School of Information Sciences (SOIS), MAHE, Manipal 576104

Second Semester Master of Engineering - ME (Big Data and Data Analytics)

Degree Examination April - 2018

BDA 610 Machine Learning for Big Data

Scheme for Evaluation

1 **Explain about the different types of learning in Artificial Neural Networks** 10

Types of Learning:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning

Teach the neural network to perform certain task (learning). Providing sequence of sample inputs – compare output with expected output. Training continues until network able to provide the expected output 3 ½ Marks

Input → adjust the weight using learning algorithm → target output

Example:

Consider a basket with different kinds of fruits. Names of the fruits in basket: apple, banana, grape and cherry. Task is to arrange them as groups. You already learned about the physical characters of fruits through training. So arranging the same type of fruits at one place is easy now.

NO.	SIZE	COLOR	SHAPE	FRUIT NAME
1	Big	Red	Rounded shape with a depression at the top	Apple
2	Small	Red	Heart-shaped to nearly globular	Cherry
3	Big	Green	Long curving cylinder	Banana
4	Small	Green	Round to oval, Bunch shape Cylindrical	Grape

You can observe in the table that a column was labeled as "FRUIT NAME" this is called as response variable. If you learn from training data and then applying that knowledge to the test data, this type of learning is called as Supervised Learning. Classification come under Supervised learning.

Unsupervised Learning

For a training input, output is not known. Modify weights to get more similar input assigned to output. Task: Arrange the fruits as groups. You don't know anything about that fruits (first time you have seen them) – not learned. 3 ½ Marks

You will take a fruit and you will arrange them by considering physical character of that particular fruit.

Suppose you have considered color. Then you will arrange them on considering base condition as color. Then the groups will be something like this. RED COLOR GROUP: apples & cherry fruits. GREEN COLOR GROUP: bananas & grapes.

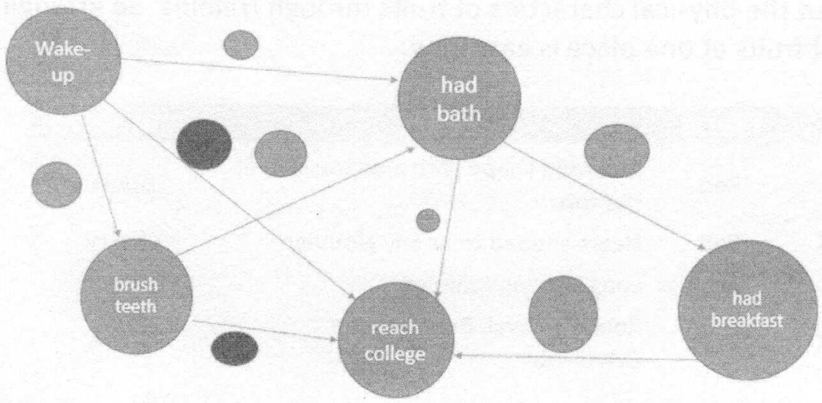
Next you will take another physical character such as **size**. RED COLOR AND BIG SIZE: apple. RED COLOR AND SMALL SIZE: cherry fruits. GREEN COLOR AND BIG SIZE: bananas. GREEN COLOR AND SMALL SIZE: grapes.

Here there is no training and learning. This type of learning is known unsupervised learning. Clustering comes under unsupervised learning.

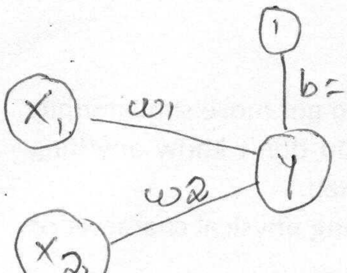
NO.	SIZE	COLOR	SHAPE	FRUIT NAME
1	Big	Red	Rounded shape with a depression at the top	Apple
2	Small	Red	Heart-shaped to nearly globular	Cherry
3	Big	Green	Long curving cylinder	Banana
4	Small	Green	Round to oval, Bunch shape Cylindrical	Grape

Reinforcement Learning
Rewards (+ve or –ve).
Example: Wake-up, brushed teeth, had bath, had breakfast, reach college

3 Marks



2 Develop a perceptron model for AND function with bipolar inputs and targets. The initial values of weights and bias are w1=w2=b=0, learning rate α = 1 and threshold θ=0. 10



x_1	x_2	b	y
1	1	1	1
1	-1	1	-1
-1	1	1	-1
-1	-1	1	-1

(1 Mark)

1. initialise weights and bias

$w_1 = w_2 = b = 0$ and $\alpha = 1$ $\theta = 0$

2. Begin (repeat until stopping condition True)

3. input (1, 1) : Target $t = 1$

set $x_i = (1, 1)$

4. Calculate $y_{in} = 0 + 1 \times 0 + 1 \times 0 = 0$

Apply activation $y = f(y_{in}) = 0$

5. $t \neq y$ $\therefore$ update weights

$w_1 (new) = w_1 (old) + \alpha t x_1$

$= 0 + 1 \times 1 \times 1 = 1$

$w_2 (new) = w_2 (old) + \alpha t x_2$

$= 0 + 1 \times 1 \times 1 = 1$

3 Marks

$b(new) = b(old) + \alpha t$

$= 0 + 1 \times 1$

$= 1$

$y_{in} = 1 + 1 \times 1 + 1 \times 1 = 3$ $y = f(y_{in}) = 1$

$t = y$ $\therefore$ go to step (2)

