

X



(<https://swayam.gov.in>)



([https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL))

bhandarishivay2001@gmail.com ▾

[NPTEL \(<https://swayam.gov.in/explorer?ncCode=NPTEL>\)](https://swayam.gov.in/explorer?ncCode=NPTEL) » Deep Learning - IIT Ropar (course)

≡

Course outline

How does an  
NPTEL online  
course work?  
( )

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

# Week 1: Assignment 1

The due date for submitting this assignment has passed.

Due on 2022-08-10, 23:59 IST.

Assignment submitted on 2022-08-07, 13:32 IST

- 1) Pick out the appropriate shape of decision boundary if the number of inputs is three. **1 point**

- Point
- Line
- Plane
- Hyperplane

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Plane*

- 2) Pick out the one in biological neuron that is responsible for receiving signal from other **1 point** neurons.

- Dendrite
- Synapse
- Soma
- Axon

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Dendrite*

[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)

3) Which of the following is considered as a drawback of Deep Learning?

**1 point**

- Numerical stability
- Overfitting never occurs
- Sharp minima
- Overfitting always occurs

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Sharp minima*

4) Neurons play a vital role in how humans respond to the outside world. When does this **1 point** occur?

- Any one neuron gets activated
- All the neurons of massively parallel interconnected network of neurons are activated.
- Specific set of these neurons fire and relay the information to other neurons
- At least 10% of the total number of neurons in the brain

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Specific set of these neurons fire and relay the information to other neurons*

5) Consider a Mc Culloch Pitts Neuron for which the inputs are  $x_1, x_2$  and  $x_3$ . Also, the **1 point** aggregate function  $g(x)$  is an OR function. What is the thresholding parameter for the same?

- 0
- 1
- 2
- 3

Yes, the answer is correct.

Score: 1

Accepted Answers:

1

6) Which of the following statements are True? **1 point**

Statement I. Mc. Culloch Pitts neuron can be used to represent any boolean function

Statement II. If any of the inputs in a Mc. Culloch Pitts Neuron is inhibitory, then output will be zero

- Only I
- Only II
- Both
- None

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Only II*

7) Pick out the boolean function that is not linearly separable.

**1 point**

- AND
- OR
- NOR
- XOR

Yes, the answer is correct.

Score: 1

Accepted Answers:

XOR

8) In a perceptron learning algorithm, what is the initial value of the weights before the algorithm starts learning?

**1 point**

- All weights set to zero
- All weights set to one
- All weights assigned random values
- All weights assigned values specific to the application in hand

Yes, the answer is correct.

Score: 1

Accepted Answers:

*All weights assigned random values*

9) What is the condition for convergence of a perceptron learning algorithm?

**1 point**

- Always converges
- Data is linearly separable
- Data is linearly non-separable
- May or may not converge depending on the data

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Data is linearly separable*

10) Select all the statements that hold TRUE for a Single Perceptron.

**1 point**

- Inputs are weighted
- Threshold is hand coded
- Only Real inputs are allowed
- Both Real and boolean inputs are allowed
- Can solve only linearly separable data

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Inputs are weighted*

*Both Real and boolean inputs are allowed*

*Can solve only linearly separable data*

X



(https://swayam.gov.in)



(https://swayam.gov.in/nc\_details/NPTEL)

bhandarishivay2001@gmail.com ▾

[NPTEL \(https://swayam.gov.in/explorer?ncCode=NPTEL\)](https://swayam.gov.in/explorer?ncCode=NPTEL) » Deep Learning - IIT Ropar (course)

≡

Course  
outline

How does an  
NPTEL online  
course work?  
( )

**Week 0** ( )

**Week 1** ( )

**Week 2** ( )

**Week 3** ( )

**week 4** ( )

**Week 5** ( )

**Week 6** ( )

**Week 7** ( )

**Week 8** ( )

**Week 9** ( )

**week 10** ( )

## Week 2: Assignment 2

The due date for submitting this assignment has passed.

**Due on 2022-08-10, 23:59 IST.**

**Assignment submitted on 2022-08-10, 15:02 IST**

1) How many Boolean functions can be designed with 3 inputs? **1 point**

- $2^3$
- $2^2$
- $2^{2^3}$
- $2^{2^2}$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$2^{2^3}$

2) Pick out the function(s) that are not linearly separable? **1 point**

- XOR
- NOT
- NOR
- !XOR

Yes, the answer is correct.

Score: 1

Accepted Answers:

XOR

!XOR

[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)

3) Out of the functions that can be designed from  $n$  inputs, how many of them are linearly **1 point** separable?

- $2^{2^n}$
- $2^{2n}$
- $2^n$
- unknown

Yes, the answer is correct.

Score: 1

Accepted Answers:

*unknown*

4) Which of the following statements are TRUE? **1 point**

Statement I. The given network of perceptrons can be used to implement any complex boolean input functions.

Statement II. Each  $W_i$  can be adjusted to get desired output for that input.

- Only I
- Only II
- Both
- None

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Both*

5) Consider you are given a Boolean function with 5 inputs. It is represented by a network **1 point** of perceptrons containing one hidden layer and one output layer with one perceptron. How many perceptrons are there in the hidden layer?

- $2^5$
- $5^2$
- $2^{2^5}$
- $2^{5^2}$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$2^5$

6) Assume you have a perceptron to solve a problem of deciding if a student is eligible for **1 point** scholarship or not. We have only one input in this case. Bias being 50%. What will be the decision of the model when the student scored 0.49 and 0.51?

- eligible, eligible
- eligible, ineligible
- ineligible, eligible

ineligible, ineligible

Yes, the answer is correct.

Score: 1

Accepted Answers:

*ineligible, eligible*

7) State True or False.

**1 point**

I. Logistic function is smooth and continuous. II. Logistic function is differentiable.

I is True and II is False

I is False and II is True

Both are True

Both are False

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Both are True*

8) Select all that applies to a learning algorithm.

**1 point**

Aims to find the value for parameter

Maximize the objective function

Aims to find all possible values for the input x

Minimize the objective function

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Aims to find the value for parameter*

*Minimize the objective function*

9) Sum of squared error is better than sum of errors. Why is this true?

**1 point**

Differential is one for SSE

Positive & Negative do not cancel in SSE

Error is magnified by squaring in SSE

Sum of errors might lead to negative values

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Positive & Negative do not cancel in SSE*

10) Consider a machine learning model, with only one input  $x$  and output  $y$ . Given training instances,  $(x, y) = (0.4, 0.3), (1.8, 0.6)$ ,  $w = 1.2$ ,  $b = -1.4$  and the function is logistic sigmoid function. Compute the loss function,  $L(w, b) = \frac{1}{2} \sum_{i=1}^N (y_i - f(x_i))^2$

0.0059

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Range) 0.003,0.004

**1 point**

X



(<https://swayam.gov.in>)



([https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL))

bhandarishivay2001@gmail.com ▾

**NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)**

≡

Course outline

How does an  
NPTEL online  
course work?  
( )

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

## Week 3: Assignment 3

The due date for submitting this assignment has passed.

