

# Quiz-1

**Due** Oct 5 at 8pm      **Points** 10      **Questions** 10  
**Available** Oct 2 at 12am - Oct 5 at 8pm 4 days      **Time Limit** 120 Minutes

## Instructions

Quiz-1 has 10 questions.

1. Each question is worth 1 point.
2. You have 2 hrs to finish the quiz once started.
3. Only one attempt is allowed.
4. For multiple choice with multiple correct answers, marks will only be given if all the right responses are selected.
5. The quiz will close on Monday Oct 5 8:00 PM.

## Attempt History

	Attempt	Time	Score
LATEST	<a href="#">Attempt 1</a>	120 minutes	7 out of 10

❗ Correct answers are hidden.

Score for this quiz: **7** out of 10

Submitted Oct 5 at 10:27am

This attempt took 120 minutes.

Incorrect

### Question 1

0 / 1 pts

A linear regression model between sales of a product (Y) and advertising budget (X) reported following statistics:  $R^2=0.6$  and TSS = 0.9. Select all that is true about this model.

Course Chat





The amount of variability in the sales data that is explained by performing the regression is 0.36 .



The amount of variability in sales is not explained by the regression is 0.45.



The proportion of variability in Y that can be explained by the model is 0.54.



The amount of variability inherent in the sales data that is not explained by the regression is performed is 0.45.



Send

## Question 2

1 / 1 pts

Here is an unordered list of different steps in a machine learning model lifecycle.

Model training, Monitoring, Model deployment, Retraining, Data preprocessing.

List these steps in correct order below, starting from the first step:

1. Data preprocessing

2. Model training

3. Model deployment

4. Monitoring

5. Retraining

**Answer 1:**

Data preprocessing

Course Chat



**Answer 2:**

Model training

**Answer 3:**

Model deployment

**Answer 4:**

Monitoring

**Answer 5:**

Retraining

**Question 3**

Send

Select all that is true for linearly separable binary dataset.

☐ All of the above.

Only a support vector machine (SVM) with non-linear kernels can guarantee 100% accuracy with this dataset.



When a single layer perceptron is trained with this dataset it may or may not converge depending on the separation margin between the two classes.



Non-linear transformations can further improve the accuracy of perceptron for this dataset.



None of the above.



The intersection of the convex hulls of the two be empty.

#### Question 4

Select the correct expression from A, B, C, and D for the least squares regression with

$$\hat{y} = a_0 + \sum_{j=1}^p a_j x_j$$

A.

$$\sum_i \left( y_i - a_0 - \sum_{j=1}^p x_j a_j \right)^2$$

B.

$$\sum_i \left( y_i - a_0 - \sum_{j=1}^p x_j a_j \right)^2 + \lambda \sum_{j=1}^p a_j$$

C.

$$\sum_i \left( y_i - a_0 - \sum_{j=1}^p x_j a_j \right)^2 + \lambda \sum_{j=0}^p a_j^2$$

D.  $\sum_i \left( y_i - \sum_{j=1}^p x_j a_j \right)^2 + \lambda \sum_{j=0}^p a_j^2$

☐ C

☒ B

☐ A



Send

Course Chat



☐ D

### Question 5

This question has 4 sub-questions each is worth 25 points.

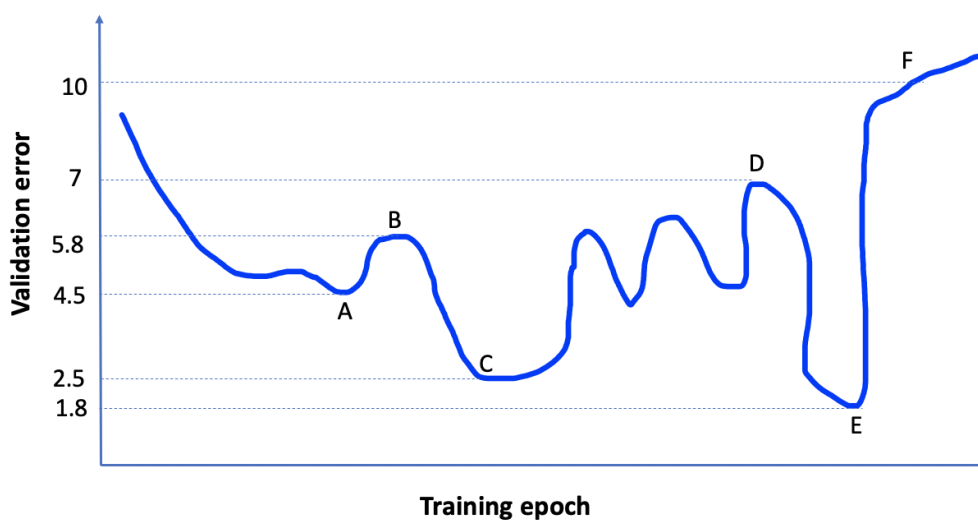
An early stopping criterion is created using the following formula: "Stop as soon as the generalization loss (GL) is greater than a threshold", where GL (expressed in percentage) is defined as:

$$GL(t) = 100 \cdot \left( \frac{E_{va}(t)}{E_{opt}(t)} - 1 \right)$$

where  $E_{va}(t)$  is the validation error at epoch  $t$  and  $E_{opt}(t)$  is the minimum validation error till time  $t$  defined as:

$$E_{opt}(t) := \min_{t' \leq t} E_{va}(t')$$

This criterion was applied to a neural network training and validation error as a function of training epoch is plotted. This plot is shown below:



From the validation error plot identify the following:

1. Early stopping epoch for 28% threshold

Course Chat

2. Early stopping epoch for 150% threshold

3. Epoch when overfitting truly begins

4. Epoch at which the model has the smallest

Your answers for the above should be in terms of B, D, E, and F only. These epochs are marked in the

**Answer 1:**

B

**Answer 2:**

D

**Answer 3:**

F

**Answer 4:**

E



Send

Incorrect

## Question 6

0 / 1 pts

Consider neural network training with gradient descent and regularization. The loss function and weight update equations are shown below. Select all that is correct from the choices below.

Course Chat

( $L_2$  regularization)  $L = \sum_{(x,y)}$

( $L_1$  regularization)  $L = \sum_{(x,y)}$

(gradient descent weight update)  $u$

☐

When lambda =1 and alpha =0.5, L2 regularization learning and the neural network weights will not converge to zero.

☐

When lambda =1 and alpha =0.5, L1 regularization learning and the neural network weights will not converge to zero.

☐

L1 regularization will always converge to non-zero values for all the weights and the resulting weight matrices are not sparse.

☒

When training model will always learn for both L2 and L1 for any choice of learning rate (alpha) regularizer (lambda).



Send

## Question 7

1 / 1 pts

Consider the outcome of a test to detect presence of a radioactive element in drinking water. Given 1000 samples of drinking water from different geographies the test results are summarized as follows (with positive being the sample has radioactive element present or is contaminated). The test results are summarized below:

Course Chat



		True value
Test outcome	Radioactive present	Radioactive present
	Radioactive absent	Radioactive absent
		100
		250



Send

Based on the test outcome answer following questions. You just need to fill the right numeric value for each question's answer. For answers with decimal parts round up to 2 places.

1. How many samples did the test correctly classified ?

2. What is the chance (in percentage) that a sample identified as not contaminated is in fact contaminated ?  %

3. What is the chance (in percentage) that a sample identified as contaminated is in fact contaminated ?  %

4. What is the F<sub>1</sub> score for this test ?

**Answer 1:**

**Course Chat**





450

**Answer 2:**

41.67

**Answer 3:**

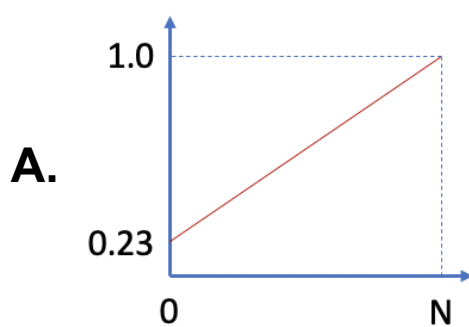
25

**Answer 4:**

0.27

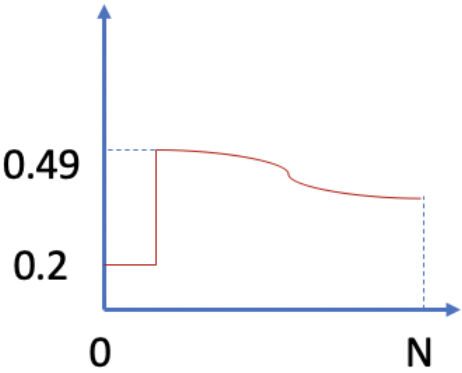
**Question 8**

There are six candidates for assigning dropout layers of a deep neural network during training. A dropout profile is represented as a function of the depth of neural network layer, with input layer being at depth 0 and last hidden layer at depth  $N$ . Correctly select all the profiles which will result in a dead neural network meaning the network will not learn anything (weights will not get updated) during training. Each profile is shown as a graph with x-axis being the depth of the layer and y axis being the drop-out probability.

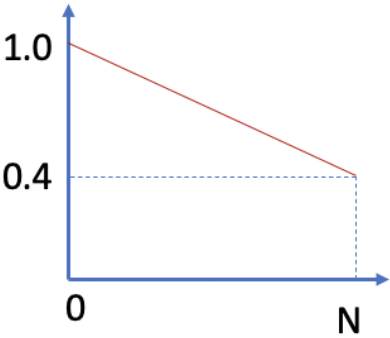


Send

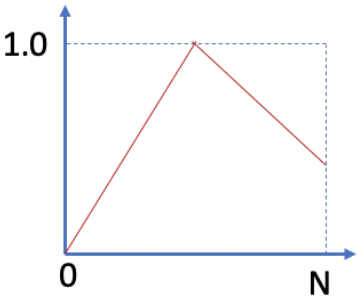
B.



