Deep Learning

Assignment- Week 4

TYPE OF QUESTION: MCQ/MSQ

Number of questions: 10 **Total mark: 10 X 1 = 10**

QUESTION 1:

A given cost function is of the form J(θ) = θ^2 - θ+2? What is the weight update rule for gradient descent optimization at step t+1? Consider, α=0.01 to be the learning rate.

a. $\theta_{t+1} = \theta_t - 0.01(2\theta - 1)$
b. $\theta_{t+1} = \theta_t + 0.01(2\theta)$
c. $\theta_{t+1} = \theta_t - (2\theta - 1)$
d. $\theta_{t+1} = \theta_t - 0.01(\theta - 1)$

Correct Answer: a

Detailed Solution:

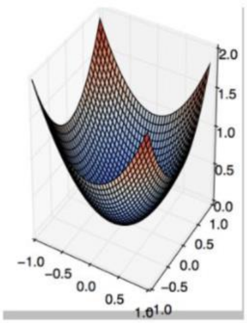
$$\frac{\partial J(\theta)}{\partial \theta} = 2\theta - 1$$

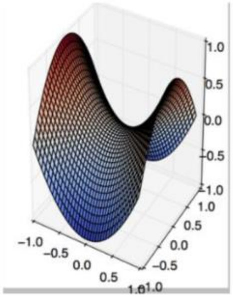
So, weight update will be

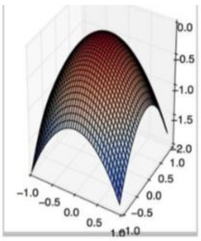
$$\theta_{t+1} = \theta_t - 0.01(2\theta - 1)$$

QUESTION 2:

Can you identify in which of the following graph gradient descent will not work correctly?







a. First figure
b. Second figure
c. First and second figure
d. Fourth figure

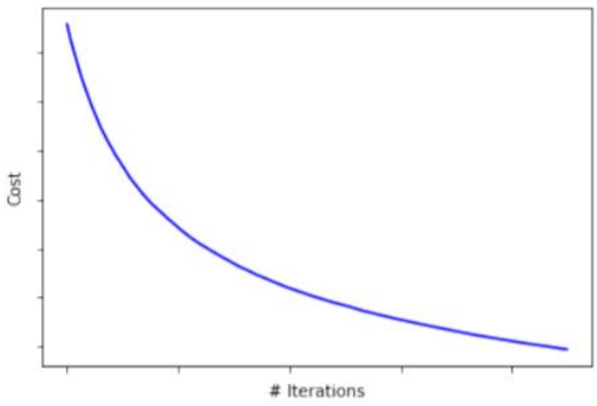
Correct Answer: b

Detailed Solution:

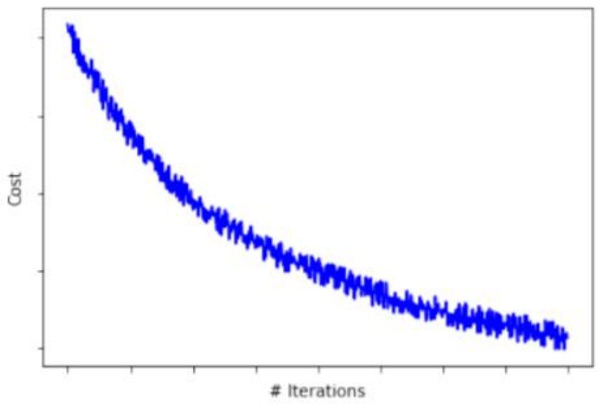
This is a classic example of saddle point problem of gradient descent. In the second graph gradient descent may get stuck in the saddle point.

QUESTION 3:

From the following two figures can you identify which one corresponds to batch gradient descent and which one to Stochastic gradient descent?