**Due on 2022-08-17, 23:59 IST.**

**Assignment submitted on 2022-08-17, 20:04 IST**

1) Assume you are developing a model to predict the probability as an output. Pick out the **1 point** appropriate Activation function.

- linear
- sigmoid
- tanh
- Relu

Yes, the answer is correct.

Score: 1

Accepted Answers:

*sigmoid*

2) The pre-activation at layer  $i$  can be best described as the

**1 point**

- weighted sum of all the inputs at layer  $i$
- sum of all the the inputs at layer  $i$
- weighted sum of all the inputs at layer  $i + 1$
- sum of all the inputs at layer  $i + 1$
- weighted sum of all the inputs at layer  $i - 1$

[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)sum of all the inputs at layer  $i - 1$ 

No, the answer is incorrect.

Score: 0

Accepted Answers:

*weighted sum of all the inputs at layer  $i - 1$* 

- 3) Consider a Machine Learning model that is applied to a specific set of inputs. Actual output being  $y_i = [10, 5, 7, 8, 6]$  and the predicted output being  $\hat{y}_i = [9, 6, 5, 7, 5]$ , Compute Mean Squared error loss.

1.60

Yes, the answer is correct.

Score: 1

Accepted Answers:

(Type: Numeric) 1.6

**1 point**

- 4) Consider a Classification problem with k classes. The output being a probability distribution, which of the following is the best output function?

- Linear
- Sigmoid
- tanh
- softmax

Yes, the answer is correct.

Score: 1

Accepted Answers:

softmax

**1 point**

- 5) Given the output  $y_j = O(a_l)_j$  and  $a_l = [2.5, 3.6, 4.2, 5]$ . If 'O' is the softmax function, compute the value of  $\hat{y} = [\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4]$ ?

- [0.046, 0.139, 0.253, 0.562]
- [0.046, 0.253, 0.562, 0.139]
- [0.253, 0.046, 0.139, 0.562]
- [0.562, 0.046, 0.139, 0.253]

Yes, the answer is correct.

Score: 1

Accepted Answers:

[0.046, 0.139, 0.253, 0.562]

**1 point**

- 6) The information content is high for an event when the probability of the event is

**1 point**

- high
- low
- 1
- maximum

Yes, the answer is correct.

Score: 1

Accepted Answers:

*low*

- 7) Assume you have four inputs to a Feed Forward neural network, the first hidden layer **1 point** also has four neurons, and there are three output classes, what is the dimension of the weight matrix,  $W_1$  between the input layer and the first hidden layer, given that there is only one hidden layer?

- $\mathbb{R}^{3 \times 3}$
- $\mathbb{R}^{4 \times 3}$
- $\mathbb{R}^{4 \times 4}$
- $\mathbb{R}^{3 \times 4}$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$\mathbb{R}^{4 \times 4}$

- 8) In a Feed Forward Neural Network, if the outputs take real values then which of the **1 point** following output activation function and error function do you prefer?

- Linear, cross entropy
- Softmax, cross entropy
- Linear, Squared error
- Softmax, Squared error

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Linear, Squared error*

- 9) The activation layer at any layer  $i$  is given by **1 point**

- $h_i(x) = b_i + W_i h_{i-1}(x)$
- $h_i(x) = g(a_i(x))$
- $h_i(x) = O(aL)$
- $h_i(x) = a_i + W_i h_{i-1}(x)$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$h_i(x) = g(a_i(x))$

- 10) Identify the loss function for a classification problem to choose one out of K Classes. **1 point**

- Squared error

Absolute error

Minimize  $\underset{\theta}{\mathcal{L}}(\theta) = -\log(\hat{y}_l)$

Maximize  $\underset{\theta}{\mathcal{L}}(\theta) = -\log(\hat{y}_l)$

Yes, the answer is correct.

Score: 1

Accepted Answers:

Minimize  $\underset{\theta}{\mathcal{L}}(\theta) = -\log(\hat{y}_l)$

X



(<https://swayam.gov.in>)



([https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL))

bhandarishivay2001@gmail.com ▾

[NPTEL](https://swayam.gov.in/explorer?ncCode=NPTEL) (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)

≡

Course outline

How does an  
NPTEL online  
course work?  
( )

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

## Week 4: Assignment 4

The due date for submitting this assignment has passed.

Due on 2022-08-24, 23:59 IST.

Assignment submitted on 2022-08-24, 21:59 IST

1) Consider the movement on the 3D error surface for Vanilla Gradient Descent Algorithm. Select all the options that are TRUE. 1 point

- Smaller the gradient, slower the movement
- Larger the gradient, faster the movement
- Gentle the slope, smaller the gradient
- Steeper the slope, smaller the gradient

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Smaller the gradient, slower the movement*

*Larger the gradient, faster the movement*

*Gentle the slope, smaller the gradient*

2) Pick out the drawback in Vanilla gradient descent algorithm. 1 point

- Very slow movement on gentle slopes
- Increased oscillations before converging
- escapes minima because of long strides
- Very slow movement on steep slopes

No, the answer is incorrect.

Score: 0

Accepted Answers:

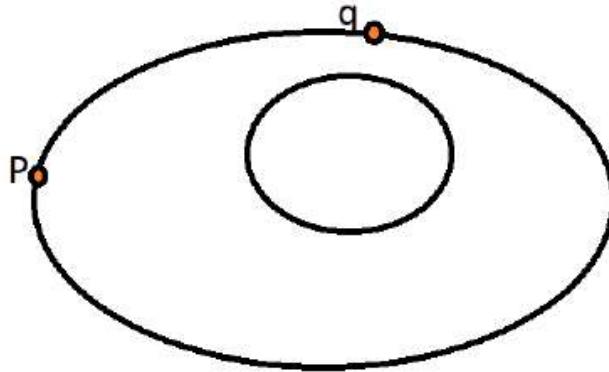
[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)*Very slow movement on gentle slopes*3) Comment on the update at the  $t^{\text{th}}$  update in the Momentum-based Gradient Descent. **1 point**

- weighted average of gradient
- Polynomial weighted average
- Exponential weighted average of gradient
- Average of recent three gradients

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Exponential weighted average of gradient*4) Given a horizontal slice of the error surface as shown in the figure below, if the error at **1 point** the position p is 0.49 then what is the error at point q?

- 0.70
- 0.69
- 0.49
- 0

Yes, the answer is correct.