New input (1, -1) : t = -1

$b = 1$; $w_1 = 1$ $w_2 = 1$ $\alpha = 1$ $\theta = 0$

$y_{in} = 1 + 1 \times 1 + 1 \times -1 = 1$

$y = f(y_{in}) = 1$

$t = -1$. $y = 1$ $t \neq y$

∴ update weights

w_1 (new) = 0

w_2 (new) = 2

b(new) = 0

$y_{in} = 0 + 0 \times 1 + 2 \times -1 = -2$

$y = f(y_{in}) = -1$

$t = y$

∴ Go to next step - input

New input (-1, 1) : t = -1

$w_1 = 0$ $w_2 = 2$ $b = 0$ (2 Marks)

$y_{in}(-1, 1) = 2$

$t = -1$; $y_{in} = 2$; $y = f(y_{in}) = 1$

$t \neq y$ ∴ update weights

$w_1(new) = 1$; $w_2(new) = 1$; $b(new) = -1$

$y_{in} = -1$; $y = f(y_{in}) = -1$

$t = y$ ∴ go to next input.

new input $(-1, -1)$; $t = -1$

$y_{in} = -1$

$y = f(y_{in}) = -1$

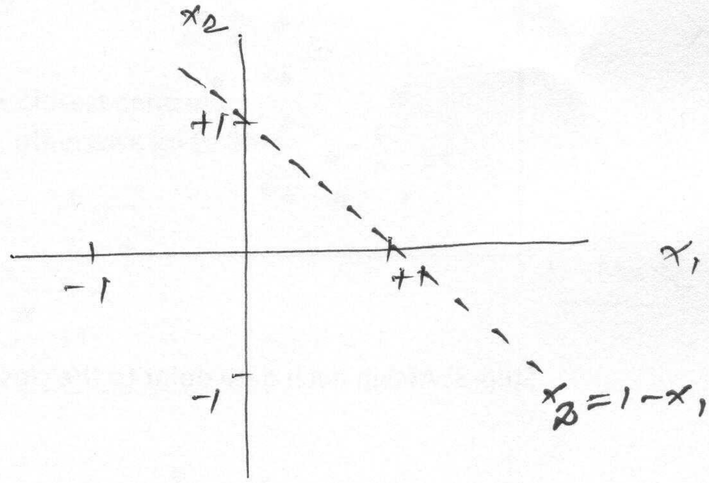
$t = y \Rightarrow$ stop the iteration (2 Marks)

<u>Find weights</u>

$w_1 = w_2 = 1 \quad b = -1$

$b + w_1 x_1 + w_2 x_2 = 0 \Rightarrow \boxed{x_2 = 1 - x_1}$

This equation separate the points



AND perceptron table

x_1	x_2	y
1	1	1
-1	1	-1
1	-1	-1
-1	-1	-1

(2 Marks)

3 Write the steps in k-means clustering algorithm. Explain with suitable example about k-means clustering 10

K-Means Algorithm

Step-1: Choose the number K of clusters

Step-2: Select at random k points, the centroids (not necessarily from your dataset) 3 Marks

Step-3: Assign each data point to the closest centroid → that forms k clusters

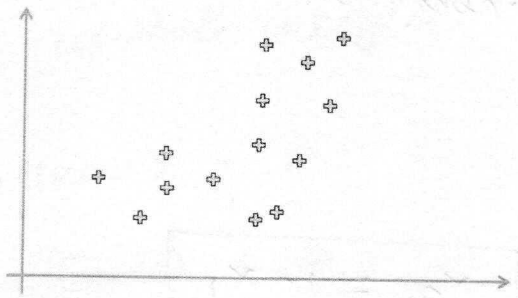
Step-4: Compute and place the new centroid of each cluster

Step-5: Reassign each data points to the new closest centroid.

If any reassignment took place, go to Step-4, otherwise go to Stop.

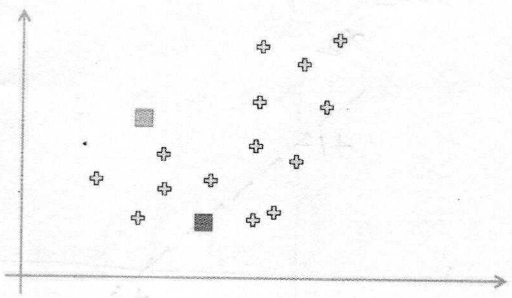
Stop: your Model is ready

STEP 1: Choose the number K of clusters: K = 2

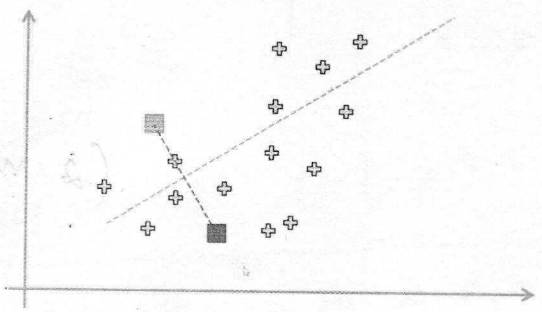


1 Mark

Step-2: Select at random k points, the centroids (not necessarily from your dataset)

