National Testing Agency

Question Paper Name:	Neural Networks and Deep Learning
Subject Name:	Neural Networks and Deep Learning
Creation Date:	2019-03-30 13:12:44
Duration:	180
Total Marks:	100
Display Marks:	Yes

Neural Networks and Deep Learning

Group Number :	1
Group Id :	90958220
Group Maximum Duration :	0
Group Minimum Duration :	120
Revisit allowed for view? :	No
Revisit allowed for edit? :	No
Break time:	0
Group Marks:	100

Neural Networks and Deep Learning

Section Id :	90958220
Section Number :	1
Section type :	Online
Mandatory or Optional:	Mandatory
Number of Questions:	53
Number of Questions to be attempted:	53
Section Marks:	100
Display Number Panel:	Yes
Group All Questions:	No

Sub-Section Number:	1
Sub-Section Id:	90958220
Question Shuffling Allowed :	Yes

Question Number : 1 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

What is the main point of difference between human & machine intelligence?

a. Humans have emotions

b. Humans have more IQ & intellect

c. Humans have sense organs

d. Humans perceive everything as a pattern while machine perceive it merely as data

Options :

1. A
2. B
3. C
4. D

Question Number : 2 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

What is unsupervised learning?

a. Features of group are explicitly stated

b. Number of groups may be known

c. Neither features of group nor number of groups is known

d. None of the above

Options :

1. A
2. B
3. C
4. D

Question Number : 3 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

What are the issues on which biological networks prove to be superior to AI networks?

a. Robustness & fault tolerance

b. Flexibility

c. Collective computation

d. All of the above

Options :

1. A
2. B

3. C

4. D

Question Number : 4 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

The fundamental unit of biological neural network is

a. Brain

b. Nucleus

c. Neuron

d. Axon

Options :

1. A

2. B

3. C

4. D

Question Number : 5 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

What is the purpose of Axons?

a. Receptors

b. Neuro-transmitter

c. Transmission (fan out)

d. None of the above

Options :

1. A

2. B

3. C

4. D

Question Number : 6 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

When a particular neuron gets repeatedly fired, the strength of neuron to fire in future increases. What is the name of the above mechanism?

a. Error correction learning

b. Memory based learning

c. Delta Rule

d. Hebb's rule learning

Options :

1. A
2. B
3. C
4. D

Question Number : 7 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

What was the name of the first neuron model whose response is non-linear activation of weighted sum of inputs?

a. McCulloch-pitts neuron model

b. Marvin Minsky neuron model

c. Hopfield model of neuron

d. None of the above

Options :

1. A
2. B
3. C
4. D

Question Number : 8 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

Ackley and Hinton are jointly known for their contribution to:

a. Perceptron

b. Boltzman machine

c. Learning algorithms

d. None of the Above

Options :

1. A

2. B
3. C
4. D

Question Number : 9 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical
Correct Marks : 1 Wrong Marks : 0

Positive sign of weight indicates?

a. Excitatory input

b. Inhibitory input

c. Can be either excitatory or inhibitory as such

d. None of the Above

Options :
1. A
2. B
3. C
4. D

Question Number : 10 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical
Correct Marks : 1 Wrong Marks : 0

The process of adjusting the weights is known as?

a. Activation

b. Synchronisation

c. Learning

d. None of the above

Options :
1. A
2. B
3. C
4. D

Question Number : 11 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical
Correct Marks : 1 Wrong Marks : 0

For image classification problem (one class out of many), what non-linearity should be used in the final layer of a deep neural network?

a. ReLU
b. Sigmoid
c. Tanh
d. softmax

Options :

1. A
2. B
3. C
4. D

Question Number : 12 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

. What property of CNN's make them particularly useful for image recognition problems?

a. Deep architecture
b. Shift invariance
c. Short term memory
d. Scale independence

Options :

1. A
2. B
3. C
4. D

Question Number : 13 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

How many trainable parameters are there in a max-pooling layer of filter size F=3 and stride size S=3?

a. 3
b. 6
c. 9
d. None of the above

Options :

1. A
2. B
3. C
4. D

Question Number : 14 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

The derivative of the sigmoid non-linearity $\sigma(x)$ is given by

a. $\sigma(x)$
b. $\sigma(x)(1-\sigma(x))$
c. $1/(1+\sigma(x))$
d. $\sigma(x)^2$

Options :

1. A
2. B
3. C
4. D

Question Number : 15 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

A classification problem has 3 output classes. With one-hot encoding, the output vectors will be represented as:

a. [0,0,1], [0,1,0], [1,0,0]
b. [0], [1], [2]
c. [0,0], [0,1], [1,0]
d. None of the above

Options :

1. A
2. B
3. C
4. D

Question Number : 16 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

A single neuron with three inputs has weights 2, 4 and 5. The transfer function is linear with the constant of proportionality being equal to 3. The inputs are 2, 8 and 10 respectively. The output will be:

a. 125
b. 258
c. 86
d. 192

Options :

1. A
2. B
3. C
4. D

Question Number : 17 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

Why is the XOR problem exceptionally interesting to neural network researchers?

a. Because it can be expressed in a way that allows you to use a neural network
b. Because it is complex binary operation that cannot be solved using neural networks
c. Because it can be solved by a single layer perceptron
d. Because it is the simplest linearly inseparable problem that exists.

Options :

1. A
2. B
3. C
4. D

Question Number : 18 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

The network that involves backward links from output to the input and hidden layers is called:

a. Self organizing maps
b. Perceptrons
c. Recurrent neural network
d. Multi layered perceptron

Options :

1. A
2. B
3. C
4. D

Question Number : 19 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

What is the advantage of a multi-layered perceptron over single layer perceptron?

a. Training a multi-layered perceptron is faster than the training of a single layer perceptron.
b. A multi-layered perceptron can classify data that are not linearly separable
c. Multi-layered perceptron has many local minima
d. Multiple neurons are needed if there is more than one input variable, as a perceptron only accepts a single input (plus a bias value).