Score: 1

Accepted Answers:

0.49

5) Identify the update rule for Nesterov Accelerated Gradient Descent. **1 point**

$$w_{t+1} = w_t - \eta \nabla w_t$$

$$b_{t+1} = b_t - \eta \nabla b_t$$



$$update_t = \gamma \cdot update_{t-1} + \eta \nabla w_t$$

$$w_{t+1} = w_t - update_t$$



$$w_{look\_ahead} = w_t - \gamma \cdot update_{t-1}$$

$$update_t = \gamma \cdot update_{t-1} + \eta \nabla w_{look\_ahead}$$

$$w_{t+1} = w_t - update_t$$



$$v_t = v_{t-1} + (\nabla w_t)^2$$

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{v_t + \epsilon}} * \nabla w_t$$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$$w_{look\_ahead} = w_t - \gamma \cdot update_{t-1}$$

$$update_t = \gamma \cdot update_{t-1} + \eta \nabla w_{look\_ahead}$$

$$w_{t+1} = w_t - update_t$$

- 6) Select all the options that are TRUE for Line search. **1 point**

- w is updated using different learning rates
- updated value of w always gives the minimum loss
- Involves minimum calculation
- Best value of Learning rate is used at every step

Yes, the answer is correct.

Score: 1

Accepted Answers:

w is updated using different learning rates

updated value of w always gives the minimum loss

Best value of Learning rate is used at every step

- 7) Assume you have 1,50,000 data points, Mini batch size being 25,000, one epoch **1 point**  
implies one pass over the data, and one step means one update of the parameters, What is the  
number of steps in one epoch for Mini-Batch Gradient Descent?

- 1

- 1,50,000
- 6
- 60

Yes, the answer is correct.

Score: 1

Accepted Answers:

6

8) Which of the following learning rate methods need to tune two hyperparameters?

**1 point**

- I. step decay
- II. exponential decay
- III.  $1/t$  decay

- I and II
- II and III
- I and III
- I, II and III

Yes, the answer is correct.

Score: 1

Accepted Answers:

*II and III*

9) How can you reduce the oscillations and improve the stochastic estimates of the gradient that is estimated from one data point at a time?

**1 point**

- Mini-Batch
- Adam
- RMSprop
- Adagrad

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Mini-Batch*

10) Select all the statements that are TRUE.

**1 point**

- RMSprop is very aggressive when decaying the learning rate
- Adagrad decays the learning rate in proportion to the update history
- In Adagrad, frequent parameters will receive very large updates because of the decayed learning rate
- RMSprop has overcome the problem of Adagrad getting stuck when close to convergence

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Adagrad decays the learning rate in proportion to the update history*

*RMSprop has overcome the problem of Adagrad getting stuck when close to convergence*

X



(https://swayam.gov.in)



(https://swayam.gov.in/nc\_details/NPTEL)

bhandarishivay2001@gmail.com ▾

**NPTEL (https://swayam.gov.in/explorer?ncCode=NPTEL) » Deep Learning - IIT Ropar (course)**

≡

**Course outline**

How does an  
NPTEL online  
course work?  
( )

**Week 0 ()****Week 1 ()****Week 2 ()****Week 3 ()****week 4 ()****Week 5 ()****Week 6 ()****Week 7 ()****Week 8 ()****Week 9 ()****week 10 ()**

# Week 5: Assignment 5

**The due date for submitting this assignment has passed.****Due on 2022-08-31, 23:59 IST.****Assignment submitted on 2022-08-31, 15:14 IST**

- 1) What is the largest dominant eigen value for a stochastic matrix? **1 point**

- 1
- 0
- 1
- 2

Yes, the answer is correct.

Score: 1

Accepted Answers:

1

- 2) Consider  $\lambda_d$  to be the dominant eigen value of a matrix and  $\lambda_d > 1$ , comment on the **1 point** given sequence:  $x_0, Ax_0, A^2x_0, \dots$

- will vanish
- will explode
- will reach its maximum
- will reach a steady state

Yes, the answer is correct.

Score: 1

Accepted Answers:

*will explode*

[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)

Find the Eigen values a & b of the given matrix.

$$A = \begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}$$

3) a = \_\_\_\_\_ (Enter the minimum eigen value)

No, the answer is incorrect.  
Score: 0

Accepted Answers:  
(Type: Numeric) 1

**1 point**

4) b = \_\_\_\_\_ (Enter the maximum eigen value)

No, the answer is incorrect.  
Score: 0

Accepted Answers:  
(Type: Numeric) 6

**1 point**

5) The eigen vectors of a matrix having distinct eigen values are

- orthogonal
- stochastic
- scalar
- linearly independent

Yes, the answer is correct.  
Score: 1

Accepted Answers:  
*linearly independent*

6) Given the vectors,  $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

**1 point**

check if they form a basis for  $\mathbb{R}^3$ ?

- Yes
- No

Yes, the answer is correct.  
Score: 1

Accepted Answers:  
Yes

7) Comment on the number of Eigen vectors of a  $n \times n$  matrix.

**1 point**



Less than or equal to  $n^2$



Less than or equal to  $n$



At least  $n$



At least  $n/2$

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Less than or equal to n*

8) Principal Component Analysis helps in representing data using fewer dimensions. **1 point**

Which of the following is TRUE for the selected dimensions?

- data has low variance along these dimensions
- dimensions are linearly independent
- dimensions are linearly dependent
- dimensions are orthogonal

Yes, the answer is correct.

Score: 1

Accepted Answers:

*dimensions are linearly independent*

*dimensions are orthogonal*

9) Pick out the one that best describes the given matrix where a and b are probabilities of **1 point** two independent events,

$$\begin{bmatrix} a & (1 - b) \\ (1 - a) & b \end{bmatrix}$$

- row stochastic
- column stochastic
- doubly stochastic
- Identity

Yes, the answer is correct.

Score: 1

Accepted Answers:

*column stochastic*

10) Consider the set of features in a Cancer dataset: Name, DateofBirth, age, contact **1 point** number, city, Weight in Kg, Height in cm, Height in inches etc. Identify the attribute(s) that can be excluded in the representation of the dataset. Select all that apply.

- Height in cm or Height in inches
- Weight in Kg or Height in inches
- Date of Birth or Age
- Age or Height in cms

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Height in cm or Height in inches*

*Date of Birth or Age*

11) Which of the following statements are TRUE for Eigen vectors of a square matrix? **1 point**

Statement I. Eigen vectors having distinct eigen values are linearly dependent Statement II. Eigen vectors of a square matrix are orthogonal

- Only I
- Only II
- Both
- None

No, the answer is incorrect.

