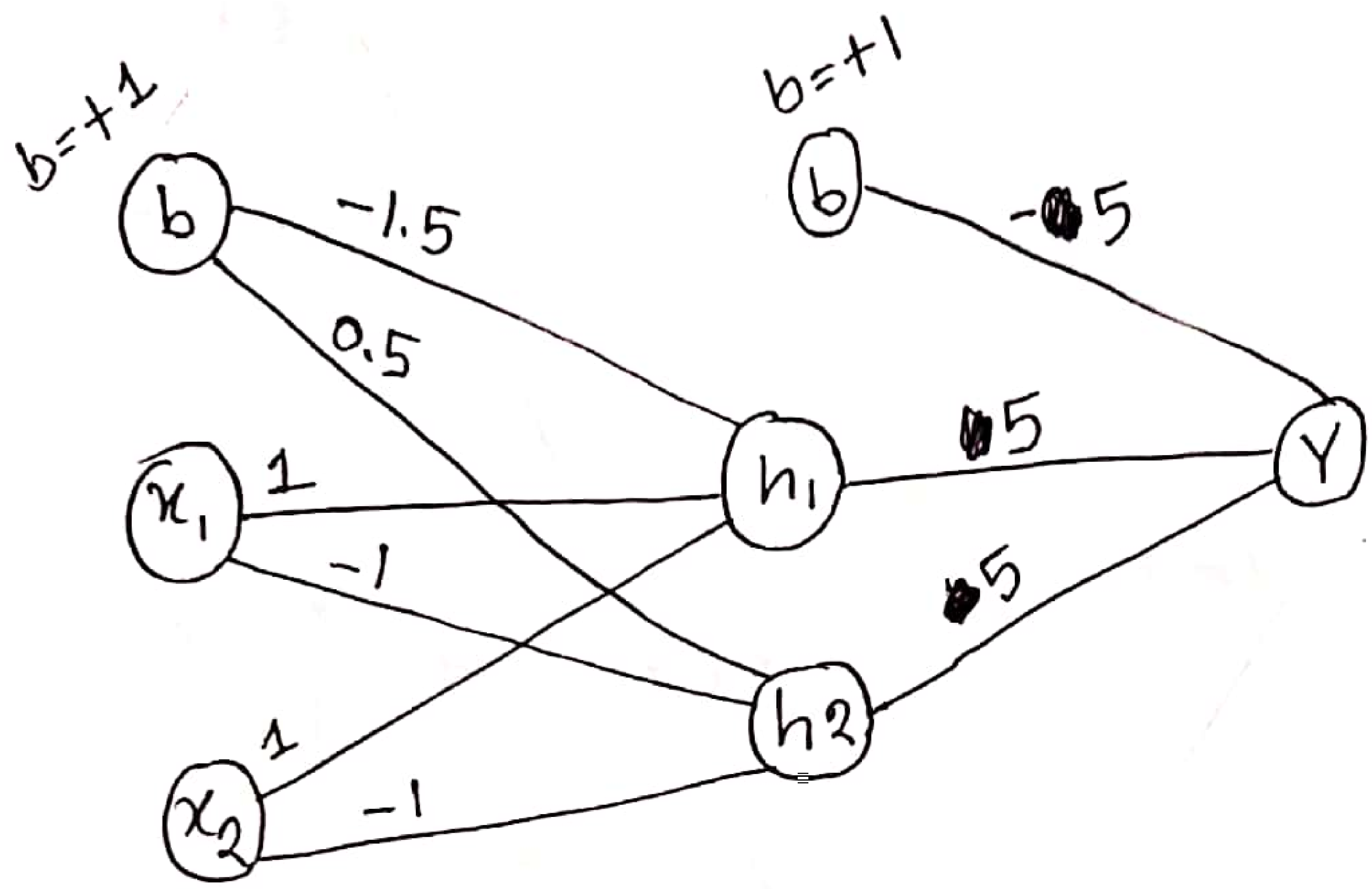
University of Asia Pacific
Department of Computer Science and Engineering

Final Term Examination: Spring-2020

Name: Mahnaz Rafia Isam	Registration No: 17101007
Roll No: 07 Year: 4th Semester: 1st	Course Code: CSE 427
Course Title: Machine Learning	Date: 07.11.2020

ID-17101007

Answer to the question no: 4(a) (OR)



For, $x_1=0, x_2=0$

$$h_1^{(1)} = \sum_{(x_1=0,\, x_2=0)} = b*(-1.5) + x_1(1) + x_2*(1)$$

$$= 1\times(-1.5) + 0\times 1 + 0\times 1$$

$$= -1.5 \qquad \boxed{h_1^{(1)} = 0}$$

$$h_2^{(1)} = \sum_{(x_1=0,\, x_2=0)} = b\times(0.5) + x_1*(-1) + x_2*(-1)$$

$$= 1\times(0.5) + 0*(-1) + 0*(-1)$$

$$= 0.5 \qquad \boxed{h_2^{(1)} = 1}$$

$$y = \sum_{(h_1^{(1)}=0,\, h_2^{(1)}=1)} = b*(-0.5) + h_1^{(1)}\times 1.5 + h_2^{(1)}\times 1.5$$

$$= 1*(-0.5) + 0\times 1.5 + 1\times 1.5$$

$= -5 + 0 + 5$

$= 0 \quad \boxed{y=1}$

$\boxed{x_1=0, x_2=1}$

$h_1^{(2)} = \Sigma_{(x_1=0, x_2=1)} = 1\times(-1.5) + 0\times1 + 1\times1$

$= -1.5+1 = -0.5 \quad \boxed{h_1^{(2)} = 0}$

$h_2^{(2)} = \Sigma_{(x_1=0, x_2=1)} = 1\times(0.5) + 0\times(-1) + 1\times(-1)$

$= 0.5 - 1$

$= -0.5 \quad \boxed{h_2^{(2)} = 0}$

$y \rightarrow \Sigma_{(h_1^{(2)}=0,\, h_2^{(2)}=0)} = 1*(-5) + 0\times5 + 0\times5$