Options :

1. A
2. B
3. C
4. D

Question Number : 20 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

How does an LSTM handle the vanishing gradient problem that occurs in traditional Recurrent Neural Networks?

a. By the use of forget gates which have sigmoidal activation
b. By maintaining a memory cell whose activation function is the identity function with a derivative of 1.0.(considering forget gate is on)
c. By the use of output gates who have tanh activation
d. None of them are true

Options :

1. A
2. B
3. C
4. D

Question Number : 21 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

The neighbourhood function in Kohonen Self-organizing map is

a. $h_\gamma = e^{-\frac{d_\gamma^2}{2\sigma^2}}$, where σ represents the width of the neuron.
b. a fixed value less than 1.0
c. $h_\gamma = e^{-\frac{d_\gamma}{2\sigma^2}}$, where σ represents the width of the neuron.
d. $h_\gamma = 1.0$

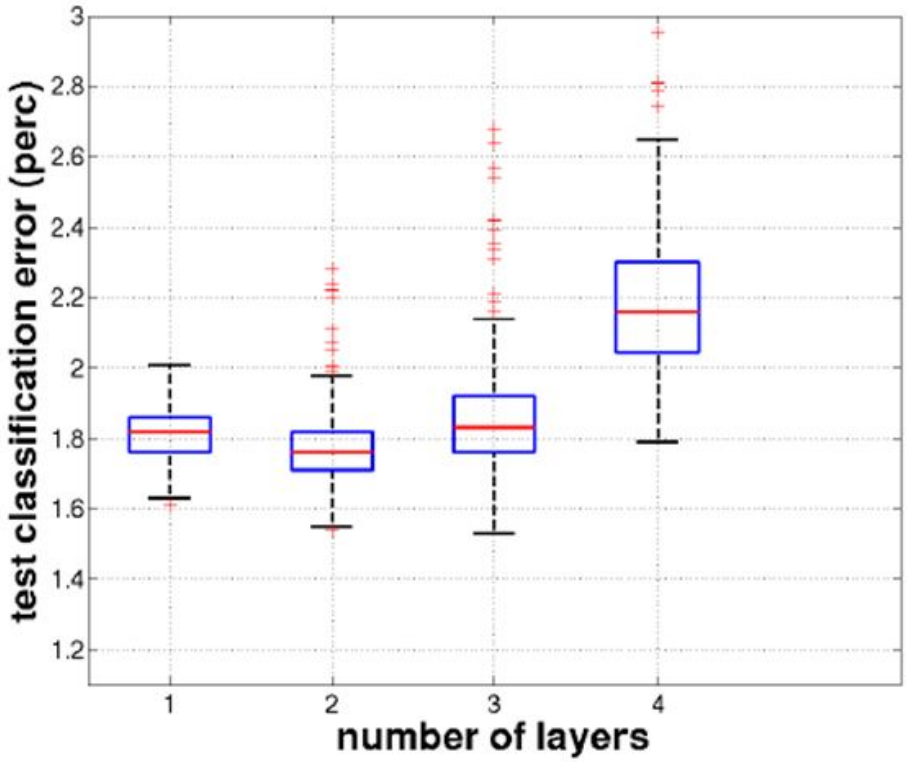
Options :

1. A
2. B
3. C
4. D

Question Number : 22 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

Mark the correct statement based on following figure:



a. Increasing depth seems to increase the probability of finding poor apparent local minima.
b. Increasing depth surely increases the probability of finding poor apparent local minima.
c. Decreasing depth surely decreases the probability of finding poor apparent local minima.
d. Increasing depth seems to increase the probability of finding global minima.

Options :

1. A
2. B
3. C
4. D

Question Number : 23 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

Who is the father of LSTM network?

a. Jürgen Schmidhuber
b. John Hopfield
c. David Rumelhart
d. Jeff Hinton

Options :

1. A
2. B

3. C
4. D

Question Number : 24 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 1 Wrong Marks : 0

In the keras library which is used to construct neural networks in python, what does the following command create?

model.add(Dense(512))

a. Fully connected layer, input size = 512, activation = sigmoid
b. Semi connected layer, output size = 512, activation = linear
c. Fully connected layer, output size = 512, activation = linear
d. Fully connected layer, output size = 512, activation = sigmoid

Options :

1. A
2. B
3. C
4. D

Sub-Section Number:	2
Sub-Section Id:	90958221
Question Shuffling Allowed :	Yes

Question Number : 25 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

In the following python code using a sequential model from keras library, what is the total number of weights to be trained(including the biases)?

```
model.add(Dense(10, input_dim= 100, activation='relu'))
model.add(Dense(8, activation='relu'))
model.add(Dense(5, activation='relu'))
model.add(Dense(5, activation='softmax'))
```

a. 1145
b. 1150
c. 1173
d. 2000

Options :

1. A
2. B
3. C
4. D

Question Number : 26 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

Given $p(x_1 = 0) = \frac{1}{4}$ and $p(x_2 = 0) = \frac{1}{2}$, the joint distribution of x_1 and x_2 in the order (0,0), (0,1), (1,0) and (1,1) are (assuming x_1 is independent of x_2):

a. $p(x_1, x_2) = \begin{pmatrix} 0.375 \\ 0.125 \\ 0.250 \\ 0.250 \end{pmatrix}$

b. $p(x_1, x_2) = \begin{pmatrix} 0.250 \\ 0.250 \\ 0.250 \\ 0.250 \end{pmatrix}$

c. $p(x_1, x_2) = \begin{pmatrix} 0.125 \\ 0.375 \\ 0.250 \\ 0.250 \end{pmatrix}$

d. $p(x_1, x_2) = \begin{pmatrix} 0.125 \\ 0.125 \\ 0.375 \\ 0.375 \end{pmatrix}$