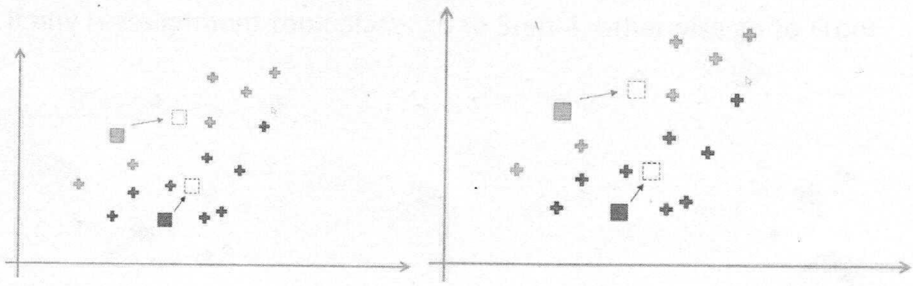


Step-3: Assign each data point to the closest centroid → that forms k clusters

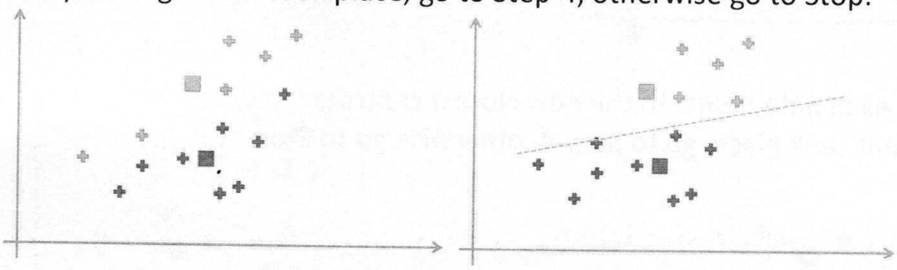


1 Mark

Step-4: Compute and place the new centroid of each cluster

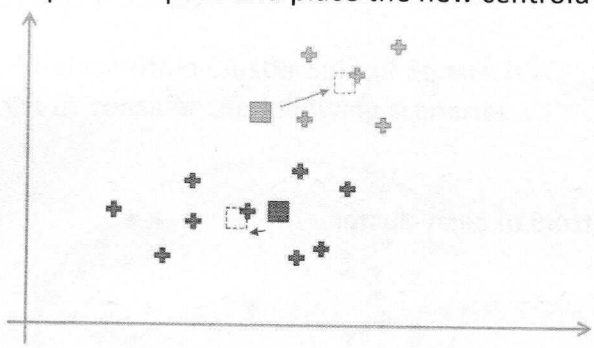


Step-5: Reassign each data points to the new closest centroid.
If any reassignment took place, go to Step-4, otherwise go to Stop.

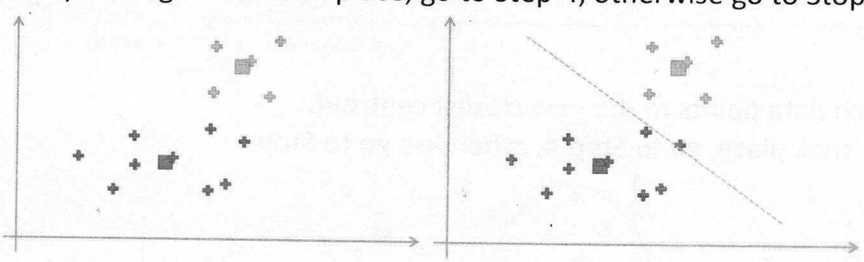


1 Mark

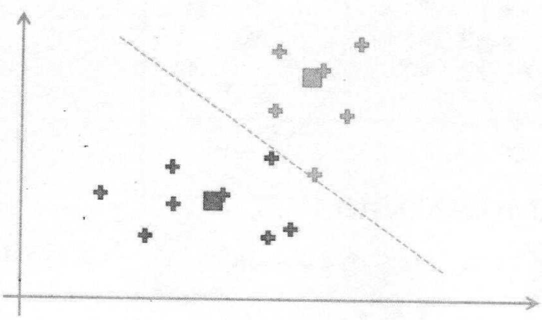
Step-4: Compute and place the new centroid of each cluster



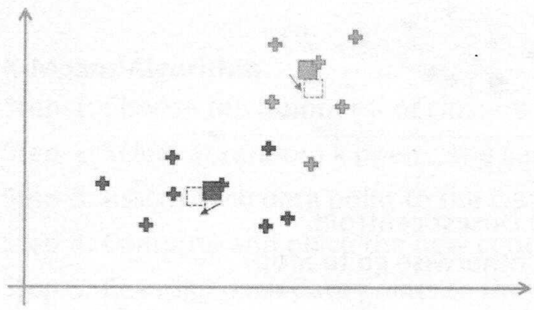
Step-5: Reassign each data points to the new closest centroid.
If any reassignment took place, go to Step-4, otherwise go to Stop.