Score: 0

Accepted Answers:

*None*

X



(<https://swayam.gov.in>)



([https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL))

bhandarishivay2001@gmail.com ▾

**NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)**

≡

Course outline

How does an  
NPTEL online  
course work?  
( )

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

# Week 6: Assignment 6

The due date for submitting this assignment has passed.

**Due on 2022-09-07, 23:59 IST.**

**Assignment submitted on 2022-09-07, 22:14 IST**

1) Identify the type of Autoencoder, that limits the amount of information flowing through **1 point** the network by reducing the number of nodes in hidden layer to learn the most important attributes of the input with minimum reconstruction error from the encoded state.

- Sparse Autoencoder
- Overcomplete Autoencoder
- Undercomplete Autoencoder
- Contractive Autoencoder

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Undercomplete Autoencoder*

2) Consider an autoencoder where the inputs are binary and outputs are also binary. **1 point** Identify the appropriate function for the encoder.

- tanh
- linear
- Binary step
- sigmoid

Yes, the answer is correct.

Score: 1

Accepted Answers:

[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)*sigmoid*

- 3) For an autoencoder where the inputs are real and the interpretation is performed in terms of real values. Pick out the most appropriate loss function?

- Squared error
- Absolute error
- Cross entropy
- None of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Squared error*

- 4) Select all the characteristics that ensure that an autoencoder is equivalent to PCA. **1 point**

- use linear encoder
- use linear decoder
- use squared error loss function
- use real value input

Yes, the answer is correct.

Score: 1

Accepted Answers:

*use linear encoder*

*use linear decoder*

*use squared error loss function*

- 5) You have a model that learns a very complex or flexible model that yields zero error on **1 point** training data. Select the options that are possible in this case.

- Overfitting has occurred
- bad performance on unseen data
- Underfitting has occurred
- Will perform good on Unseen data

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Overfitting has occurred*

*bad performance on unseen data*

- 6) Which of the following is True? **1 point**

Statement I. An overcomplete autoencoder has a large number of parameters which in turn leads to overfitting.

Statement II. An undercomplete Autoencoder does not require regularization as it shrinks the number of parameters

- I only
- II only

- Both
- None

Yes, the answer is correct.  
Score: 1

Accepted Answers:  
*I only*

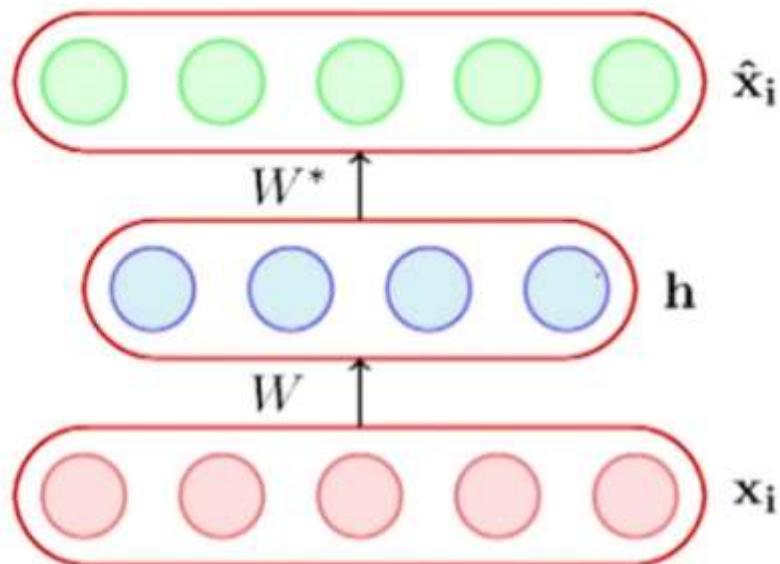
7) Which of the following Autoencoders purposefully corrupts randomly picked input data **1 point** to improve the performance?

- Sparse
- Contractive
- Denoising
- Overcomplete
- Undercomplete

Yes, the answer is correct.  
Score: 1

Accepted Answers:  
*Denoising*

8) For the given Autoencoder, observe the encoder function. The dimension of input **1 point** vector,  $x_i$  is  $n \times 1$  and dimension of bias is  $d \times 1$ . What is the dimension of weight  $w$  given that the hidden layer has a dimension  $d$ ?



$$\mathbf{h} = g(\mathbf{Wx}_i + \mathbf{b})$$



$d \times d$

$d \times n$

$n \times d$

$n \times 1$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$d \times n$

9) Which of the following statements are True? **1 point**

S1: L1-regularization enforces sparsity on model weights leading to implicit feature selection. S2:

L2-regularization prevents most weights from taking large values.

I only

II only

Both

None

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Both*

10) Which of the following statements are TRUE for Sparse Autoencoders? **1 point**

Statement I. Tries to ensure that the neuron is inactive most of the time.

Statement II. Uses a probabilistic process to corrupt the input data.

Only I

Only II

Both

None

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Only I*

X



(<https://swayam.gov.in>)



([https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL))

bhandarishivay2001@gmail.com ▾

**NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)**

≡

Course outline

How does an  
NPTEL online  
course work?  
( )

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

## Week 7: Assignment 7

The due date for submitting this assignment has passed.

**Due on 2022-09-14, 23:59 IST.**

**Assignment submitted on 2022-09-14, 23:36 IST**

1) Bias enables shifting the activation function by adding a constant to the input. How is it **1 point** related to model complexity? Select all that apply.

- Complex model has a low bias
- Simple model has a low bias
- Complex model has a high bias
- Simple model has a high bias

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Complex model has a low bias*

*Simple model has a high bias*

2) What are the preferred values for Bias and Variance to yield a low Mean Squared **1 point** Error?

- Low bias and high variance
- Low variance and high bias
- Low bias and low variance
- Optimal value for both bias and variance

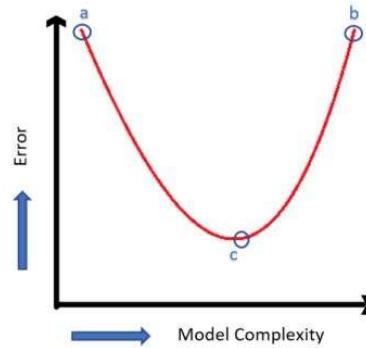
Yes, the answer is correct.

Score: 1

Accepted Answers:

[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)*Optimal value for both bias and variance*

- 3) Given the relation between Validation Error and Model complexity for a Deep Neural Network to classify dogs and cats, Which of the points correspond to underfitting? **1 point**



- a
- b
- c
- both a and b

Yes, the answer is correct.  
Score: 1

Accepted Answers:  
a

- 4) Consider the Deep Neural Network as described in Question 3, Which of the following **1 point** instances marked on graph marks the preferred optimal complexity?