Options :

1. A
2. B
3. C
4. D

Question Number : 27 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

Data is fitted to the following linear model,

$$Aw = y, A \epsilon \mathbb{R}^{n \times m}; y \epsilon \mathbb{R}^{n}; w \epsilon \mathbb{R}^{m}$$

Given $n \gg m$,the model parameter vector w has least square estimate as

a. $w = A^T(A^{-1}A^T)y$
b. $w = (A^TA)^{-1}A^Ty$
c. $w = A^T(AA^T)^{-1}y$
d. $w = (A^TA)A^Ty^{-1}$

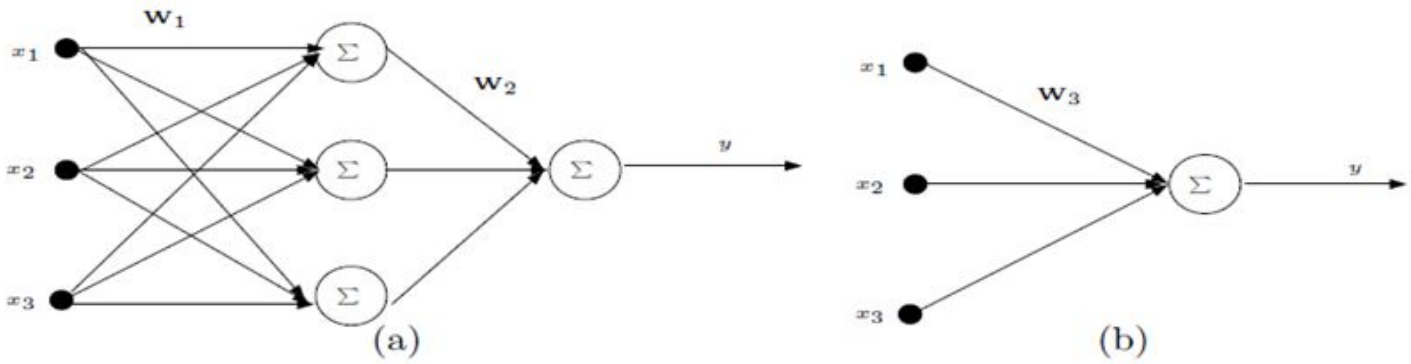
Options :

1. A
2. B
3. C
4. D

Question Number : 28 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

The network in Figure (a) is equivalent to the network in Figure (b). Note that the output of each neuron is just the sum of all its inputs.



Then which of the following statements is true:

a. $W_3 = W_1 + W_2$

b. $W_3 = W_1 \times W_2$

c. $W_3 = W_1 \otimes W_2$

d. $W_3 = W_1^T W_2$

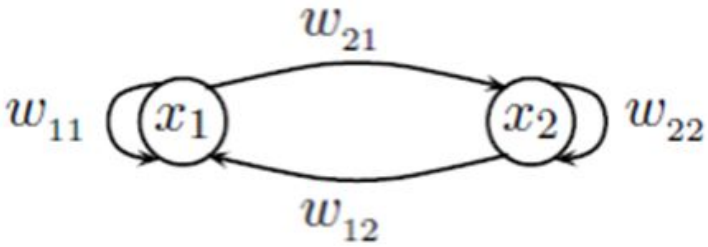
Options :

1. A
2. B
3. C
4. D

Question Number : 29 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

In the following recurrent network, the activation functions are linear. The instantaneous cost function is $E(t+1) = \frac{1}{2}\left(x_1^d(t+1) - x_1(t+1)\right)^2 + \frac{1}{2}\left(x_2^d(t+1) - x_2(t+1)\right)^2$. The weight update equation for w_{21} is



a. $w_{21} \leftarrow w_{21} + \boldsymbol{\eta}\left(x_1^d(t+1) - x_1(t+1)\right)\frac{\partial x_1(t+1)}{\partial w_{21}} + \boldsymbol{\eta}\left(x_2^d(t+1) - x_2(t+1)\right)\frac{\partial x_2(t+1)}{\partial w_{21}}$
where $\frac{\partial x_1(t+1)}{\partial w_{21}} = w_{11}\frac{\partial x_1(t)}{\partial w_{21}} + w_{12}\frac{\partial x_2(t)}{\partial w_{21}}$ and $\frac{\partial x_2(t+1)}{\partial w_{21}} = x_1(t) + w_{21}\frac{\partial x_1(t)}{\partial w_{21}} + w_{22}\frac{\partial x_2(t)}{\partial w_{21}}$

b. $w_{21} \leftarrow w_{21} + \boldsymbol{\eta}\left(x_1^d(t+1) - x_1(t+1)\right)\frac{\partial x_1(t+1)}{\partial w_{21}} + \boldsymbol{\eta}\left(x_2^d(t+1) - x_2(t+1)\right)\frac{\partial x_2(t+1)}{\partial w_{21}}$
where $\frac{\partial x_1(t+1)}{\partial w_{21}} = w_{11}\frac{\partial x_1(t)}{\partial w_{21}} + w_{12}\frac{\partial x_2(t)}{\partial w_{21}}$ and $\frac{\partial x_2(t+1)}{\partial w_{21}} = w_{21}\frac{\partial x_1(t)}{\partial w_{21}} + w_{22}\frac{\partial x_2(t)}{\partial w_{21}}$

c. $w_{21} \leftarrow w_{21} + \boldsymbol{\eta}\left(x_1^d(t+1) - x_1(t+1)\right)\frac{\partial x_1(t+1)}{\partial w_{21}}$
where $\frac{\partial x_1(t+1)}{\partial w_{21}} = w_{11}\frac{\partial x_1(t)}{\partial w_{21}} + w_{12}\frac{\partial x_2(t)}{\partial w_{21}}$

d. $w_{21} \leftarrow w_{21} + \boldsymbol{\eta}\left(x_2^d(t+1) - x_2(t+1)\right)\frac{\partial x_2(t+1)}{\partial w_{21}}$
where $\frac{\partial x_2(t+1)}{\partial w_{21}} = w_{21}\frac{\partial x_1(t)}{\partial w_{21}} + w_{22}\frac{\partial x_2(t)}{\partial w_{21}}$