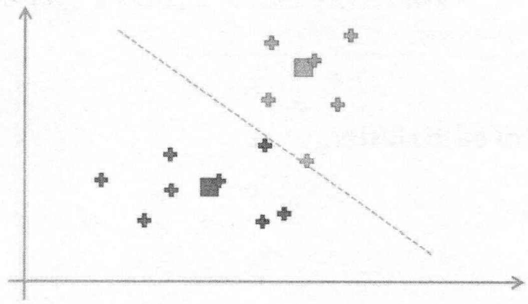
1 Mark



Step-4: Compute and place the new centroid of each cluster

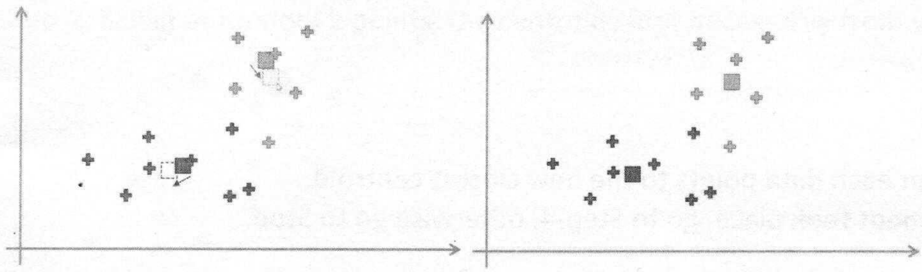


Step-5: Reassign each data points to the new closest centroid.
If any reassignment took place, go to Step-4, otherwise go to Stop.

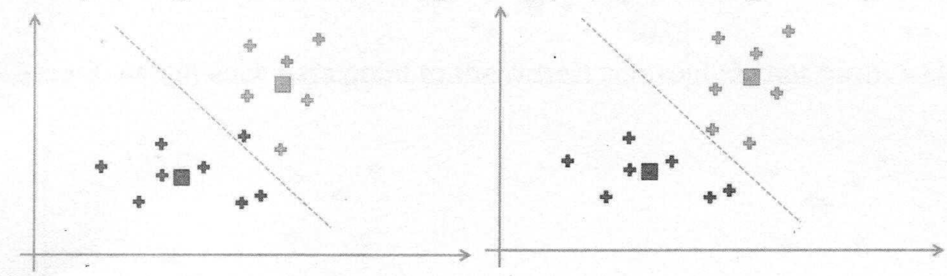


1 Mark

Step-4: Compute and place the new centroid of each cluster

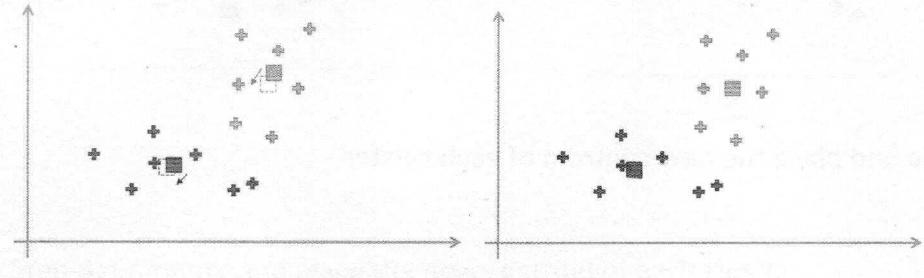


Step-5: Reassign each data points to the new closest centroid.
If any reassignment took place, go to Step-4, otherwise go to Stop.



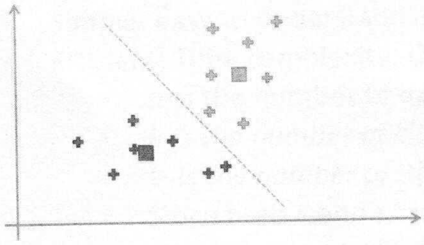
1 Mark

Step-4: Compute and place the new centroid of each cluster



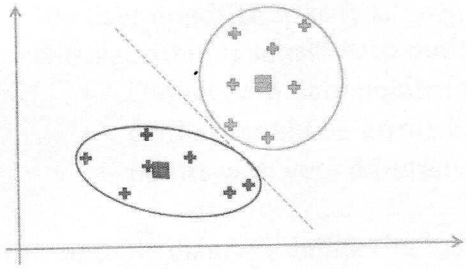
Step-5: Reassign each data points to the new closest centroid.

If any reassignment took place, go to Step-4, otherwise go to Stop.

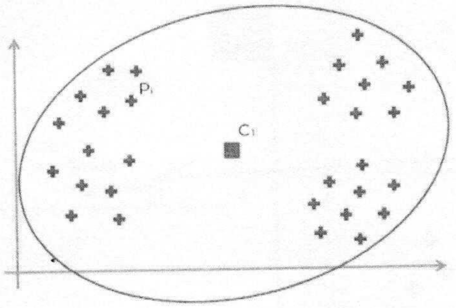


1 Mark

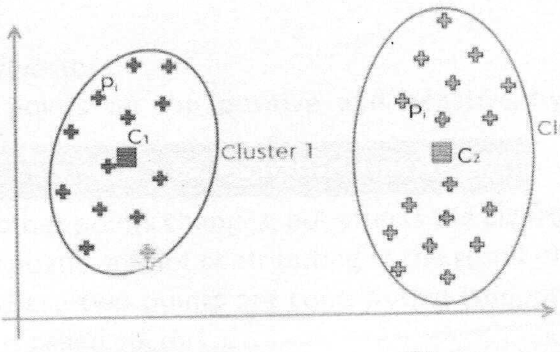
Stop: your model is ready



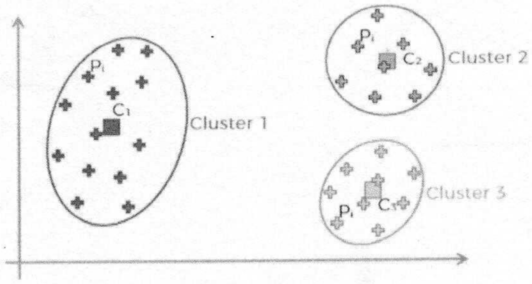
4 What is Within Cluster Sum of Square (WCSS)? Write the use of WCSS in clustering. 6
Let us consider the following scenarios...