$= -5 \quad \boxed{y=0}$

$\boxed{x_1=1, x_2=0}$

$h_1^{(3)} \rightarrow \Sigma_{(x_1=1, x_2=0)} = 1\times(-1.5) + 1\times1 + 0\times1$

$= -1.5+1 = -0.5 \quad \boxed{h_1^{(3)} = 0}$

$h_2^{(3)} \rightarrow \Sigma_{(x_1=1, x_2=0)} = 1\times(0.5) + 1\times(-1) + 0\times(-1)$

$= 0.5 - 1 = -0.5 \quad \boxed{h_2^{(3)} = 0}$

$$y \to \sum_{(h_1^{(3)}=0,\, h_2^{(3)}=0)} = 1\times(-5) + 0\times 5 + 0\times 5 = -5$$

$$\boxed{y=0}$$

For $\boxed{x_1=1,\ x_2=1}$

$$h_1^{(4)} \to \sum_{(x_1=1,\, x_2=1)} = 1\times(-1.5) + 1\times 1 + 1\times 1$$
$$= -1.5+1+1 = 0.5 \quad \boxed{h_1^{(4)}=1}$$

$$h_2^{(4)} \to \sum_{(x_1=1,\, x_2=1)} = 1\times(0.5) + 1\times(-1) + 1\times(-1)$$
$$= 0.5-1-1 = -1.5 \quad \boxed{h_2^{(4)}=0}$$

$$y \to \sum_{(h_1^{(4)}=1,\, h_2^{(4)}=0)} = 1\times(-5) + 1\times(5) + 0\times 5$$
$$= -5+5 = 0 \quad \boxed{y=1}$$

So, the name of the logic gate is XNOR gate.

x_1	x_2	y(XNOR)
0	0	1
0	1	0
1	0	0
1	1	1

Answer to the question no: 4(b) or

Yes I agree with the given statement because logistic regression is a one layer neural network. It is used as activation function, in the hidden layer of a neural network. It follows the single layer perception which is the simplest Neural Network with only one neuron. Logistic Regression is also a binary classification. Output can be 0 and 1, which follow single neuron perception

Answer to the question no: 4 (c)

Last digit of my id is 7, one greater is 8.

$f(x,y,z) = x+y+2z$

$x^2+y^2+z^2=8 \quad \therefore g(x,y,z) = x^2+y^2+z^2-8$

We know, Lagrance function,

$F(x,y,z,\lambda) = f(x,y,z) - \lambda * g(x,y,z)$

$F_x = 0$

$F_y = 0$

$F_z = 0$

$F_\lambda = 0$

$F_x = \frac{\partial}{\partial x} F(x,y,z,\lambda)$

$= \frac{\partial}{\partial x} \{ f(x,y,z) - \lambda * g(x,y,z) \}$

$= \frac{\partial}{\partial x} \{ x+y+2z - \lambda(x^2+y^2+z^2-8) \}$

$= \frac{\partial}{\partial x} \{ x+y+2z - \lambda x^2 - \lambda y^2 - \lambda z^2 + 8\lambda \}$

$= 1+0+0-2\lambda x - 0 - 0 + 0$

$= 1-2\lambda x$

$F_x = 0$

$1 - 2\lambda x = 0$

$x = \frac{1}{2\lambda}$

$F_y = \frac{\partial}{\partial y}\{F(x, y, z, \lambda)\}$

$= \frac{\partial}{\partial y}\{x + y + 2z - \lambda x^2 - \lambda y^2 - \lambda z^2 + 8\lambda\}$

$= 0 + 1 - 0 - 0 - 2\lambda y - 0 + 0$

$= 1 - 2\lambda y$

$F_y = 0$

$1 - 2\lambda y = 0$

$y = \frac{1}{2\lambda}$

$F_z = \frac{\partial}{\partial z}\{F(x, y, z, \lambda)\}$

$= \frac{\partial}{\partial z}\{x + y + 2z - \lambda x^2 - \lambda y^2 - \lambda z^2 + 8\lambda\}$

$= 0 + 0 + 2 - 0 - 0 - 2z\lambda + 0$

$= 2 - 2\lambda z$

$F_z = 0$

$2 - 2\lambda z = 0$

$2 = 2\lambda z$

$z = \frac{1}{\lambda}$.