Options :

1. A
2. B
3. C
4. D

Question Number : 30 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

Given a 4 layered neural network whose neurons are sigmoidally activated, the weight update law for the 2nd layer using back-propagation is given as: [Use the notation convention as taught in the class]

a. $w_{i_2 i_1}(t+1) = w_{i_2 i_1}(t) + \eta \delta_{i_2}\, v_{i_1}$ where $\delta_{i_2} = v_{i_2}(1 - v_{i_2}) \sum_{i_3=1}^{n_3} w_{i_3 i_2} \delta_{i_3}$

b. $w_{i_2 i_1}(t+1) = w_{i_2 i_1}(t) + \eta \delta_{i_2}\, v_{i_1}$ where $\delta_{i_2} = v_{i_2}(1 - v_{i_2}) \sum_{i_2=1}^{n_2} w_{i_3 i_2} \delta_{i_3}$

c. $w_{i_2 i_1}(t+1) = w_{i_2 i_1}(t) + \eta \delta_{i_1}\, v_{i_2}$ where $\delta_{i_1} = v_{i_1}(1 - v_{i_1}) \sum_{i_3=1}^{n_3} w_{i_3 i_2} \delta_{i_2}$

d. $w_{i_2 i_1}(t+1) = w_{i_2 i_1}(t) + \eta \delta_{i_1}\, v_{i_2}$ where $\delta_{i_1} = v_{i_1}(1 - v_{i_1}) \sum_{i_2=1}^{n_2} w_{i_3 i_2} \delta_{i_2}$

Options :

1. A
2. B
3. C
4. D

Question Number : 31 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

A nonlinear function is given as $y = (1-x)\, x^2;\ 0 < x < 1$. Representation of this function using 1-D KSOM based function approximation model consisting of three neurons whose weights are 0.25, 0.5 and 0.75 respectively is:

$$y = \frac{\sum h_\gamma (y_\gamma + A_\gamma (x - w_\gamma))}{\sum h_\gamma}$$

Neighborhood function value is 1 for the winner and nearest neighbour is 0.25. The Jacobian associated with each neuron are

a. $A_\gamma = 0.625$ at $w_\gamma = 0.25$; $A_\gamma = 0.5$ at $w_\gamma = 0.5$; $A_\gamma = 0.1875$ at $w_\gamma = 0.75$

b. $A_\gamma = 0.0468$ at $w_\gamma = 0.25$; $A_\gamma = 0.125$ at $w_\gamma = 0.5$; $A_\gamma = 0.14$ at $w_\gamma = 0.75$

c. $A_\gamma = 0.5$ at $w_\gamma = 0.25$; $A_\gamma = 0.2$ at $w_\gamma = 0.5$; $A_\gamma = 0.5$ at $w_\gamma = 0.75$

d. $A_\gamma = 0.3125$ at $w_\gamma = 0.25$; $A_\gamma = 0.25$ at $w_\gamma = 0.5$; $A_\gamma = -0.1875$ at $w_\gamma = 0.75$

Options :

1. A
2. B
3. C
4. D

Question Number : 32 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

The function approximation using an n-D KSOM network is given as:

$$y = \frac{\sum_\gamma h_\gamma (y_\gamma + A_\gamma (x - w_\gamma))}{\sum_\gamma h_\gamma}$$

Given the cost function as:

$$E = \frac{1}{2}\tilde{y}^T\tilde{y}, \tilde{y} = y^d - y$$

The weight update for the Jacobian is

a. $A_\gamma \leftarrow A_\gamma + \eta\tilde{y}(x - w_\gamma)^T\left(\frac{h_\gamma}{\sum_\gamma h_\gamma}\right)$

b. $A_\gamma \leftarrow A_\gamma - \eta\tilde{y}\frac{\sum_\gamma h_\gamma (x - w_\gamma)}{\sum_\gamma h_\gamma}$

c. $A_\gamma \leftarrow A_\gamma + \eta\tilde{y}\frac{\sum_\gamma h_\gamma (x - w_\gamma)^T}{\sum_\gamma h_\gamma}$

d. $A_\gamma \leftarrow A_\gamma - \eta\tilde{y}(x - w_\gamma)^T\left(\frac{h_\gamma}{\sum_\gamma h_\gamma}\right)$

Options :

1. A
2. B
3. C
4. D

Question Number : 33 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

The cost function to update the Jacobian in kinematic control of a robot arm is

a. $E = \frac{1}{2} \| \theta_0^{out} - \frac{1}{s}\sum_\gamma h_\gamma \left(\theta_\gamma + A_\gamma (v_0 - w_\gamma)\right) \|^2; s = \sum_\gamma h_\gamma$

b. $E = \frac{1}{2} \| \Delta\theta^{out} - \frac{1}{s}\sum_\gamma h_\gamma A_\gamma \Delta v \|^2; s = \sum_\gamma h_\gamma$

c. $E = \frac{1}{2}(y^d - y)^2$

d. $E = \frac{1}{2} \| y^d - y \|^2$

Options :