$$WCSS_1 = \sum_{P_i \text{ in cluster } 1} \text{distance}(P_i, C1)^2$$



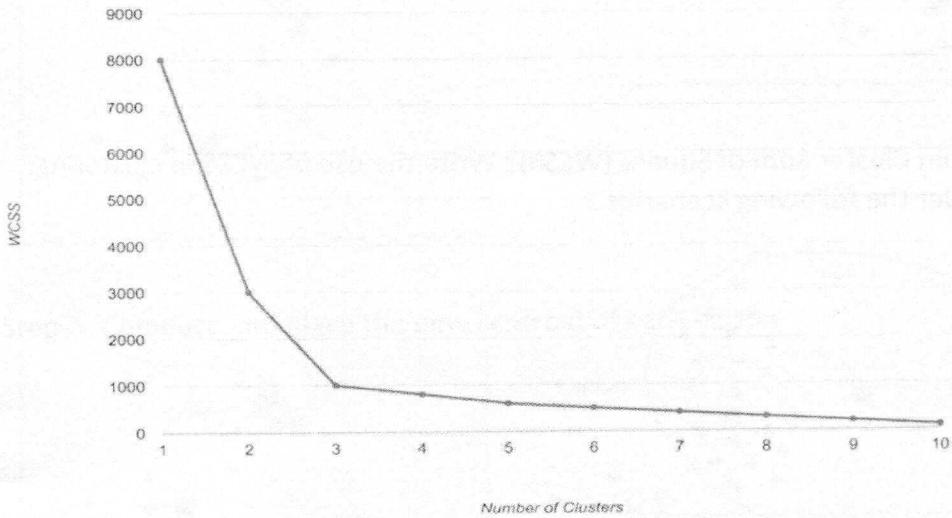
$$WCSS = \sum_{P_i \text{ in cluster } 1} \text{distance}(P_i, C1)^2 + \sum_{P_i \text{ in cluster } 2} \text{distance}(P_i, C_2)^2$$

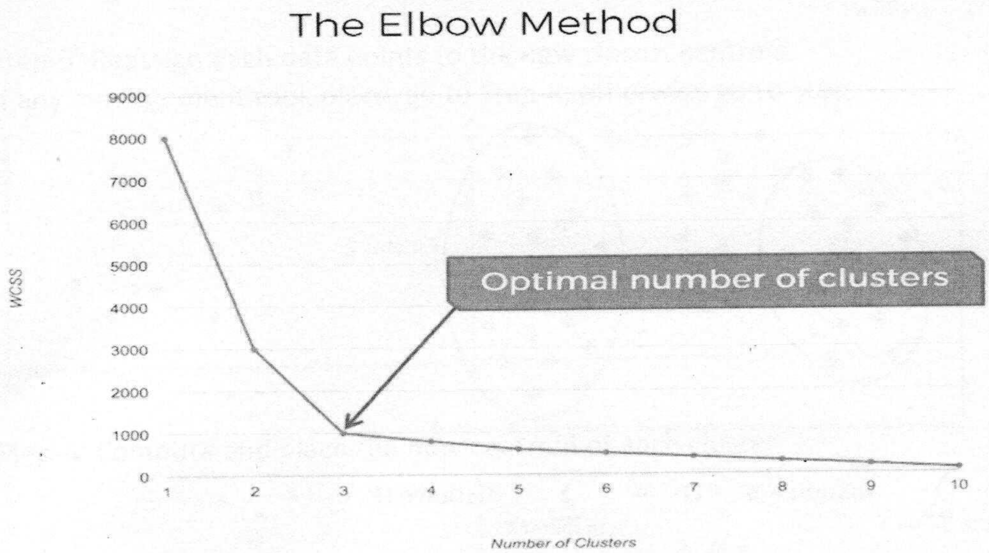


3 Marks

$$WCSS = \sum_{P_i \text{ in cluster } 1} \text{distance}(P_i, C1)^2 + \sum_{P_i \text{ in cluster } 2} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in cluster } 3} \text{distance}(P_i, C_3)^2$$

3 Marks



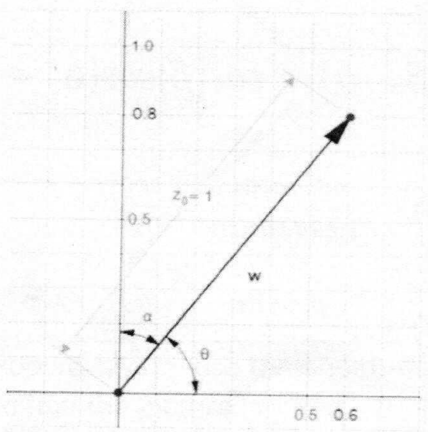


Use: to determine optimal number of clusters for k-means clustering.