- a
- b
- c
- both a and b

Yes, the answer is correct.  
Score: 1

Accepted Answers:  
c

- 5) Which of the following statements are True? **1 point**

**Statement I.** Training a model to drive the training error to zero ensures high model performance.  
**Statement II.** Tuning for Validation error ensures that the model does not overfit

- Only I
- Only II
- Both I and II
- None

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*Only II*

6) Pick out the strategies to prevent overfitting.

**1 point**

- Reduce the number of hidden layers
- Decrease the number of samples
- Switch off some of the neurons in the neural network
- Increase the number of hidden layers

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Reduce the number of hidden layers*

*Switch off some of the neurons in the neural network*

7) Which of the following is CORRECT for Bagging?

**1 point**

- Combines predictions from different neural network models on same data
- Used to reduce variance within the dataset
- Leads to Overfitting
- Ensemble on models trained on same classifier using different instances

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Used to reduce variance within the dataset*

*Ensemble on models trained on same classifier using different instances*

8) Suppose there are 20 nodes in a Deep Neural network and we implement Dropout by **1 point** removing few nodes to obtain a thinned network. What is total number of such thinned networks that can be formed?

- 
- 20
- $2^{20}$
- 
- $20^2$
- 
- $2^{2^{20}}$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$2^{20}$

9) Identify the need for regularization. Select all that apply.

**1 point**

- Deep Neural networks are complex
- Involves non-linearities
- High possibility of overfitting
- Involves many Parameters

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Deep Neural networks are complex*

*Involves non-linearities*

*High possibility of overfitting*

*Involves many Parameters*

10) You are training a neural network model using Early stopping technique. Given that the **1 point** patience parameter is 2, When will you stop training?

| Epochs | Training loss | Validation loss |
|--------|---------------|-----------------|
| 1      | 3.4           | 2.0             |
| 2      | 2.0           | 1.9             |
| 3      | 1.9           | 1.8             |
| 4      | 1.8           | 1.8             |
| 5      | 1.7           | 1.9             |
| 6      | 1.6           | 2.1             |

3

4

5

6

No, the answer is incorrect.

Score: 0

Accepted Answers:

5

X

[\(https://swayam.gov.in\)](https://swayam.gov.in)[\(https://swayam.gov.in/nc\\_details/NPTEL\)](https://swayam.gov.in/nc_details/NPTEL)

bhandarishivay2001@gmail.com ▾

[NPTEL \(https://swayam.gov.in/explorer?ncCode=NPTEL\)](https://swayam.gov.in/explorer?ncCode=NPTEL) » Deep Learning - IIT Ropar (course)

≡

**Course outline**

How does an  
NPTEL online  
course work?  
( )

**Week 0** ()

**Week 1** ()

**Week 2** ()

**Week 3** ()

**week 4** ()

**Week 5** ()

**Week 6** ()

**Week 7** ()

**Week 8** ()

**Week 9** ()

**week 10** ()

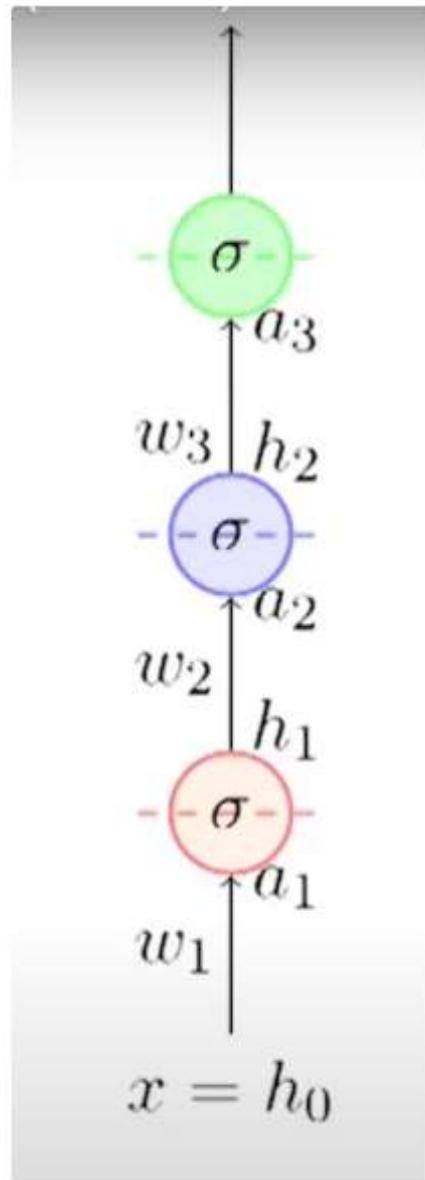
# Week 8: Assignment 8

The due date for submitting this assignment has passed.

**Due on 2022-09-21, 23:59 IST.**

As per our records you have not submitted this assignment.

- 1) Identify the statements that are True for training the following deep neural network. **1 point**

[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)

- Implement Gradient Descent Algorithm
- Apply chain rule to compute gradients
- 
- Gradient from any layer depends only on the input  $x_i$
- Chain rule applied across multiple paths

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Implement Gradient Descent Algorithm*

*Apply chain rule to compute gradients*

2) Which of the following are True for Unsupervised pre-training?

**1 point**

- Pre-training is more robust than random initialization

- Layers capture the internal characteristics of the data
- Ensures Learning is not greedy
- constrains most of the weights to zero

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Pre-training is more robust than random initialization*

*Layers capture the internal characteristics of the data*

*Ensures Learning is not greedy*

3) What would be the gradient when sigmoid neuron saturates? **1 point**

- maximum
- keeps decreasing
- vanishes
- keeps increasing

No, the answer is incorrect.

Score: 0

Accepted Answers:

*vanishes*

4) Identify the problem with Sigmoid neurons. **1 point**

- zero-centered
- computationally expensive
- update directions of gradient is not restricted
- neurons do not saturate

No, the answer is incorrect.

Score: 0

Accepted Answers:

*computationally expensive*

5) Pick out the Activation function that converges much faster. **1 point**

- Linear
- tanh
- sigmoid
- ReLU

No, the answer is incorrect.

Score: 0

Accepted Answers:

*ReLU*

6) One of the disadvantages of ReLU is that it can lead to dead neurons. Select all the conditions that causes more neurons to die. **1 point**

- bias receives a large negative update
- learning rate is high

- too many inputs
- initial weights are not good

No, the answer is incorrect.  
Score: 0

Accepted Answers:

*bias receives a large negative update*

*learning rate is high*

7) Which of the following variants of ReLU is computationally expensive? **1 point**

- Leaky ReLU
- Parametric ReLU
- Exponential Linear Unit
- None

No, the answer is incorrect.  
Score: 0

Accepted Answers:

*Exponential Linear Unit*

8) Identify the cause of Symmetry Breaking problem. **1 point**

- weights in a network initialised to random values.
- weights in a network initialised to same value.
- weights in a network initialised by pre-training.
- weights in a network initialised to unequal values.

