Deep Learning

Assignment- Week 9

TYPE OF QUESTION: MCQ/MSQ

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

QUESTION 1:

What can be a possible consequence of choosing a very small learning rate?

a. Slow convergence
b. Overshooting minima
c. Oscillations around the minima
d. All of the above

Correct Answer: a

Detailed Solution:
Choosing a very small learning rate can lead to slower convergence and thus option (a) is correct.

QUESTION 2:

The following is the equation of update vector for momentum optimizer. Which of the following is true for γ?

$$V_t = \gamma V_{t-1} + \eta \nabla_\theta J(\theta)$$

a. γ is the momentum term which indicates acceleration
b. γ is the step size
c. γ is the first order moment
d. γ is the second order moment

Correct Answer: a

Detailed Solution:

A fraction of the update vector of the past time step is added to the current update vector. γ is that fraction which indicates how much acceleration you want and its value lies between 0 and 1.

QUESTION 3:

Which of the following is true about momentum optimizer?

a. It helps accelerating Stochastic Gradient Descent in right direction
b. It helps prevent unwanted oscillations
c. It helps to know the direction of the next step with knowledge of the previous step
d. All of the above

Correct Answer: d

Detailed Solution:

Option (a), (b) and (c) all are true for momentum optimiser. Thus, option (d) is correct.

QUESTION 4:

Let $J(\theta)$ be the cost function. Let the gradient descent update rule for θ_i be,

$$\theta_{i+1} = \theta_i + \nabla\theta_i$$

What is the correct expression of $\nabla\theta_i$. α is the learning rate.

a. $-\alpha \frac{dJ(\theta_i)}{d\,\theta_i}$
b. $\alpha \frac{dJ(\theta_i)}{d\,\theta_i}$
c. $-\frac{dJ(\theta_i)}{d\theta_{i+1}}$
d. $\frac{dJ(\theta_i)}{d\theta_i}$

Correct Answer: a

Detailed Solution:

Gradient descent update rule for θ_i is,
$\theta_{i+1} = \theta_i - \alpha \frac{dJ(\theta_i)}{d\,\theta_i}$, α is the learning rate

---

QUESTION 5:

A given cost function is of the form J(θ) =6 θ^2 - 6θ+6? What is the weight update rule for gradient descent optimization at step t+1? Consider, α to be the learning rate.

a. $\theta_{t+1} = \theta_t - 6\alpha(2\theta - 1)$
b. $\theta_{t+1} = \theta_t + 6\alpha(2\theta)$
c. $\theta_{t+1} = \theta_t - \alpha(12\theta - 6 + 6)$

d. $\theta_{t+1} = \theta_t - 6\alpha(2\theta + 1)$

Correct Answer: a

Detailed Solution:

$\frac{\partial J(\theta)}{\partial \theta} = 12\theta - 6$

So, weight update will be

$\theta_{t+1} = \theta_t - 6\alpha(2\theta - 1)$

QUESTION 6:

If the first few iterations of gradient descent cause the function $f(\theta_0,\theta_1)$ to increase rather than decrease, then what could be the most likely cause for this?

a. we have set the learning rate to too large a value
b. we have set the learning rate to zero
c. we have set the learning rate to a very small value
d. learning rate is gradually decreased by a constant value after every epoch

Correct Answer: a

Detailed Solution:

If learning rate were small enough, then gradient descent should successfully take a tiny small downhill and decrease $f(\theta_0,\theta_1)$ at least a little bit. If gradient descent instead increases the objective value that means learning rate is too high.

QUESTION 7:

For a function $f(\theta_0,\theta_1)$, if θ_0 and θ_1 are initialized at a global minimum, then what should be the values of θ_0 and θ_1 after a single iteration of gradient descent?

a. θ_0 and θ_1 will update as per gradient descent rule
b. θ_0 and θ_1 will remain same
c. Depends on the values of θ_0 and θ_1
d. Depends on the learning rate

Correct Answer: b

Detailed Solution:

At a local minimum, the derivative (gradient) is zero, so gradient descent will not change the parameters.

QUESTION 8:

What can be one of the practical problems of exploding gradient?

a. Too large update of weight values leading to unstable network
b. Too small update of weight values inhibiting the network to learn
c. Too large update of weight values leading to faster convergence
d. Too small update of weight values leading to slower convergence

Correct Answer: a

Detailed Solution:

Exploding gradients are a problem where large error gradients accumulate and result in very large updates to neural network model weights during training. This has the effect of your model being unstable and unable to learn from your training data.

QUESTION 9:

What are the steps for using a gradient descent algorithm?

1. Calculate error between the actual value and the predicted value
2. Update the weights and biases using gradient descent formula
3. Pass an input through the network and get values from output layer
4. Initialize weights and biases of the network with random values
5. Calculate gradient value corresponding to each weight and bias

a. 1, 2, 3, 4, 5
b. 5, 4, 3, 2, 1
c. 3, 2, 1, 5, 4
d. 4, 3, 1, 5, 2

Correct Answer: d

Detailed Solution:

Initialize random weights, and then start passing input instances and calculate error response from output layer and back-propagate the error through each subsequent layers. Then update the neuron weights using a learning rate and gradient of error. Please refer to the lectures of week 4.

QUESTION 10:

You run gradient descent for 15 iterations with learning rate $\eta = 0.3$ and compute error after each iteration. You find that the value of error decreases very slowly. Based on this, which of the following conclusions seems most plausible?

a. Rather than using the current value of a, use a larger value of η
b. Rather than using the current value of a, use a smaller value of η
c. Keep $\eta = 0.3$
d. None of the above

Correct Answer: a

Detailed Solution:

Error rate is decreasing very slowly. Therefore increasing the learning rate is a most plausible solution.

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