Computing the direction vector

We will now compute the direction of the vector $\mathbf{u}$ from Figure

$$cos(\theta) = \frac{u_1}{\|u\|} = \frac{3}{5} = 0.6$$



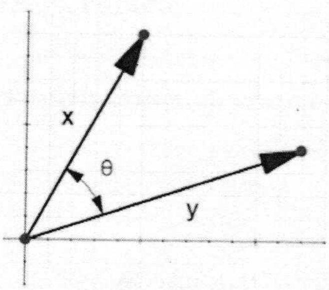
$$cos(\alpha) = \frac{u_2}{\|u\|} = \frac{4}{5} = 0.8$$

The direction of $\mathbf{u}(3,4)$ is the vector $\mathbf{w}(0.6, 0.8)$

8 If we have two vectors x and y and there is an angle θ between them, find their dot product. 5

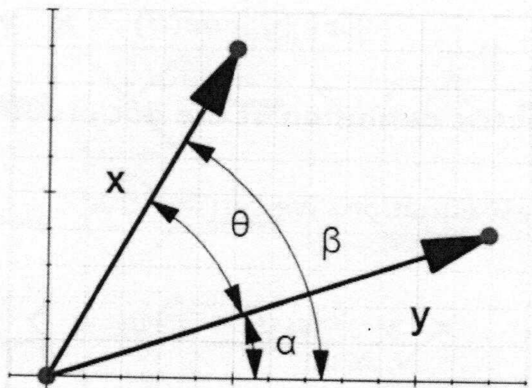
if we have two vectors X and Y and there is an angle θ between them, their dot product is :

$$x \cdot y = \|x\| \ \|y\| \cos(\theta)$$



$$\cos(\theta) = \frac{adjacent}{hypotenuse}$$

$$\theta = \beta - \alpha$$



2 Marks

$$cos(\beta - \alpha) = cos(\beta)cos(\alpha) + sin(\beta)sin(\alpha)$$

$$cos(\beta - \alpha) = cos(\beta)cos(\alpha) + sin(\beta)sin(\alpha)$$

$$cos(\beta) = \frac{adjacent}{hypotenuse} = \frac{x_1}{\|x\|} \qquad cos(\alpha) = \frac{adjacent}{hypotenuse} = \frac{y_1}{\|y\|}$$

$$sin(\beta) = \frac{opposite}{hypotenuse} = \frac{x_2}{\|x\|} \qquad sin(\alpha) = \frac{opposite}{hypotenuse} = \frac{y_2}{\|y\|}$$

$$cos(\theta) = cos(\beta - \alpha) = cos(\beta)cos(\alpha) + sin(\beta)sin(\alpha)$$

$$cos(\theta) = \frac{x_1}{\|x\|}\frac{y_1}{\|y\|} + \frac{x_2}{\|x\|}\frac{y_2}{\|y\|}$$

$$cos(\theta) = \frac{x_1y_1 + x_2y_2}{\|x\|\|y\|}$$

3 Marks

If we multiply both sides by $\|x\|\|y\|$ we get:

$$\|x\|\|y\|cos(\theta) = x_1y_1 + x_2y_2$$

Which is the same as :

$$\|x\|\|y\|cos(\theta) = \mathbf{x} \cdot \mathbf{y}$$

We just found the geometric definition of the dot product !

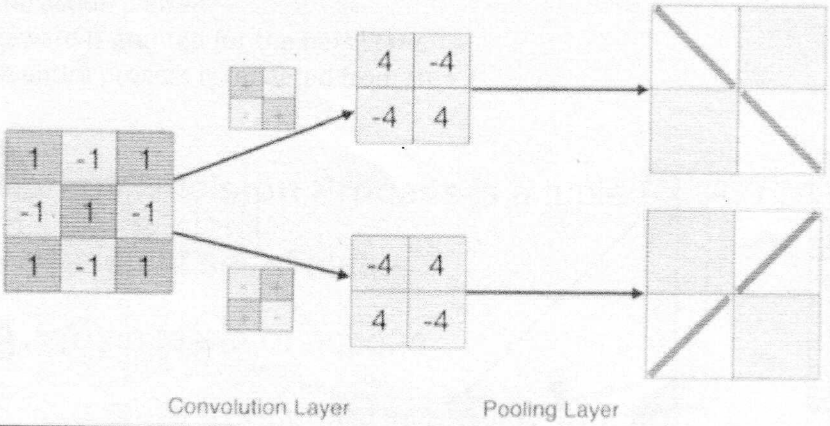
Eventually from the two last equations we can see that :

$$\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 = \sum_{i=1}^{2}(x_iy_i)$$

9 Show the convolutional, pooling and fully connected layers model for the image given below 12

1	-1	1
-1	1	-1
1	-1	1

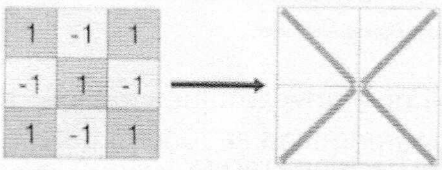
Convolutional and Pooling layer break the images into smaller pieces. Whereas the fully connected layer use some logic to classify the image. 1 Mark