No, the answer is incorrect.  
Score: 0

Accepted Answers:

*weights in a network initialised to same value.*

9) Pick out the disadvantage of Maxout neuron. **1 point**

- No saturation
- Gradients vanish
- Doubles the number of parameters
- Not zero-centered

No, the answer is incorrect.  
Score: 0

Accepted Answers:

*Doubles the number of parameters*

10) Consider a Deep neural network with three inputs and five hidden layers  $h_0$  to  $h_4$  with **1 point** three neurons each. The Learning process would be hard if across the mini-batches, the distribution of  $h_3$  keeps changing. Identify a possible solution.

- 
- Pre-activation at layer  $h_0$  were unit gaussian
- Pre-activation at every layer were unit gaussian

- Pre-activation at any layer is not a unit gaussian
- Pre-activation at layer  $h_4$  were unit gaussian

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Pre-activation at every layer were unit gaussian*

X



(<https://swayam.gov.in>)



([https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL))

bhandarishivay2001@gmail.com ▾

**NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)**

≡

Course outline

How does an  
NPTEL online  
course work?  
( )

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

# Week 9: Assignment 9

The due date for submitting this assignment has passed.

**Due on 2022-09-28, 23:59 IST.**

**Assignment submitted on 2022-09-28, 21:02 IST**

- 1) Given the following metrics based on one hot representation of cat, dog and Truck as **1 point** mentioned in the Lecture, select all the options that hold True.

$\text{euclid\_dist}(\text{cat}, \text{dog}) = \sqrt{2}$   
 $\text{euclid\_dist}(\text{dog}, \text{truck}) = \sqrt{2}$   
 $\text{cosine\_sim}(\text{cat}, \text{dog}) = 0$   
 $\text{cosine\_sim}(\text{dog}, \text{truck}) = 0$

- Cosine similarity captures the fact that both cat and dog are domestic animals
- Euclidean distance captures the fact that Truck is different from the domestic animals
- Both measures do not capture the similarity or difference between the words.
- Cosine similarity between any two words in the corpus is always zero

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*Both measures do not capture the similarity or difference between the words.*

*Cosine similarity between any two words in the corpus is always zero*

- 2) What is the maximum size of the co-occurrence matrix, given the number of words in **1 point** corpus being ' $n$ '?

$n \times n$

[Week 11 \(\)](#) $n - 1 \times n - 1$  $n/2 \times n/2$  $2n \times 2n$ 

Yes, the answer is correct.

Score: 1

Accepted Answers:

 $n \times n$ [Download Videos \(\)](#)

3) Select all the drawbacks of distributed representation of words.

**1 point** sparse high dimension size of co-occurrence matrix increases with vocabulary size size of co-occurrence matrix is very high even for a very small vocabulary size[Books \(\)](#)

Yes, the answer is correct.

Score: 1

Accepted Answers:

*sparse**high dimension**size of co-occurrence matrix increases with vocabulary size*[Text Transcripts \(\)](#)

4) Pick out the stop word among the following options.

**1 point** an common data stop

Yes, the answer is correct.

Score: 1

Accepted Answers:

*an*[Live Sessions \(\)](#)5) Given the quantity count  $(w, c) = 0$ , this drives the value of  $PMI(w, c)$  to  $-\infty$ .**1 point**

Which of the following can be a possible solution to this issue?

 Ignore high frequency stop words PPMI - restricts the value of PMI to 0 and above PPMI - restricts the value of PMI to -1 and above Ignore low frequency stop words

Yes, the answer is correct.

Score: 1

Accepted Answers:

*PPMI - restricts the value of PMI to 0 and above*[Problem Solving Session \(\)](#)

6) Which of the following is True for SVD?

**1 point** gives best rank k approximation of given data discovers latent semantics in the corpus

- SVD reduces the number of dimensions  
 all the above

Yes, the answer is correct.  
 Score: 1

Accepted Answers:  
*all the above*

7) In learning word representations using prediction based models, the softmax function is **1 point** computationally expensive. Which of the following are alternate solution for the above problem?

- I. Use negative sampling  
 II. Use contrastive estimation  
 III. Use hierarchical softmax

- I and II  
 II and III  
 I and III  
 I, II and III

Yes, the answer is correct.  
 Score: 1

Accepted Answers:  
*I, II and III*

8) Pick out the model that is NOT Count-based.

**1 point**

- PMI  
 Word count  
 PPMI  
 Continuous bag of words

Yes, the answer is correct.  
 Score: 1

Accepted Answers:  
*Continuous bag of words*

9) Which of the following models directly learn word representations?

**1 point**

- Count-based model  
 Prediction-based model  
 Both  
 None

Yes, the answer is correct.  
 Score: 1

Accepted Answers:  
*Prediction-based model*

10) Given a term and  $k$  different words, select a candidate out of  $k$  which has the largest **1 point** similarity with the term. Which of the following methods follows the above approach in evaluation of word representation?

- Semantic relatedness
- Synonym detection
- Semantic analogy
- Any of the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Synonym detection*

X



(<https://swayam.gov.in>)



([https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL))

bhandarishivay2001@gmail.com ▾

[NPTEL](https://swayam.gov.in/explorer?ncCode=NPTEL) (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)

≡

Course outline

How does an  
NPTEL online  
course work?  
( )

[Week 0](#) ( )

[Week 1](#) ( )

[Week 2](#) ( )

[Week 3](#) ( )

[week 4](#) ( )

[Week 5](#) ( )

[Week 6](#) ( )

[Week 7](#) ( )

[Week 8](#) ( )

[Week 9](#) ( )

[week 10](#) ( )

# Week 10: Assignment 10

The due date for submitting this assignment has passed.

**Due on 2022-10-05, 23:59 IST.**

As per our records you have not submitted this assignment.

- 1) Consider the following: Weight array,  $W = [0.2, 0.7, 0.05, 0.75, 0.86, 0.21]$  and Input  $X = [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$ . What is the next revised estimate  $S_t$  that is obtained by sliding the filter  $W_t$  over the input  $X_t$ ?

**NOTE:** Include two digits after the decimal.

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Range) 1.8,1.9

**1 point**

- 2) Given the following Input matrix and kernel, What is the (0,0) element in the output matrix of the convolution operation if the stride rate is 1?

**1 point**

Input Matrix:

|   |   |   |   |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 2 | 1 | 4 | 3 |
| 1 | 3 | 2 | 4 |
| 4 | 1 | 2 | 3 |



[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)

Kernel Matrix:

|   |   |
|---|---|
| 1 | 2 |
| 2 | 1 |

- 10
- 0
- 4
- 8

No, the answer is incorrect.