$F_\lambda = \frac{\partial}{\partial \lambda} \{x + y + 2z - \lambda x^2 - \lambda y^2 - \lambda z^2 + 8\lambda\}$

$= 0 + 0 + 0 - x^2 - y^2 - z^2 + 8$

$F_\lambda = 0$

$8 - x^2 - y^2 - z^2 = 0$

$8 = x^2 + y^2 + z^2$

$8 = \left(\frac{1}{2\lambda}\right)^2 + \left(\frac{1}{2\lambda}\right)^2 + \left(\frac{1}{\lambda}\right)^2$

$8 = \frac{1}{4\lambda^2} + \frac{1}{4\lambda^2} + \frac{1}{\lambda^2}$

$8 = \frac{1+1+4}{4\lambda^2}$

$8 = \frac{6}{4\lambda^2}$

$\lambda^2 = \frac{6}{4 \times 8} = \frac{6}{32}$

$\lambda = \pm 0.43$

When,

$\lambda = -0.43$

$x = \frac{1}{2\lambda} = \frac{1}{2\times(-0.43)} = -1.16$

$y = \frac{1}{2\lambda} = -1.16$

$z = \frac{1}{\lambda} = \frac{1}{-0.43} = -2.33$

$\lambda = 0.43$

$x = \frac{1}{2\lambda} = \frac{1}{2\times 0.43} = 1.16$

$y = \frac{1}{2\lambda} = 1.16$

$z = \frac{1}{\lambda} = \frac{1}{0.43} = 2.33$

$(x,y,z) = (-1.16, -1.16, -2.33)$

$(x,y,z) = (1.16, 1.16, 2.33)$

For, $(x,y,z) = (-1.16, -1.16, -2.33)$

$f(x,y,z) = x+y+2z$

$= -1.16 - 1.16 + 2(-2.33) = -6.98$

For $(x,y,z) = (1.16, 1.16, 2.33)$

$f(x,y,z) = x+y+2z$

$= 1.16 + 1.16 + 2(2.33)$

$= 6.98 \longrightarrow$ This is the maximum value.

Answer to the question no: 1(a)

Total data points = 20

A has - 7

B has - 8

C has - 20 - (7+8)

= 5

my id - 17101007

$\therefore X = 7+1 = 8$

Here,

$P_A = \frac{7}{20}$, $P_B = \frac{8}{20}$, $P_C = \frac{5}{20}$

So, entropy $H(S) = -P_A \log_2 P_A - P_B \log_2 P_B - P_C \log_2 P_C$

$= -\frac{7}{20}\log_2\left(\frac{7}{20}\right) - \frac{8}{20}\log_2\left(\frac{8}{20}\right) - \frac{5}{20}\log_2\left(\frac{5}{20}\right)$

$= 1.559$

(Ans)

Answer to the question no: 1(b)

No, I don't agree with the given statement.

Kmeans clusterin is a unsupervised learning because the dataset has no label/class defined. K-means bundles the data points around centroids which are the mean distance from each of the data points clustered together. This algorithm partitions data into K distinct clusters based on distance to the centroid of a cluster.

K-NN is a supervised learning algorithm, which is used for classification. It takes some labeled data points and uses them to learn how to label other data points or a new data point taking the maximum value of the class/label. It classify a data point based on the known classification of other points.

Answer to the question no: 1(c)

Here, my id is 17101007

$\therefore \theta_0 = 0$

$\theta_1 = 7$

Given, learning rate $\alpha = 0.1$

(i) Dataset:

x	y
2	17
4	28
6	42

(ii) Hypothesis function:

$$h_\theta(x) = \theta_0 + \theta_1 * x$$

(iii) Parameter initialization:

$$\theta_0 = 0, \ \theta_1 = 7$$

$$h_\theta^{(1)}(x) = \theta_0 + \theta_1 * x^{(1)}$$

$$= 0 + 7 * 2 = 14$$

$$h_\theta^{(2)}(x) = \theta_0 + \theta_1 * x^{(2)}$$

$$= 0 + 7 * 4 = 28$$

$$h_\theta^{(3)}(x) = \theta_0 + \theta_1 * x^{(3)}$$
$$= 0 + 7 * 6 = 42$$

(iv) cost function:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta^{(i)}(x) - y^{(i)} \right)^2$$

$$= \frac{1}{2 \times 3} \left\{ (14-17)^2 + (28-28)^2 + (42-42)^2 \right\}$$

$$= \frac{1}{6} (9+0+0)$$

$= 1.5$ error/cost

(v) Gradient Descent:

repeat untill convergence

$$\left\{ \theta_j := \theta_j - \frac{\alpha}{m} \frac{d}{d\theta_j} \{J(\theta)\} \right.$$

$\}$

$$\theta_0 := \theta_0 - \frac{\alpha}{m} \sum_{i=1}^{m} \left(h_\theta^{(i)} - y^{(i)} \right)$$

$$= 0 - \frac{0.1}{3} \left\{ (14-17) + (28-28) + (42-42) \right\}$$

$$= 0 - \frac{0.1}{3} (-3) = 0.1$$

$$\theta_1 := \theta_1 - \frac{\alpha}{m}\sum_{i=1}^{m}\left(h_\theta(i) - y^{(i)}\right) \cdot x^{(i)}$$

$$= 7 - \frac{0.1}{3}\left\{(14-17) \cdot 2 + (28-28) \cdot 4 + (42-42) \cdot 6\right\}$$