1. A
2. B
3. C
4. D

Question Number : 34 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

The adaptive learning rate for a feedforward network with associated Lyapunov function is

a. $\eta = \mu \frac{\|\tilde{y}\|}{\|J^{-1}\tilde{y}\|}$

b. $\eta = \mu \frac{1}{\|J^{T}\tilde{y}\|}$

c. $\eta = \mu \frac{\|\tilde{y}\|}{\|J^{T}\tilde{y}\|}$

d. $\eta = \mu \frac{\|\tilde{y}\|^2}{\|J^{T}\tilde{y}\|^2}$

Options :

1. A
2. B
3. C
4. D

Question Number : 35 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

Write the ReLU activation function f with input $x_i \in \mathbb{R}, \forall i = 0,1,\ldots,N$. Here, $exp(x)$ is exponential function and $\delta(x)$ is the Dirac delta function

a. $\frac{exp(x_i)}{\sum_{j=1}^{N} exp(x_j)}$

b. x_i

c. $\frac{1}{1+exp(-x_i)}$

d. $x_i \int_{-\infty}^{x_i} \delta(x)dx$

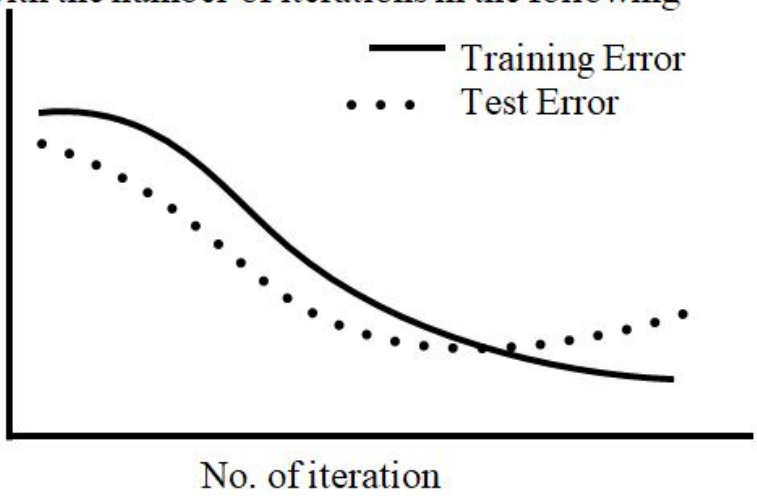
Options :

1. A
2. B
3. C
4. D

Question Number : 36 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

What is the cause of the test error increasing with the number of iterations in the following plot?



a. Vanishing gradients
b. Underfitting
c. Overfitting
d. Local minima

Options :

1. A
2. B
3. C
4. D

Question Number : 37 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

What is the total number of tunable parameters in a fully-connected layer with 100 nodes and sigmoid non-linearity, when it gets the input from a layer with 10 nodes?

a. 100
b. 110
c. 1000
d. 1100

Options :

1. A
2. B
3. C
4. D

Question Number : 38 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

The effective receptive field of a stack of three 3x3 convolution layers, with stride S=2 at each layer, is same as that of

a. One 3x3 convolution layer
b. One 7x7 convolution layer
c. One 15x15 convolution layer
d. None of the above

Options :

1. A
2. B
3. C

4. D

Question Number : 39 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

In the following figure, the total number of tunable parameters is approximately

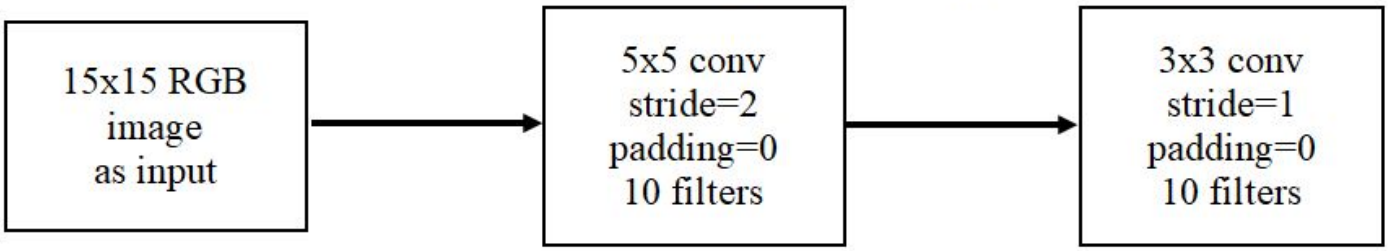


Fig. CNN

a. 1700
b. 1500
c. 1300
d. 1000

Options :

1. A
2. B
3. C
4. D

Question Number : 40 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

See the following code fragment (python).

```
x = Input(shape=(18,18,3))
y = Conv2D(filters=3,kernel=(9,9), strides=1, padding='same', activation='relu')(x)
```

Here, padding='same' means the size (spatially) of the input is preserved in the output. How many rows/columns of zeros are padded along the border?

a. P=2
b. P=4
c. P=6
d. P=8

Options :

1. A
2. B
3. C
4. D

Question Number : 41 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

In the following pseudo code of stochastic gradient descent:

```
X = input data with 50 samples
for i in range(20):
    shuffle X and divide samples equally into 5 batches X_0, X_1,…, X_4
    for j in range(10):
        x = X_k, where k=(j modulo 5)
        update NN weights with gradients calculated using x
```

What is the number of epochs?

a. 10
b. 20
c. 40
d. 50

Options :

