# Practice

# Common Instructions for questions 1-3

Consider the following data set containing one feature:

| x | y |
|---|---|
| 1 | +1 |
| -1 | +1 |
| -3 | -1 |
| -2 | -1 |

Consider $w_0 = 1$ and $w_1 = 1$.

# Question-1

## Statement

If $\hat{y} = sign(w^T x)$, what will be the value of 0-1 loss for this data?

## Answer

0 (No range required)

## Solution

Zero-one loss: if $y = sign(w^T x)$, then loss = 0, else 1

Here, $w^T x = w_0 + w_1 x$

| $x$ | $w^T x$ | $sign(w^T x)$ | y |
|---|---|---|---|
| 1 | 1+1 = 2 | +1 | +1 |
| -1 | -1+1 = 0 | +1 | +1 |
| -3 | -3+1 = -2 | -1 | -1 |
| -2 | -2+1 = -1 | -1 | -1 |

Since $sign(w^T x) = y$ for all data points, zero-one loss = 0

# Question-2

## Statement

What will be the value of squared loss, i.e., $(\hat{y} - y)^2$ where $\hat{y} = w^T x$?

Note: For this loss, use $y = 0$ wherever $y = -1$.

## Answer

7 (No range required)

## Solution

| $x$ | $\hat{y} = w^T x$ | y | $(\hat{y} - y)^2$ |
|-----|-------------------|---|-------------------|
| 1 | 1+1 = 2 | 1 | 1 |
| -1 | -1+1 = 0 | 1 | 1 |
| -3 | -3+1 = -2 | 0 | 4 |
| -2 | -2+1 = -1 | 0 | 1 |

Hence squared loss = 1+1+4+1 = 7

# Question-3

## Statement

What will be the value of hinge loss, i.e., $max(0, 1 - w^T xy)$?

## Answer

1 (No range required)

## Solution

| $x$ | $w^T x$ | $y$ | $w^T xy$ | $1 - w^T xy$ | **Hinge loss** |
|---|---|---|---|---|---|
| 1 | 1+1 = 2 | +1 | 2 | -1 | 0 |
| -1 | -1+1 = 0 | +1 | 0 | 1 | 1 |
| -3 | -3+1 = -2 | -1 | 2 | -1 | 0 |
| -2 | -2+1 = -1 | -1 | 1 | 0 | 0 |

Total error = 0+1+0+0 = 1

# Question-4

## Statement

Given a cat image, you want to classify which of the 10 cat breeds it belongs to, using a neural network.

Which loss function will be appropriate?

## Options

### (a)

0-1 Loss

### (b)

Cross entropy Loss

### (c)

Squared Loss

### (d)

Hinge Loss

## Answer

(b)

## Solution

It's a multi-class classification problem.  Hence, cross entropy loss will be correct.

(0-1 loss is non-continuous and non-differentiable

Squared loss is not an appropriate loss for classification.

Hinge loss is used for binary classification.)

# Question-5

## Statement

Assume that $y_i = [1, 0, 0]$ and $\hat{y}_i = [0.7, 0.2, 0.1]$ for a data point $x_i$. What will be the value of the cross-entropy loss?

## Answer

0.5146 (Range: 0.50 to 0.55)

## Solution

Cross- entropy loss = $-\sum_i y_i log_2 \hat{y}_i$

$= -1 * log_2(0.7) - 0 * log_2(0.2) - 0 * log_2(0.1)$

$= -log_2(0.7)$

$= 0.5146$

# Question-6

## Statement

Following is the output produced by an activation function at some hidden layer of a neural network:

[0, 4.9, 0, 5.2, 7.4, 0]

Which of the following could possibly be the activation function?

## Options

**(a)**

Sigmoid

**(b)**

ReLU

**(c)**

Tanh

## Answer

(b)

## Solution

Sigmoid transforms values in the range -1 to 0. So, it may not be correct.

Tanh transforms values between -1 and 1, so, it may not be correct.

ReLU transforms negative values to zero, and keeps the positive values as it is, so it may be the activation function used.
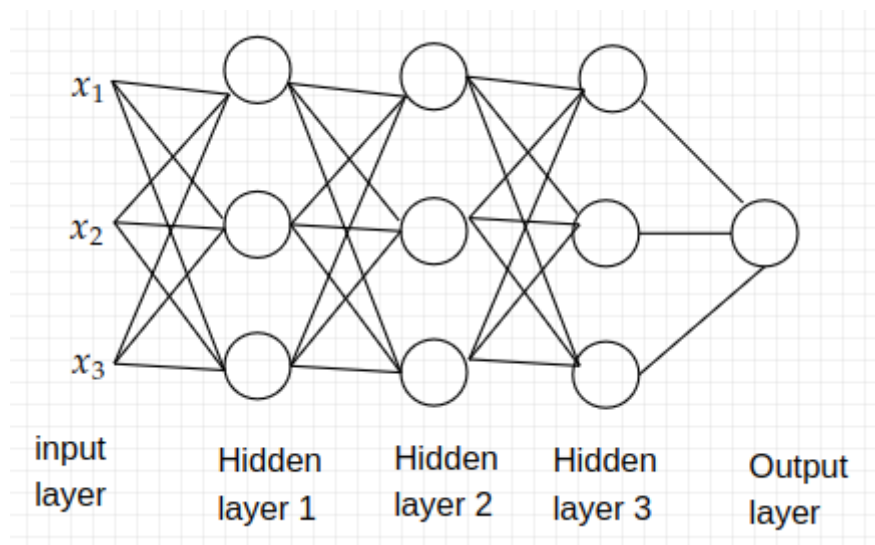
# Question-7

## Statement

Consider a neural network with 3 inputs and one output. If there are 3 hidden layers each with 3 neurons, how many parameters need to be learnt by the back-propagation algorithm?

Note: Assume that each hidden and output layer neuron also contains a bias.

## Answer

40 (No range required)

## Solution



#weights from input to hidden layer 1= 3*3 = 9

#weights from hidden layer 1 to hidden layer 2= 3*3 = 9

#weights from hidden layer 2 to hidden layer 3= 3*3 = 9

#weights from hidden layer 3 to output layer= 3*1 = 3


Total number of neurons = 10

Hence, number of bias terms = 10


Therefore, total number of parameters to be computed = 9+9+9+3+10 = 40

# Question-8

## Statement

Which of the following is/are true?

## Options

### (a)

Both Sigmoid and Softmax are activation functions.

### (b)

Sigmoid is used for binary classification tasks, while SoftMax applies to multiclass problems.

### (c)

SoftMax function is an extension of the Sigmoid function.

### (d)

Sigmoid function is also called Logistic function.

### (e)

Both functions transform a real value to a number between 0 and 1.

## Answer

(a), (b), (c), (d), (e)

## Solution

Softmax function is a generalization of sigmoid function, to be used in multi-class classification, such that of there are k classes, the sum of probability values returned for each of these classes is equal to 1.