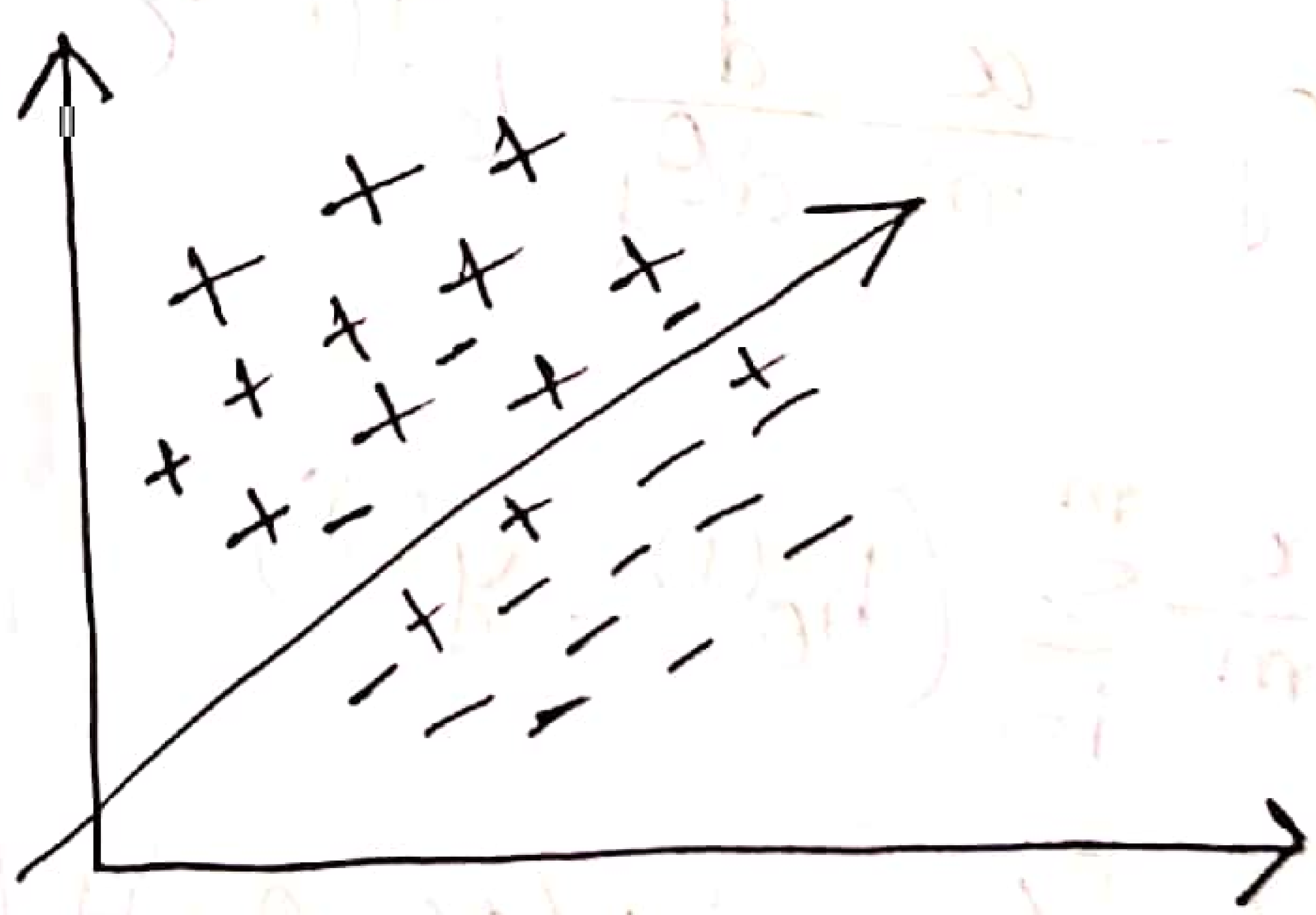
$$= 7 - \frac{0.1}{3}(-6+0+0)$$

$$= 7.2$$

$\therefore \theta_0 = 0.1, \ \theta_1 = 7.2.$

Answer to the question no:2(a)

Underfitting/High Bias:



Here, (+) and (-) define diffent label.

Over fitting / High variance:

To prevent High variance / Over fitting –

- Getting more training example.
  → fixes high variance.
- Trying smaller set of features.
  – Suppose I have 50 features. I will take 10 features. It fixes high variance.

To prevent High Bias / Under fitting:

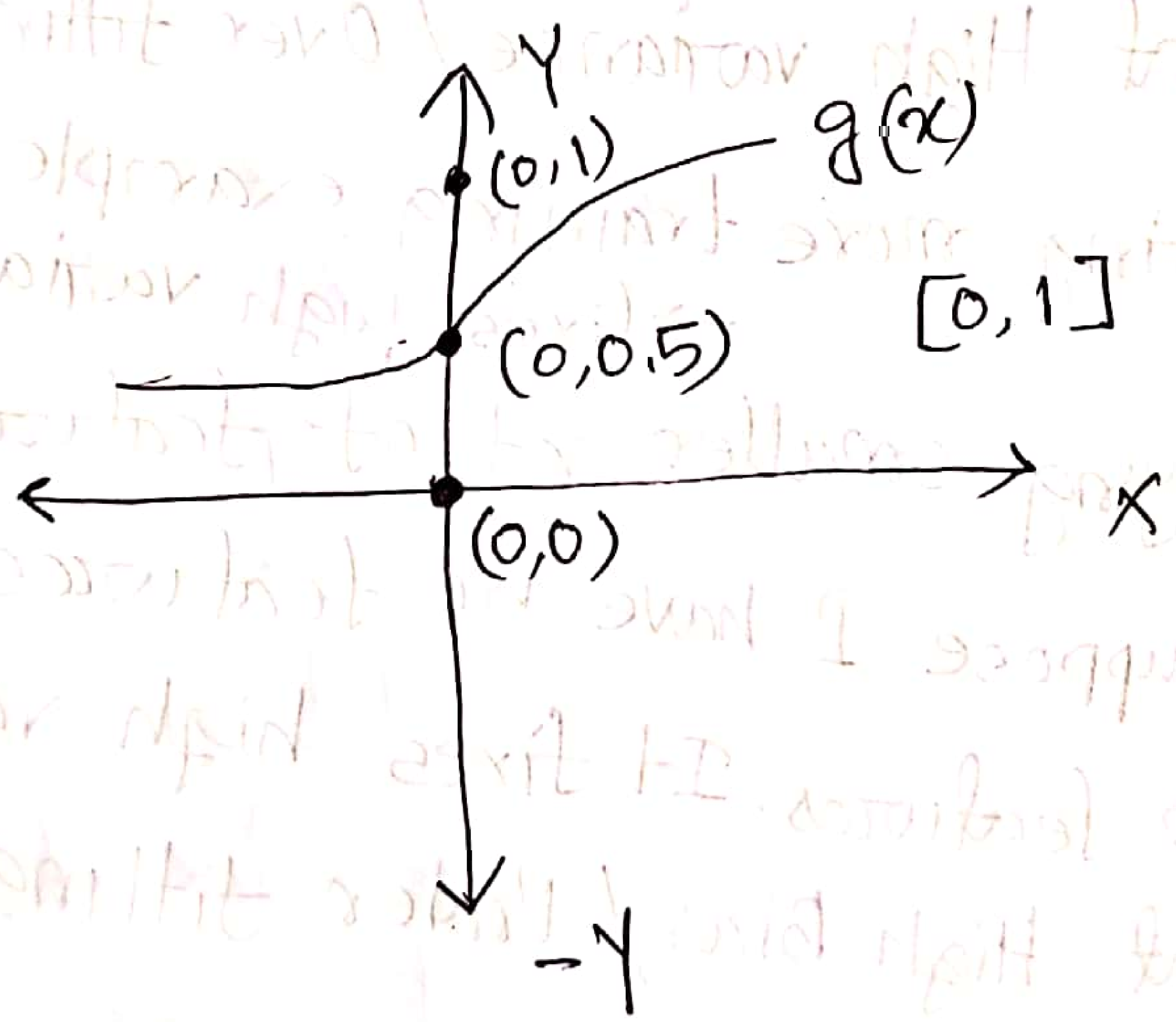
- Adding feature.
  – Suppose I have 10 features. I will take 20 features.
- Adding polynomial feature.

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 * x_1 + \theta_2 \cdot x_2 \quad X$$

$$= \theta_0 + \theta_1 \cdot x_1^2 + \theta_2^2 \cdot x_2^3$$

Answer to the question no: 2(b)

① Sigmoid function: $g(x) = \dfrac{1}{1+e^{-x}}$



$$g(0) = \frac{1}{1+e^{-0}} = \frac{1}{2} = 0.5$$

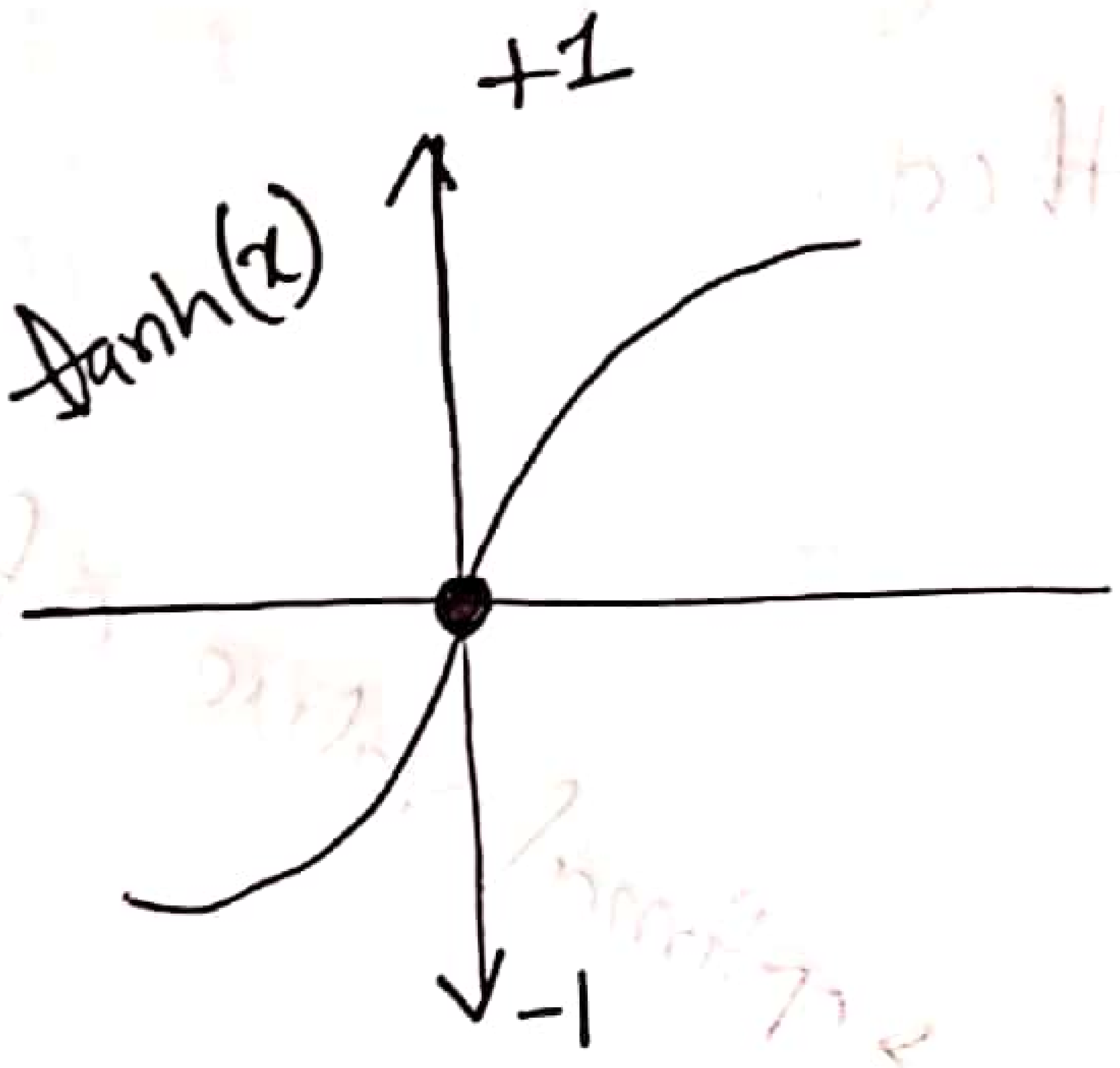
$$g(+\infty) = \frac{1}{1+e^{-\infty}} = \frac{1}{1+\frac{1}{e^{\infty}}} = \frac{1}{1+0} = 1$$