3 Marks

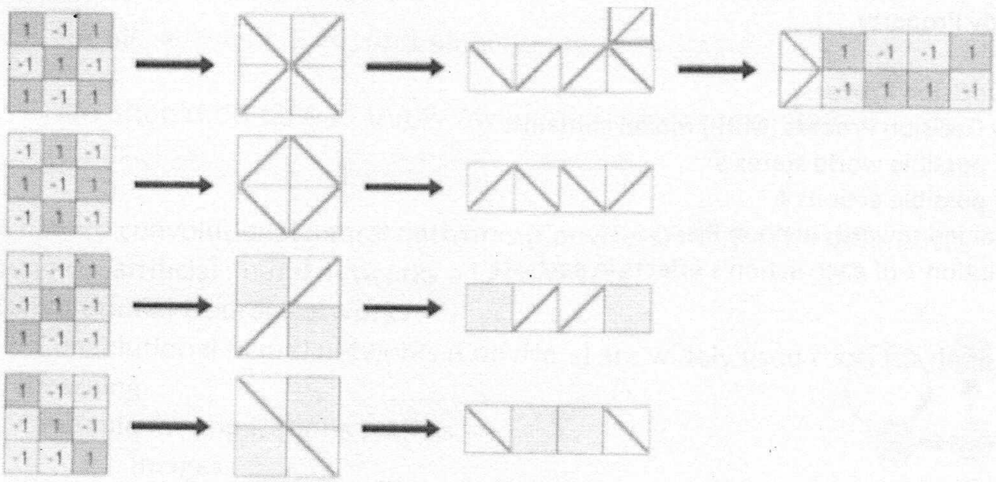
Convolutional layer – filters and slides through the image

Pooling layer – use threshold to find the picture (+4 find picture / -4 find nothing)



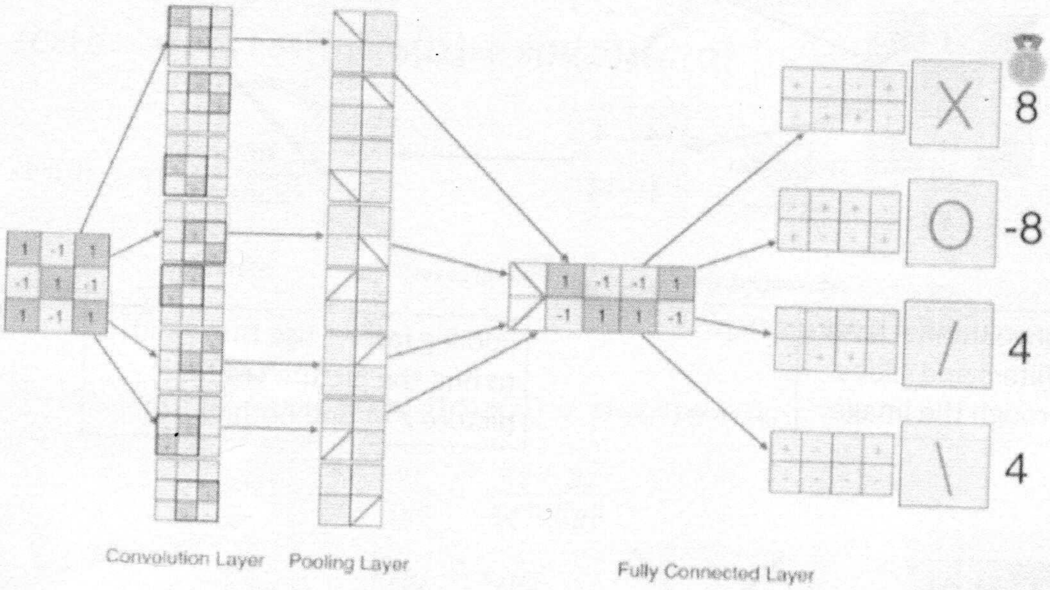
2 Marks

Finally take two images on the right and over impose them, now 3 x 3 pixel



3 Marks

The pooling and convolutional layers encoded our 3x3 images in the left as images in the right.



3 Marks

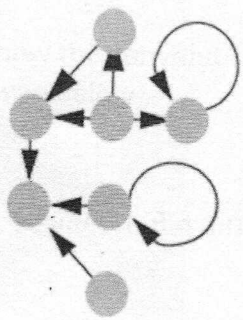
10 Write a short note on Markov Property and Markov Decision Process 8

Markov Property: The effects of an action taken in a state depend only on that state and not on the prior history.

For example, if an unmanned aircraft is trying to remain level, all it needs to know is its current state, which might include how level it currently is, and what influences (momentum, wind, gravity) are acting upon its state of level-ness. This analysis displays the Markov Property. 2 Marks

Markov Decision Process:
A Markov Decision Process (MDP) model contains:
- A set of possible world states S
- A set of possible actions A
- A real valued reward function R(s,a)
- A description T of each action's effects in each state.

2 Marks



Once the states, actions, probability distribution, and rewards have been determined, the last task is to run the process. A time step is determined and the state is monitored at each time step.
In a simulation,
1. The initial state is chosen randomly from the set of possible states. 1 Mark
2. Based on that state, an action is chosen.
3. The next state is determined based on the probability distribution for the given state

and the action chosen.
4. A reward is granted for the next state.
5. The entire process is repeated from step 2

A **Markov Decision Process** is a tuple $(\mathcal{S}, \mathcal{A}, T, r, \gamma)$ 3 Marks

- $\mathcal{S}$ is a finite set of states
- $\mathcal{A}$ is a finite set of actions
- T is a state transition probability function

$$T(s'|s,a) = \mathbb{P}[S_{t+1} = s' | S_t = s, A_t = a]$$

- r is a reward function

$$r(s,a) = \mathbb{E}[R_{t+1} | S_t = s, A_t = a]$$

- γ is a discount factor $\gamma \in [0,1]$

- A state captures whatever information is available to the agent at step t about its environment. The state can include immediate "sensations," highly processed sensations, and structures built up over time from sequences of sensations, memories etc.