1. A
2. B
3. C
4. D

Question Number : 42 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

In the following network, what is the shape of output y? Here, padding is along the border.

```
x = Input(shape=(125,125,3))
x = Conv2D(filters=20, kernel=(5,5), strides=2, padding=1, activation='relu')(x)
y = MaxPooling2D(pool_size=(2, 2), strides=2)(x)
```

a. (31, 31, 20)
b. (62, 62, 20)
c. (61, 61, 20)
d. None of the above

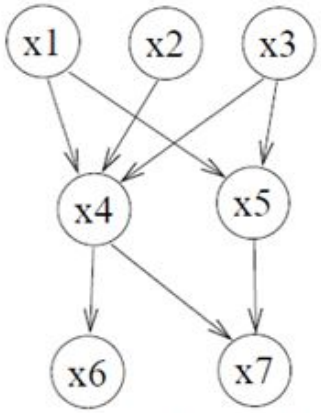
Options :

1. A
2. B
3. C
4. D

Question Number : 43 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

Given the following dependency graph, the joint distribution is given as



a. $p(x_1,x_2,x_3,x_4,x_5,x_6,x_7) = p(x_1)p(x_2)p(x_3)p(x_4)p(x_5)p(x_6)p(x_7)$

b. $p(x_1,x_2,x_3,x_4,x_5,x_6,x_7) =$
$p(x_1)p(x_2)p(x_3)p(x_4,x_1,x_2,x_3)p(x_5,x_1,x_3)p(x_6,x_4)p(x_7,x_4,x_5)$

c. $p(x_1,x_2,x_3,x_4,x_5,x_6,x_7) =$
$p(x_1|x_4,x_5)p(x_2|x_4)p(x_3|x_4,x_5)p(x_4|x_6,x_7)p(x_5|x_7)p(x_6)p(x_7)$

d. $p(x_1,x_2,x_3,x_4,x_5,x_6,x_7) =$
$p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3)p(x_5|x_1,x_3)p(x_6|x_4)p(x_7|x_4,x_5)$

Options :

1. A
2. B
3. C
4. D

Question Number : 44 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 2 Wrong Marks : 0

Assume a 4×3 RBM and a training vector $x^{(t)}$. Assuming a random number generator generates uniform random vectors $U_1 = \begin{pmatrix} 0.27 \\ 0.62 \\ 0.41 \end{pmatrix}$, $U_2 = \begin{pmatrix} 0.03 \\ 0.68 \\ 0.52 \\ 0.29 \end{pmatrix}$. Using these random vectors in the same order for one iteration of training of a 4×3 RBM given the training vector $x^{(t)}$. Assume initial weights and biases are zero vectors. Given $x^{(t)} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$, the values of probability of hidden unit activation, $h\left(x^{(t)}\right)$, and the reconstructed state, $\tilde{x}^{(t)}$, are respectively

a. $\begin{pmatrix} 0.5 \\ 0.5 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$

b. $\begin{pmatrix} 0.5 \\ 0.5 \\ 0.5 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

c. $\begin{pmatrix} 1 \\ 0.5 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$

d. $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$

Options :

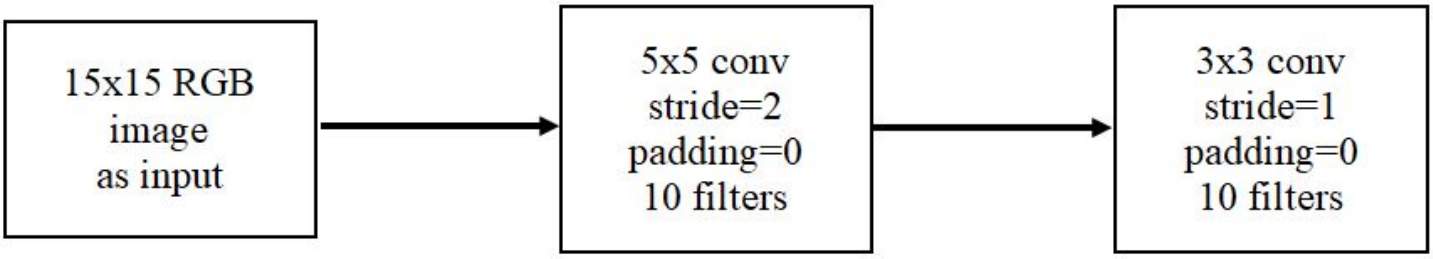
1. A
2. B
3. C
4. D

Sub-Section Number:	3
Sub-Section Id:	90958222
Question Shuffling Allowed :	Yes

Question Number : 45 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

In the following CNN architecture, the computational complexity (number of multiplications) in the forward pass is close to (here K=1,000):



a. 30 K
b. 25 K
c. 20 K
d. 15 K

Options :

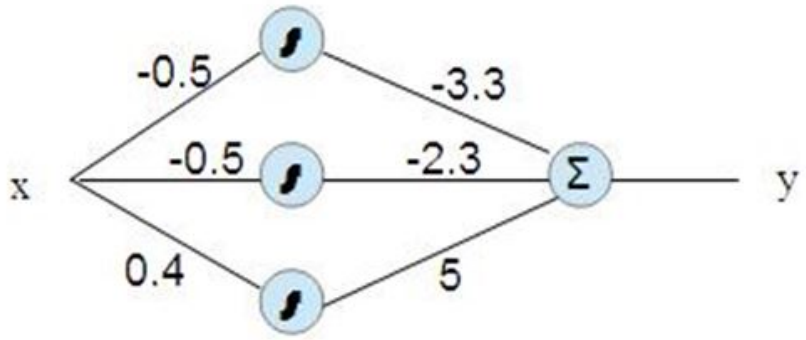
1. A
2. B
3. C
4. D

Question Number : 46 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

The following figure is a two layered network whose hidden neurons are sigmoidally activated while the output neuron is linear i.e. output is summation of all inputs to this neuron. The input to the neuron x is not known but the output is $y = 0.5$. You need to search the input space starting with $x = 0.5$ using gradient descent algorithm. Take $\eta = 0.5$;

$$x(t+1) = x(t) - \eta \frac{dE}{dx}; \; E = \frac{1}{2}(y^d - y)^2$$



The value of x after two iterations is

a. 0.65

b. 0.47

c. 1.02

d. 0.17

Options :