$$g(-\infty) = \frac{1}{1+e^{-(-\infty)}} = \frac{1}{1+e^{\infty}} = \frac{1}{\infty} = 0$$

$x \rightarrow +\infty \quad g(x) = 1$

$x \rightarrow -\infty \quad g(x) = 0$

(ii) Hyperbolic tangent function:



$[-1, 1]$

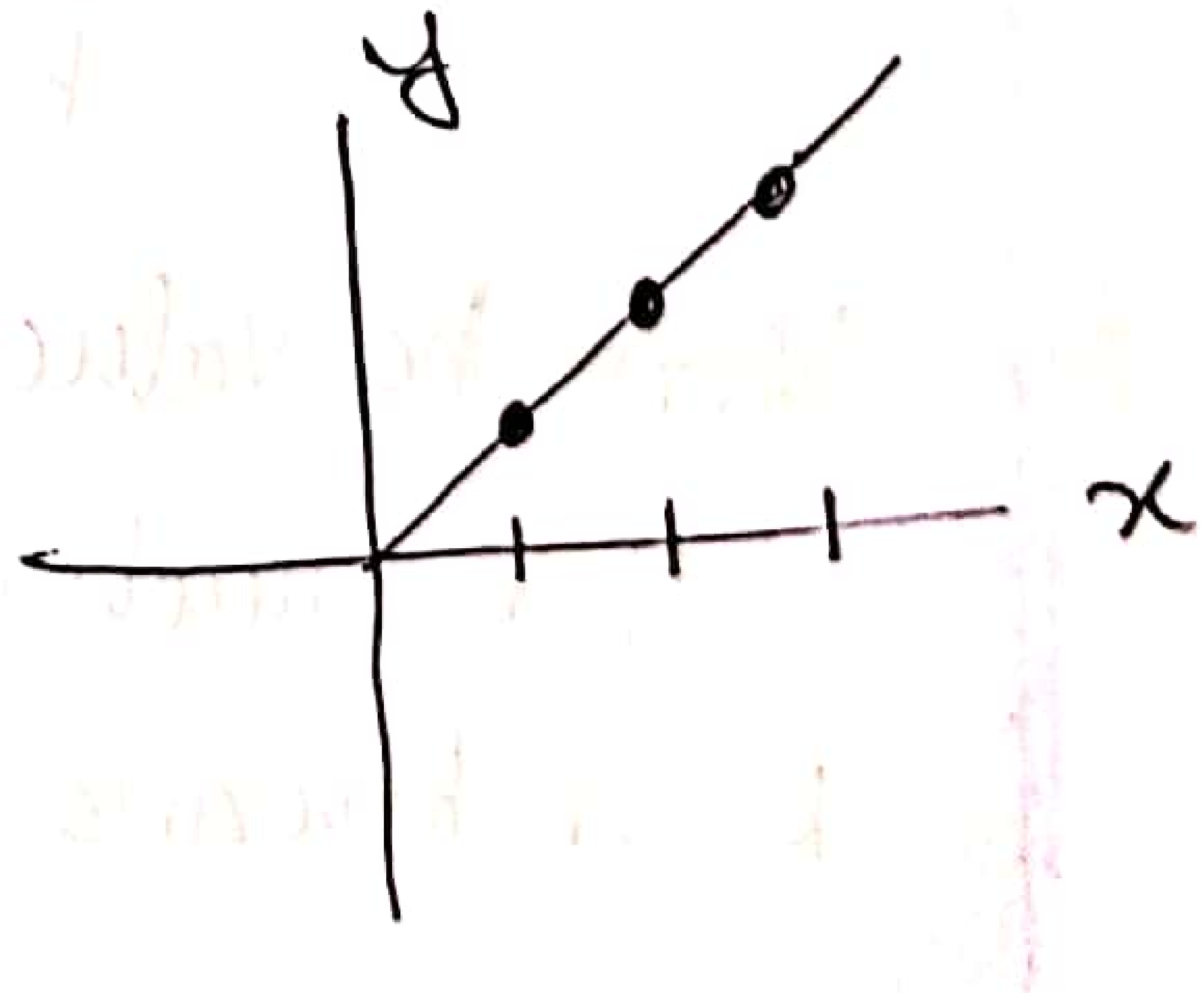
$$h_\theta^{(i)}(x) = g(\theta_0 + \theta_1 \cdot x_1^{(i)} + \theta \cdot x_2^{(i)})$$