Score: 0

Accepted Answers:

10

3) Predict the output image for the convolution operation on the given input image and **1 point** filter.

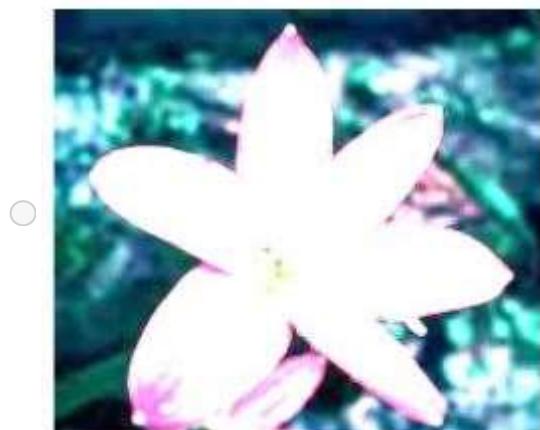
Input:



Kernel:

|   |   |   |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |





No, the answer is incorrect.

Score: 0

Accepted Answers:





4) Identify the kernel that can be used to sharpen an image.

**1 point**

|   |    |   |
|---|----|---|
| 1 | 1  | 1 |
| 1 | -8 | 1 |
| 1 | 1  | 1 |

|    |    |    |
|----|----|----|
| 0  | -1 | 0  |
| -1 | 5  | -1 |
| 0  | -1 | 0  |

|   |   |   |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |

|   |   |   |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 0 | 0 |
| 0 | 0 | 0 |

No, the answer is incorrect.

Score: 0

Accepted Answers:

|    |    |    |
|----|----|----|
| 0  | -1 | 0  |
| -1 | 5  | -1 |
| 0  | -1 | 0  |

5) Consider a 2D filter convolution over a 3D input. What would be the dimension of the output?

**1 point**

one



- two
- three
- point

No, the answer is incorrect.

Score: 0

Accepted Answers:

*three*

- 6) Given the input image with dimensions  $125 \times 49$ , filter of size  $5 \times 5$ , padding  $P = 2$  **1 point** and stride  $S = 2$ , what is the dimension of the output image of this convolution operation?

- $62 \times 24$
- $63 \times 25$
- $64 \times 26$
- None

No, the answer is incorrect.

Score: 0

Accepted Answers:

*$63 \times 25$*

- 7) Pick out the technique(s) that involve static feature extraction using handcrafted kernels to learn the weights of the classifier. **1 point**

- SIFT
- HOG
- Both SIFT and HOG
- Neither SIFT nor HOG

No, the answer is incorrect.

Score: 0

Accepted Answers:

*Both SIFT and HOG*

- 8) Given the stride rate 1, compute the output for the Average pooling operation on the given input matrix. **1 point**

|   |   |   |   |
|---|---|---|---|
| 3 | 7 | 6 | 4 |
| 2 | 6 | 4 | 8 |
| 1 | 8 | 5 | 2 |
| 1 | 9 | 7 | 5 |

|      |      |      |
|------|------|------|
| 4.5  | 5.75 | 5.5  |
| 4.25 | 5.75 | 4.75 |
| 4.75 | 7.25 | 4.75 |



|                       |      |      |      |
|-----------------------|------|------|------|
| <input type="radio"/> | 4.5  | 5.75 | 4.75 |
| <input type="radio"/> | 4.25 | 5.75 | 5.5  |
| <input type="radio"/> | 4.75 | 4.75 | 4.75 |

|                       |     |      |
|-----------------------|-----|------|
| <input type="radio"/> | 4.5 | 4.75 |
| <input type="radio"/> | 5.5 | 4.75 |

|                       |      |      |
|-----------------------|------|------|
| <input type="radio"/> | 4.5  | 4.25 |
| <input type="radio"/> | 4.75 | 5.75 |

No, the answer is incorrect.

Score: 0

Accepted Answers:

|      |      |      |
|------|------|------|
| 4.5  | 5.75 | 5.5  |
| 4.25 | 5.75 | 4.75 |
| 4.75 | 7.25 | 4.75 |

9) What is the number of parameters to be computed for the convolution operation given **1 point** the following notations?

Input image dimension =  $a \times a \times d$

Number of filters =  $n$

Filter dimension =  $f \times f$

Stride =  $s$

Padding =  $p$

$(f \times f \times d) \times n$

$(n \times n \times d) \times f$

$(n \times f \times n) \times d$

$(d \times d \times f) \times n$

No, the answer is incorrect.

Score: 0

Accepted Answers:

$(f \times f \times d) \times n$

10) Which of the following is True for Deep Dream Images? **1 point**

A single neuron fires more

A single filter fires more

A single filter does not fire

A single neuron does not fire

No, the answer is incorrect.

Score: 0

Accepted Answers:



*A single neuron fires more*



X



(https://swayam.gov.in)



(https://swayam.gov.in/nc\_details/NPTEL)

bhandarishivay2001@gmail.com ▾

[NPTEL \(https://swayam.gov.in/explorer?ncCode=NPTEL\)](https://swayam.gov.in/explorer?ncCode=NPTEL) » Deep Learning - IIT Ropar (course)

≡

Course outline

How does an  
NPTEL online  
course work?  
( )

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

# Week 11: Assignment 11

The due date for submitting this assignment has passed.

Due on 2022-10-12, 23:59 IST.

Assignment submitted on 2022-10-12, 00:34 IST

1) Identify the model that accepts input of different dimensions and the inputs that are dependent on each other in the process of Learning. **1 point**

- FeedForward Neural Network
- Convolutional Neural Network
- Recurrent Neural Network
- None

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Recurrent Neural Network*

2) Select all the characteristics that should be possessed by a Recurrent Neural Network. **1 point**

- Should account for the dependence of inputs
- Should account for variable number of inputs
- Should ensure that each timestep implements a suitable function
- Should account for same size of inputs

No, the answer is incorrect.

Score: 0

Accepted Answers:



[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)

*Should account for the dependence of inputs  
Should account for variable number of inputs*

3) Pick out the application that does not belong to the sequence learning approach. **1 point**

- Translate a single input language to several languages
- Speech recognition
- Image captioning
- Object detection
- Create classical Music
- Stock market projections

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Object detection*

4) Which of the statements is TRUE for Backpropagation Through Time(BPTT)? **1 point**

- I. Unlike backprop, in BPTT we sum up gradients for corresponding weight for each time step
- II. Unlike backprop, in BPTT we find sum or product of gradients for corresponding weight at each time step

- I only is True
- II only is True
- Both I and II are True
- Both I and II are False

No, the answer is incorrect.