- A state should summarize past sensations so as to retain all "essential" information, i.e., it should have the **Markov Property**:

$$\mathbb{P}[R_{t+1} = r, S_{t+1} = s' | S_0, A_0, R_1, ..., S_{t-1}, A_{t-1}, R_t, S_t, A_t] = \mathbb{P}[R_{t+1} = r, S_{t+1} = s' | S_t, A_t]$$

for all $s' \in \mathcal{S}, r \in \mathcal{R}$, and all histories

- We should be able to throw away the history once state is known

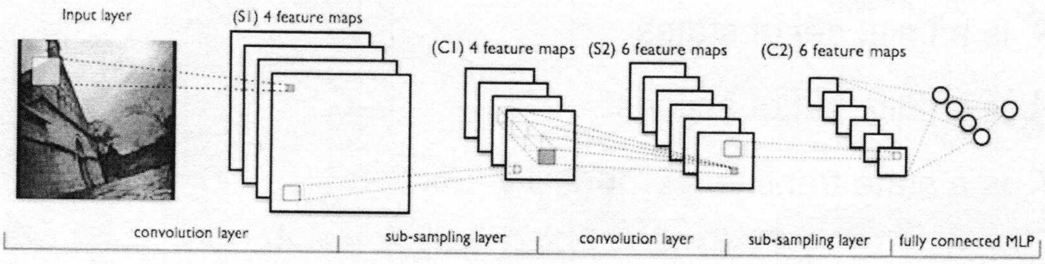
11 What are convolutional neural networks (ConvNets)? Write the difference between artificial neural networks and ConvNets 10

Convolutional Neural Networks

- Convolutional neural networks (ConvNets) are widely used tools for deep learning.
- Suitable for applications such as 2 Marks
  - Images
  - Text
  - Signals, and
  - other continuous responses - as inputs

- In a neural network with fully-connected neurons, the number of parameters (weights) can increase quickly as the size of the input increases.
- A convolutional neural network reduces the number of parameters with the reduced number of connections, shared weights, and downsampling.

- A ConvNet consists of multiple layers, such as convolutional layers, max-pooling or average-pooling layers, and fully-connected layers.



3 Marks

- They typically have five, six or seven layers
- Number of layers which makes fully-connected neural networks almost impossible to train
- Compared to standard feedforward neural networks with same sized layers,
  - CNNs have much fewer connections and parameters
  - so they are easier to train,
  - while their theoretically-best performance is likely to be only slightly worse.

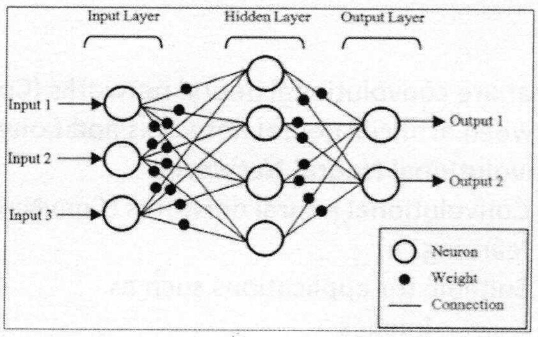
Artificial Neural Network:

ANN is a computing system made up of a number of simple, *highly interconnected processing elements*, which process information by their dynamic state response to external *inputs*.

2 Marks

- ANNs are modeled on the ***parallel architecture of animal brains***, not necessarily human ones.
- The network is based on a simple form of ***inputs*** and ***outputs***.
-

Artificial Neural Network
- Pool of simple processing units
- Communication to each other over a large number of weighted connections.



3 Marks

12 What is Deep Learning? Explain any four applications of deep learning 10

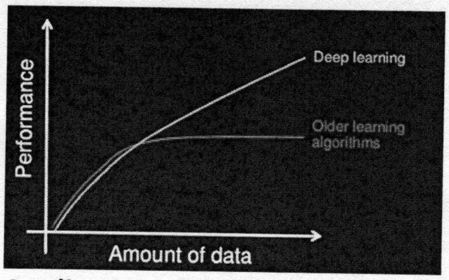
Deep Learning:

"very large neural networks and huge amounts of data that we have access to"

The term ***deep learning****, is just a large deep neural net.*

Deep refers to the ***number of layers***

Deep Learning is a deep neural networks generally. 2 Marks



Applications (1): Sentiment Analysis 2 Marks

The process of identifying and categorizing **opinions expressed in a piece of text**, to determine whether the writer's attitude towards a particular topic, product, etc. is **positive, negative, or neutral**.

The best way to hope for any chance of enjoying this film is by lowering your expectations.

The show starts out as competent but unremarkable and gradually grows into something of considerable power.

Applications (2): Question Answering 2 Marks

Yesterday Joy travelled to school.
Yesterday Happy went back to the beach.
This morning Joy travelled to market.
Bill went back to the cinema yesterday.
Happy went to the Hotel this morning.
Joy went back to the dining room this afternoon.

Question: Where was Happy before went to Hotel?

Applications (3): Language Translation 2 Marks

English: Can a man live without food?

Translated -- Kannada - ಒಬ್ಬ ಮನುಷ್ಯನು ಆಹಾರವಿಲ್ಲದೆಯೇ ಜೀವಿಸಬಹುದೇ?

Applications (4): Object Detection 2 Marks