(iii) Relu (Rectified Linear Unit)

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

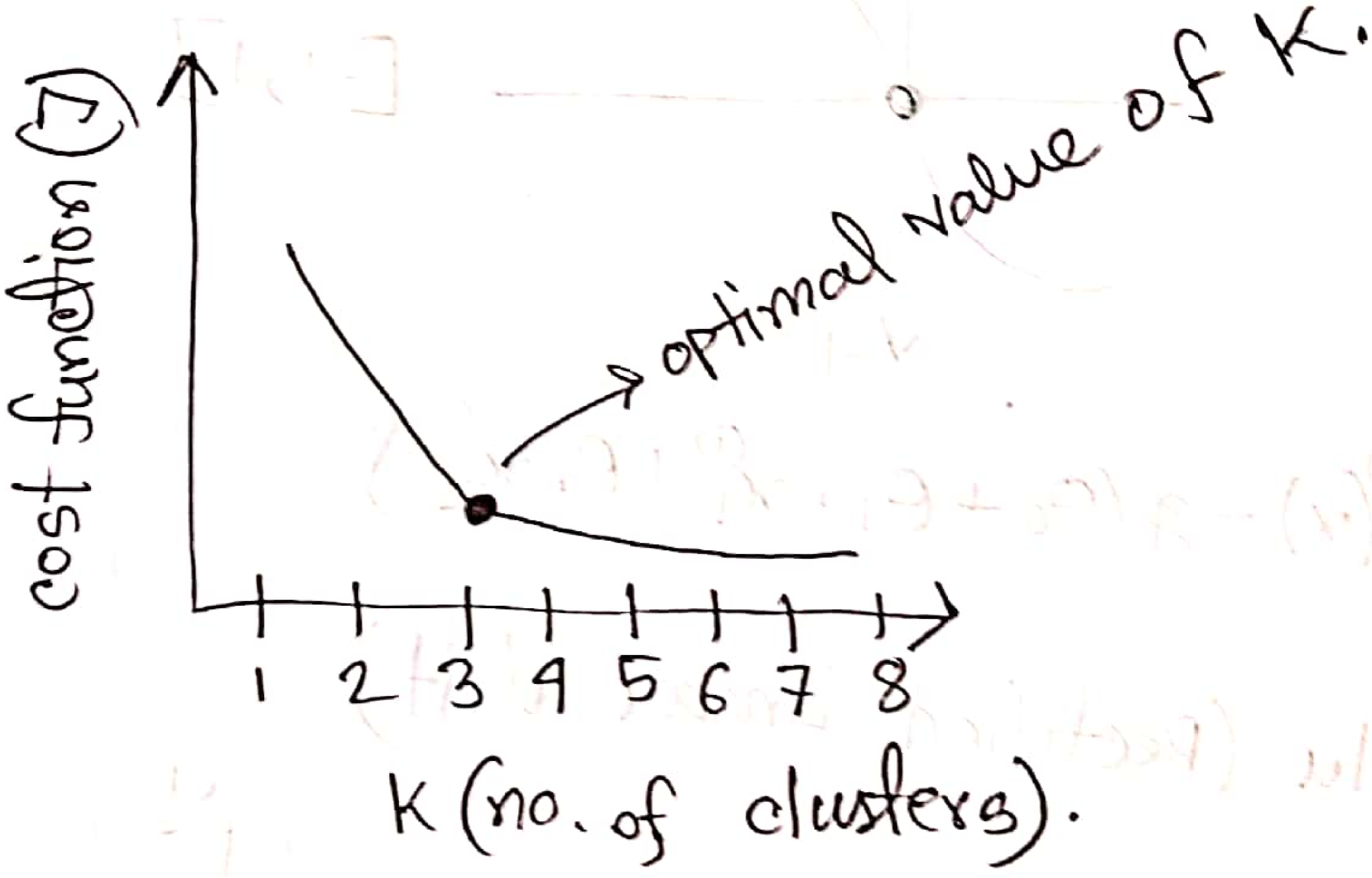
if $x > 0$ $f(x) = x$

if $x \leq 0$ $f(x) = 0$



Answer to the question no: 2(a)

We can determine the value of k in k-means clustering algorithm by using elbow plot/method.



When the value of cost functions does not differ much, that's the optimal value of k in kmeans.