(a) Graph A



(b) Graph B

a. Graph-A: Batch gradient descent, Graph-B: Stochastic gradient descent

b. Graph-B: Batch gradient descent, Graph-A: Stochastic gradient descent

c. Graph-A: Batch gradient descent, Graph-B: Not Stochastic gradient descent

d. Graph-A: Not batch gradient descent, Graph-B: Not Stochastic gradient descent

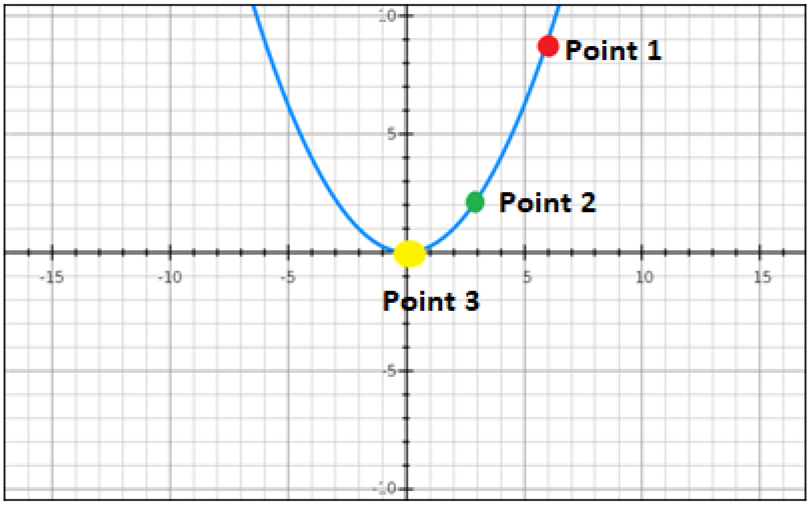
Correct Answer: a

Detailed Solution:

The graph of cost vs epochs is quite smooth for batch gradient descent because we are averaging over all the gradients of training data for a single step. The average cost over the epochs in Stochastic gradient descent fluctuates because we are using one example at a time.

QUESTION 4:

Suppose for a cost function $J(\theta) = 0.25\theta^2$ as shown in graph below, in which point do you feel magnitude of weight update will be more? θ is plotted along horizontal axis.



a. Red point (Point 1)
b. Green point (Point 2)
c. Yellow point (Point 3)
d. Red (Point 1) and yellow (Point 3) have same magnitude of weight update

Correct Answer: a

Detailed Solution:

Weight update is directly proportional to the magnitude of the gradient of the cost function. In our case, $\frac{\partial J(\theta)}{\partial \theta} = 0.5\theta$. **So, the weight update will be more for higher values of** θ.

QUESTION 5:

Which logic function can be performed using a 2-layered Neural Network?

a. AND
b. OR
c. XOR
d. All

Correct Answer: d

Detailed Solution:

A two layer neural network can be used for any type logic Gate (linear or non linear) implementation.

QUESTION 6:

Let **X** and **Y** be two features to discriminate between two classes. The values and class labels of the features are given hereunder. The minimum number of neuron-layers required to design the neural network classifier

X	Y	#Class
0	2	Class-II
1	2	Class-I
2	2	Class-I
1	3	Class-I
1	-3	Class-II

a. 1
b. 2
c. 4
d. 5

Correct Answer: a.

Detailed Solution:

Plot the feature points. They are linearly separable. Hence single layer is able to do the classification task.

QUESTION 7:

Which among the following options give the range for a logistic function?

a. -1 to 1
b. -1 to 0
c. 0 to 1
d. 0 to infinity

Correct Answer: c

Detailed Solution:

Refer to lectures, specifically the formula for logistic function.

QUESTION 8:

The number of weights (including bias) to be learned by the neural network having 3 inputs and 2 classes and a hidden layer with 5 neurons is:

a. 12
b. 15
c. 25
d. 32

Correct Answer: d

Detailed Solution:

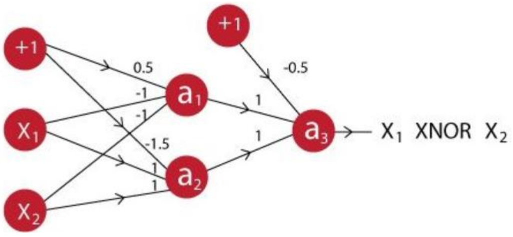
Please refer to lecture note week 4

(#input=3)+1(bias)x(#Hidden nodes=5) =(3+1)x5= 20 (#weights in 1st layer)
(#Hidden Nodes+1(bias))x(#classes=2)=(5+1)x2=12 (#weights in 2nd layer)

Hence, total weights= 20+12 =32

QUESTION 9:

For a XNOR function as given in the figure below, activation function of each node is given by: $f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases}$. Consider $X_1 = 1$ and$X_2 = 0$, what will be the output for the above neural network?



a. 1.5
b. 2
c. 0
d. 1

Correct Answer: c

Detailed Solution:

Output of $\boldsymbol{a_1}$: $\boldsymbol{f(0.5*1 + -1*1 + -1*0) = f(-0.5) = 0}$

Output of $\boldsymbol{a_2}$: $\boldsymbol{f(-1.5*1 + 1*1 + 1*0) = f(-0.5) = 0}$

Output of $\boldsymbol{a_3}$: $\boldsymbol{f(-0.5*1 + 1*0 + 1*0) = f(-0.5) = 0}$

So, the correct answer is c.

QUESTION 10:

Which activation function is more prone to vanishing gradient problem?

a. ReLU

b. Tanh

c. sigmoid

d. Threshold

Correct Answer: b

Detailed Solution:

Please refer to the lectures of week 4.

************END*******