1. A
2. B
3. C
4. D

Question Number : 47 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

A process is described by $\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} 0.5 & -1 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$ where matrix parameters are unknown. Weights are updated using real-time recurrent learning. Take $u(0) = x_1(0) = x_2(0) = 1$ and $\eta = 1$. Assume initial parameter values to be

$$W_1 = \begin{bmatrix} 0.3 & -0.8 \\ 0 & 0.4 \end{bmatrix}; W_2 = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}.$$

Assume initial values of gradients to be zeros. The parameter values after one iteration are

a. $W_1 = \begin{bmatrix} 0.3 & -0.8 \\ 0.6 & -1 \end{bmatrix}; W_2 = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$

b. $W_1 = \begin{bmatrix} 0.3 & -0.8 \\ 0.6 & 1 \end{bmatrix}; W_2 = \begin{bmatrix} 0 \\ 1.1 \end{bmatrix}$

c. $W_1 = \begin{bmatrix} 0.3 & -0.8 \\ -0.6 & -1 \end{bmatrix}; W_2 = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$

d. $W_1 = \begin{bmatrix} 0.3 & 0.8 \\ -0.6 & -1 \end{bmatrix}; W_2 = \begin{bmatrix} 0 \\ 1.1 \end{bmatrix}$

Options :

1. A
2. B
3. C
4. D

Question Number : 48 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

Consider a ($4 \times 3 \times 2$ Deep Neural Network and a training vector $v_1 = [1010]$.

Assume a random number generator generates uniform random vectors $U_1 = \begin{pmatrix} 0.27 \\ 0.62 \\ 0.41 \end{pmatrix}$, $U_2 = \begin{pmatrix} 0.03 \\ 0.68 \\ 0.52 \\ 0.29 \end{pmatrix}$. Use these random vectors in the same order for one iteration of training of a (4×3 RBM) using $v_1 = [1010]$. The weights and biases of the 4×3 RBM after one iteration assuming initial weights and biases are zero vectors are [Use $\alpha = 1$ in the update rule]

a. $W = \begin{pmatrix} 0 & 0 & 0.5 & -0.5 \\ 0 & 0.1 & 0.5 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \end{pmatrix}, b = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, c = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$

b. $W = \begin{pmatrix} 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & 0.5 & -0.5 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, c = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$

c. $W = \begin{pmatrix} 0.5 & -0.5 & 0 & 0 \\ 0.5 & -0.5 & 0 & 0 \\ 0.5 & -0.5 & 0 & 0 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, c = \begin{pmatrix} 0.5 \\ 0 \\ 1 \\ -1 \end{pmatrix}$

d. $W = \begin{pmatrix} 0 & 0 & 0.5 & -0.5 \\ 0 & 0.1 & 0.5 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \end{pmatrix}, b = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix}, c = \begin{pmatrix} 0 \\ 0.5 \\ 1 \\ -1 \end{pmatrix}$

Options :

1. A
2. B
3. C
4. D

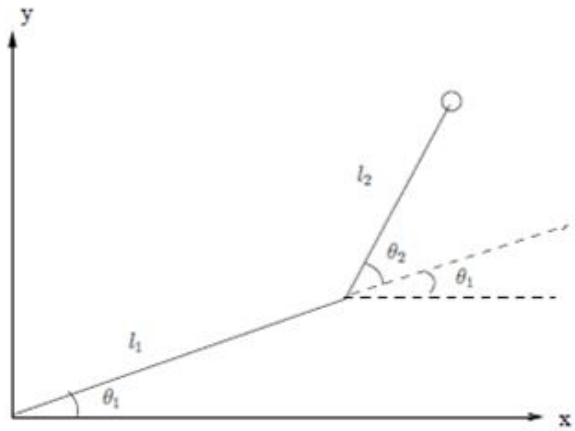
Question Number : 49 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

The forward kinematics of a two-link planar robot is given as

$$x = l_1 \cos\theta_1 + l_2 \cos(\theta_1 + \theta_2)$$
$$y = l_1 \sin\theta_1 + l_2 \sin(\theta_1 + \theta_2)$$

where θ_1, θ_2 are link 1 and 2 angles as shown in the figure.



The corresponding inverse kinematics is learned using a 2-D KSOM based network of a lattice size 10×10. For every KSOM neuron, the associated linear model is:

$$\theta(u) = \theta_\lambda + A_\lambda (u - w_\lambda)$$

where $u = \begin{bmatrix} x \\ y \end{bmatrix}$; w_λ is the cluster center of the λ^{th} neuron. Given $l_1 = l_2 = 1m$.

After training, $u = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -0.5 \\ 1.0 \end{bmatrix}$ is presented to the network to find out the corresponding joint angle vector $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$. The winning neuron cluster center for this input is $w_{\lambda_w} = \begin{bmatrix} -0.42 \\ 0.92 \end{bmatrix}$. Ideal values for $\theta_{\lambda_w}; A_{\lambda_w}$ associated with this winning neuron are

a. $\theta_{\lambda_w} = \binom{0.10}{2.08}; A = \begin{pmatrix} -0.92 & -0.82 \\ 0.42 & -0.58 \end{pmatrix}$

b. $\theta_{\lambda_w} = \binom{-2.18}{2.08}; A = \begin{pmatrix} 0.92 & 0.10 \\ 0.42 & 0.99 \end{pmatrix}$

c. $\theta_{\lambda_w} = \binom{0.56}{0.82}; A = \begin{pmatrix} -0.82 & -0.90 \\ 0.56 & -0.44 \end{pmatrix}$

d. $\theta_{\lambda_w} = \binom{0.96}{2.08}; A = \begin{pmatrix} -0.92 & -0.10 \\ -0.42 & -0.99 \end{pmatrix}$

Options :

1. A
2. B
3. C
4. D

Question Number : 50 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

A mobile robot with a single Sonar while avoiding an obstacle recorded following data. Please note that the robot either goes straight or turns based on the sonar readings which are estimates of the distance of the obstacle up-front. When both velocities are equal, the robot moves straight while unequal velocities lead to left or right turn.