Score: 0

Accepted Answers:

*I only is True*

5) Choose the appropriate solution for the “Exploding Gradient” problem in RNNs? **1 point**

- Dropout
- Gradient Clipping
- Batch normalization
- RELU

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Gradient Clipping*

6) Which of the following is true about Gated Recurrent units? **1 point**

- I. Explicit Forget gate
- II. Gate directly depends on  $s_{t-1}$
- III. Gate depends on the intermediate  $h_{t-1}$

- I only
- II only



- III only
- All the above

Yes, the answer is correct.

Score: 1

Accepted Answers:

*II only*

7) LSTM equations involve the computation of gates and states for selective read, **1 point** selective write and selective forget. What is the total number of computations at every timestep in LSTM?

- 3
- 4
- 5
- 6

Yes, the answer is correct.

Score: 1

Accepted Answers:

6

8) Which of the following statements is(are)True for LSTM? **1 point**

- I. The flow of information and gradients is controlled by Gates
- II. The gradients vanish only when they actually should vanish.

- Both are True
- I is False hence II is False
- I is True but II is False
- I is False but II is True

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Both are True*

9) Identify the correct equation for the output gate of LSTM network. Given  $U$  and  $W$  are **1 point** weight matrices,  $b$  is the bias,  $t$  denotes timestep,  $x$  denotes input and  $h$  denotes output from previous cell.

- $O_t = \sigma(W_0 h_t + U_0 x_t + b_0)$
- $O_t = \sigma(W_0 h_{t-1} + U_0 x_t + b_0)$
- $O_t = \sigma(W_0 h_{t-1} + U_0 x_{t-1} + b_0)$
- $O_t = \sigma(W_0 h_t + U_0 s_t + b_0)$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$O_t = \sigma(W_0 h_{t-1} + U_0 x_t + b_0)$



**1 point**

10) Which operation does the given set of equations represent?

$$O_{t-1} = \sigma(W_0 h_{t-2} + U_0 x_{t-1} + b_0)$$

$$h_{t-1} = O_{t-1} \odot \sigma(s_{t-1})$$

- Selective read
- Selective write
- Selective forget
- GRU's Gates

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Selective write*



X



(<https://swayam.gov.in>)



([https://swayam.gov.in/nc\\_details/NPTEL](https://swayam.gov.in/nc_details/NPTEL))

bhandarishivay2001@gmail.com ▾

**NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)**

≡

Course  
outline

How does an  
NPTEL online  
course work?  
( )

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

Week 8 ()

Week 9 ()

week 10 ()

# Week 12: Assignment 12

The due date for submitting this assignment has passed.

**Due on 2022-10-19, 23:59 IST.**

**Assignment submitted on 2022-10-19, 18:43 IST**

1) Consider the task of generating a caption for a given image. Which of the following **1 point** model would you choose to encode the image and decode the caption respectively?

- RNN, RNN
- CNN, CNN
- RNN, CNN
- CNN, RNN

Yes, the answer is correct.

Score: 1

Accepted Answers:

CNN, RNN

2) Consider the textual entailment problem, where the objective is to predict the **1 point** hypothesis given a premise. Pick out the models you choose to encode and decode the sequence of words.

- RNN, RNN
- CNN, CNN
- RNN, CNN
- CNN, RNN

[Week 11 \(\)](#)[Week 12 \(\)](#)[Download Videos \(\)](#)[Books \(\)](#)[Text Transcripts \(\)](#)[Live Sessions \(\)](#)[Problem Solving Session \(\)](#)

Yes, the answer is correct.

Score: 1

Accepted Answers:

*RNN, RNN*

- 3) Consider the task of Image Question Answering, where the input consists of an image **1 point** and a question and the output is a finite word from a vocabulary. What is the model that can be used in the encoder?

- Combination of two CNNs
- Combination of two RNNs
- Combination of a CNN and RNN
- RNN

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Combination of a CNN and RNN*

- 4) Consider the task of Video captioning, where the input is a video and output is the **1 point** caption. Which of the following is the correct equation for the encoder?

- $h_t = RNN(h_{t-1}, CNN(x_{it}))$
- $h_t = CNN(h_{t-1}, RNN(x_{it}))$
- $h_t = RNN(h_{t-1}, RNN(x_{it}))$
- $h_t = CNN(h_{t-1}, CNN(x_{it}))$

Yes, the answer is correct.

Score: 1

Accepted Answers:

$h_t = RNN(h_{t-1}, CNN(x_{it}))$

- 5) Which of the following statements is True? **1 point**

- I. Encoder Decoder model can be made more expressive by adding an “attention” mechanism  
 II. Plotting the attention weights as a heatmap helps to check if the attention model learns something meaningful.

- I only
- II only
- Both
- None

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Both*

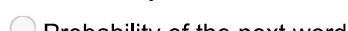
- 6) Given the first seven words of a sentence, our model has to predict the next word for a **1 point** Autocompletion task. Which of the following is the output at every timestep?



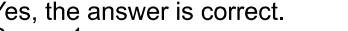
Probability of that  $i_{th}$  word in the vocabulary



Probability distribution over the vocabulary



Probability of the next letter



Probability of the next word

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Probability distribution over the vocabulary*

- 7) Consider the Language modelling problem, What is the size of the vector,  $\text{softmax}(V_{s_t} + c)$  in the equation,  $P(y_t|y_1^{t-1}) = \text{softmax}(V_{s_t} + c)_j$  ?

**1 point**



length of the  $n^{th}$  word in the vocabulary



length of the longest word in the vocabulary



length of the shortest word in the vocabulary

No, the answer is incorrect.

Score: 0

Accepted Answers:

*number of words in the vocabulary*

- 8) What should be the decoder used for the following:

**1 point**

- I. Document Classification  
II. Document Summarization



I - RNN

II- feedforward with softmax



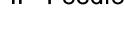
I - feedforward with softmax

II - RNN



I - feedforward with softmax

II-CNN



I - CNN

II - Feedforward with softmax

Yes, the answer is correct.

Score: 1

Accepted Answers:

*I - feedforward with softmax*

*II - RNN*

- 9) The Third convolutional layer of VGGNet is a  $56 \times 56 \times 256$  size feature map. If you **1 point** are asked to design a encoder-decoder model with attention mechanism then how many locations will the model have to learn to attend to?



196



784



256

3136

Yes, the answer is correct.

Score: 1

Accepted Answers:

3136

10) Consider a dialog between a User and a Bot as given below. Here, the objective is to **1 point** predict the response for this context.What is the encoder for given task?

Context

U: Can you suggest a good movie?

B: Yes, sure. How about Logan?

U: Okay, who is the lead actor?

Hierarchical RNN

Hierarchical CNN

Sequence of RNNs

Sequence of CNNs

Yes, the answer is correct.

Score: 1

Accepted Answers:

*Hierarchical RNN*