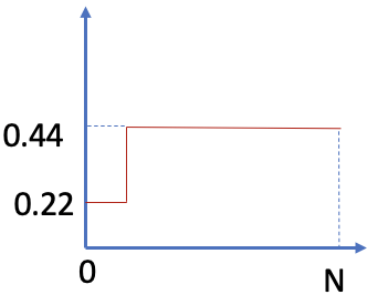
C.



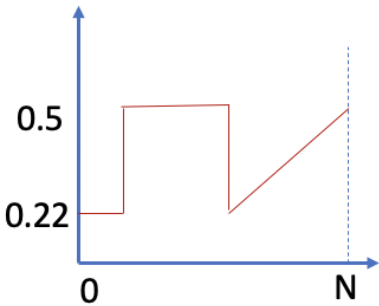
D.



E.



F.

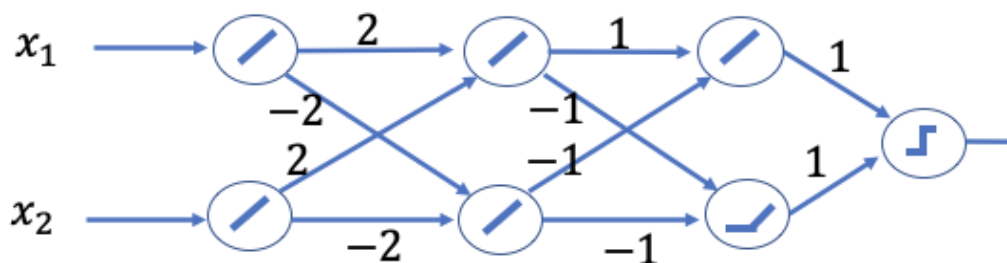


Send



☐ E☒ D☐ B☐ F☒ A☒ C**Question 9**

Consider a neural network as shown below.



The signum function is defined as  $\text{sign}(x) = +1$  for positive  $x$  (including 0) and  $-1$  for negative  $x$ . The output of neural network is  $+1$  for positive classes and  $-1$  for negative classes. Select all the values of  $(x_1, x_2)$  which will be classified as negative classes by this neural network.

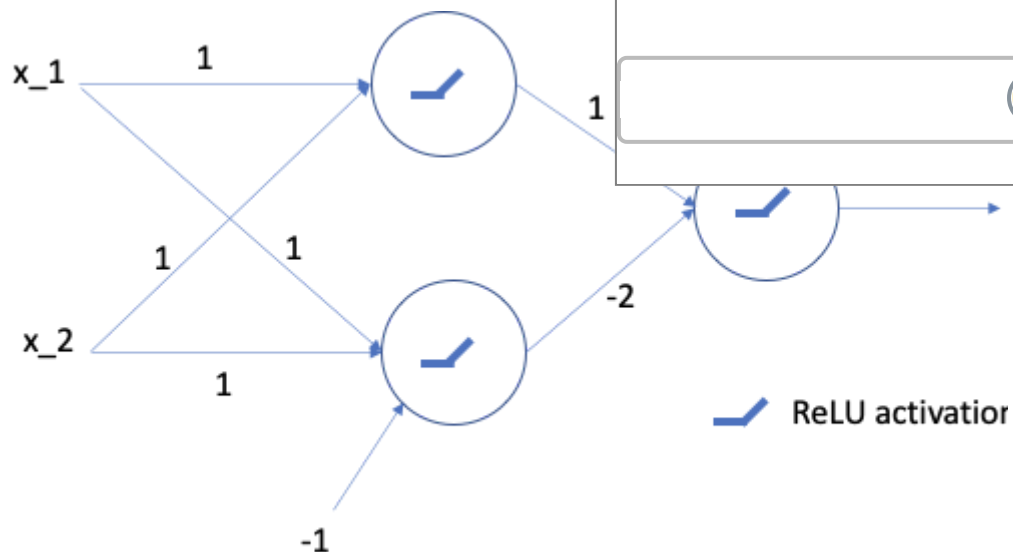
☐  $x_1=1, x_2=1$ ☒  $x_1=5, x_2=-8$ ☐  $x_1=0, x_2=0$ **Course Chat**

☒  $x_1 = -1, x_2 = -1$ ☐  $x_1 = 1, x_2 = -1$ ☒  $x_1 = -8, x_2 = 5$ 

Incorrect

**Question 10**

Consider a neural network with 3 ReLU activation layer and one in output layer as show below. this network is computing ?

☒ NOT☐ XOR☐ AND☐ ORQuiz Score: **7** out of 10

Course Chat