Sonar S	Left Wheel Velocity v_l	Right Wheel Velocity v_r
40cm	8cm/sec	8cm/sec
20cm	4cm/sec	0cm/sec

A radial basis function network is designed that can mimic this behavior where sonar data is input and velocities are output. Take two centers as 0.4 m, and 0.2 m. The radial function is given as $e^{-x^2/2\sigma^2}$ where $\sigma = 0.2$ for both the centers. There is a bias at the output node. The parameters w_1, w_2 and bias θ for predicting left wheel velocity are

a. $w_1 = 11.56, w_2 = -8.76$ and bias $\theta = 1.75$
b. $w_1 = 8.2, w_2 = -1.97$ and bias $\theta = 1.0$
c. $w_1 = 7.18, w_2 = -2.97$ and bias $\theta = 2.62$
d. $w_1 = 7.57, w_2 = -2.59$ and bias $\theta = 2.0$

Options :

1. A
2. B
3. C
4. D

Question Number : 51 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

The weight update law for parameters in KSOM based kinematic control of a robot arm is given as:

$$w_\gamma \leftarrow w_\gamma + \eta_w h_\gamma (x - w_\gamma)$$
$$\theta_\gamma \leftarrow \theta_\gamma + \eta_\theta \Delta\theta_\gamma$$
$$A_\gamma \leftarrow A_\gamma + \eta_A \Delta A_\gamma$$

Here $\Delta\theta_\gamma$ and ΔA_γ are

a. $\Delta\theta_\gamma = \frac{h_\gamma}{s}\left[\theta_0^{out} - s^{-1}\sum_\gamma h_\gamma(\theta_\gamma - A_\gamma(v_0 + w_\gamma))\right]$,

$$\Delta A_\gamma = \frac{h_\gamma}{s \parallel \Delta v \parallel^2}\left[\Delta\theta^{out} - s\sum_\gamma h_\gamma A_\gamma \Delta v\right]\Delta v^T$$

b. $\Delta\theta_\gamma = \frac{h_\gamma}{s}\left[\theta_0^{out} - s\sum_\gamma h_\gamma(\theta_\gamma + A_\gamma(v_0 + w_\gamma))\right]$,

$$\Delta A_\gamma = \frac{h_\gamma}{s \parallel \Delta v \parallel^2}\left[\Delta\theta^{out} - s\sum_\gamma h_\gamma A_\gamma \Delta v\right]\Delta v^T$$

c. $\Delta\theta_\gamma = \frac{h_\gamma}{s}\left[\theta_0^{out} - s^{-1}\sum_\gamma h_\gamma(\theta_\gamma + A_\gamma(v_0 - w_\gamma))\right]$,

$$\Delta A_\gamma = \frac{h_\gamma}{s \parallel \Delta v \parallel^2}\left[\Delta\theta^{out} - s^{-1}\sum_\gamma h_\gamma A_\gamma \Delta v\right]\Delta v^T$$

d. $\Delta\theta_\gamma = \frac{h_\gamma}{s}\left[\theta_0^{out} - s\sum_\gamma h_\gamma(\theta_\gamma + A_\gamma(v_0 - w_\gamma))\right]$,

$$\Delta A_\gamma = \frac{h_\gamma}{s \parallel \Delta v \parallel^2}\left[\Delta\theta^{out} - s\sum_\gamma h_\gamma A_\gamma \Delta v\right]\Delta v^T$$

Options :

1. A
2. B
3. C
4. D

Question Number : 52 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

Consider the following recurrent neural network: y(t+1)=-0.5*y(t)+1*y(t-1)+0.5*u(t); with y(0) = 0.3, y(1) = 0.1, u(1) = 1, u(2) = -1. Take the adaptive learning rate as $\eta = \frac{0.2}{||g||^2}$where $||g||^2 = (\frac{\partial y}{\partial w_1})^2 + (\frac{\partial y}{\partial w_2})^2 + (\frac{\partial y}{\partial w_3})^2$. Taking initial weights and gradients to be zero, the updated weights after two iterations using RTRL algorithm are:

a. $w_1 = 0.544, w_2 = 0.86, w_3 = 0.108$
b. $w_1 = 0.0224, w_2 = 0.0505, w_3 = 0.39$
c. $w_1 = -0.54, w_2 = 0.826, w_3 = 0.001$
d. $w_1 = 0.024, w_2 = 0.562, w_3 = -0.211$

Options :

1. A
2. B
3. C
4. D

Question Number : 53 Question Type : MCQ Option Shuffling : No Display Question Number : Yes Single Line Question Option : No Option Orientation : Vertical

Correct Marks : 4 Wrong Marks : 0

The confusion matrix for evaluation on iris dataset is given by

		Predicted		
		Setosa	**Virginica**	**Versicolor**
Actual	**Setosa**	6	1	1
	Virginica	1	5	1
	Versicolor	1	2	3

The F-measure of the class Virginica is:

a. 5/7
b. 5/8
c. 5/21
d. 2/3

Options :

1. A
2. B
3. C
4. D