Answer to the question no: 3(a)

$n = (17101007 \% 2) + 2 = 1 + 2 = 3$

stride is 2

pooling window is $3 * 3$.

Given 2D array is 7×7 pixels. So window size $= \frac{49}{4} = 12.25$

12 cannot take all points. so we will take 4×4 matrix.

~~0.77~~ ~~1.00~~	~~0.33~~	~~0.55~~
~~0.33~~		

1.00	0.33	0.55	0.33
0.33	1.00	0.33	0.55
0.55	0.33	1.0	0.11
0.33	0.55	0.11	0.77

This is the actual pooling window.

If we take 3×3 window size, the pooling result will be —

0.77	0.33	0.55
0.33	0.33	1.00
0.33	0.55	0.55

This is the pooling result.

Answer to the question no: 3(b)

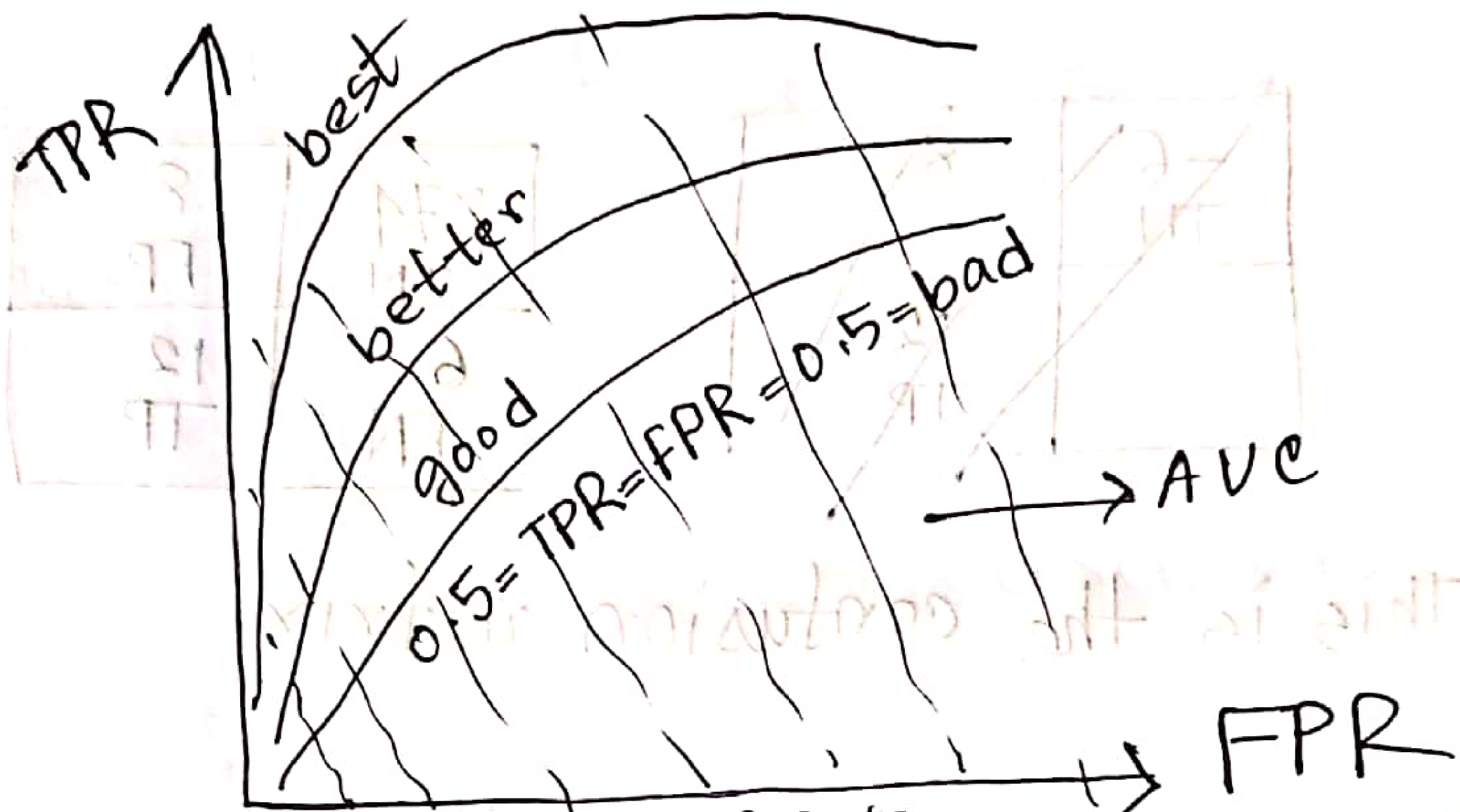


Fig: ROC curve.

The bigger the Area Under the curve, the model is much better.

Answer to the question no:3(c)

~~6 TN~~	~~8~~
	~~12 TP~~

974 TN	8 FP
6 FN	12 TP

This is the confusion matrix.

$$\text{Precission} = \frac{TP}{TP+FP} = \frac{12}{12+8} = 0.6$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{12}{12+6} = 0.67$$

$$\text{f1 scrore} = \frac{2 \times P \times R}{P+R}$$

$$= \frac{2 \times 0.6 \times 0.67}{0.6+0.67}$$

$$= 0.633$$

So, the model is very bad/poor as the value of f1 scrore is less than 0.7.