Introduction to Machine Learning -IITKGP

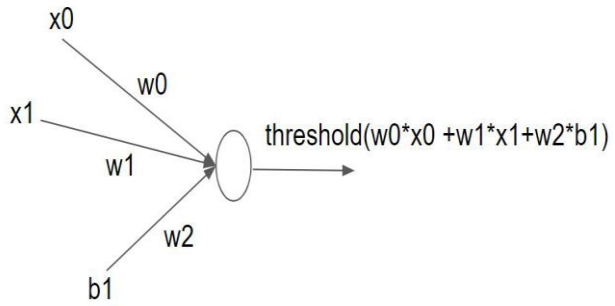
Assignment - 6

TYPE OF QUESTION: MCQ/MSQ

Number of questions: 15 Total mark: 2 * 15 = 30

Do answer questions 1-2 with the data provided below

1. Given below the neural network, find the appropriate weights for w0, w1, and w2 to represent the AND function. Threshold function = {1, if output >0; 0 otherwise}. x0 and x1 are the inputs and b1=1 is the bias.



a. w0=1, w1=1, w2=1
b. w0=1, w1=1, w2=-1
c. w0=-1, w1=-1, w2=-1
d. w0=2, w1=-2, w3=-1

Correct Answer: b

Detailed Solution: for x0=1, x1=1 and b1=1,
option (b) gives 1*1+1*1+-1=1>0=1
Similarly, check for others.

---

2. Fill in the correct weights to represent OR function:

a. w0=1, w1=1, w2=0
b. w0=1, w2=1, w3=1
c. w0=1, w1=1, w2=-1
d. w0=-1, w1=-1, w2=-1

Correct Answer: a

Detailed Solution: Same as Solution to Q. No 1.

---

3. Which of the following gives non-linearity to a neural network

    a. Gradient descent
    b. Bias
    c. ReLU Activation Function
    d. None

Correct Answer: c
Detailed Solution: An activation function such as ReLU gives a non-linearity to the neural network.

---

4. Suppose you are to design a system where you want to perform word prediction also known as language modeling. You are to take the output from the previous state and also the input at each step to predict the next word. The inputs at each step are the words for which the next words are to be predicted. Which of the following neural network would you use?

    a. Multi-Layer Perceptron
    b. Recurrent Neural Network
    c. Convolutional Neural Network
    d. Perceptron

Correct Answer: b
Detailed Solution:
Recurrent Neural Network (RNN) is a type of Neural Network where the output from the previous step is fed as input to the current step. Refer to lecture nodes for detailed explanation.

---

5. For a fully-connected deep network with one hidden layer, increasing the number of hidden units should have what effect on bias and variance?

    a. Decrease bias, increase variance
    b. Increase bias, increase variance
    c. Increase bias, decrease variance
    d. No change

Correct Answer: a
Detailed Solution: Adding more hidden units should decrease bias and increase variance. In general, more complicated models will result in lower bias but larger variance, and adding more hidden units certainly makes the model more complex

---

<u>Do answer questions 6-7 with the data provided below</u>

6. You are given the task of predicting the price of a house given the various features of a house such as number of rooms, area (sq ft), etc.

    How many neurons should you have at the output?

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

Correct Answer: c
Detailed Solution: The price of a house is a single value. Hence, one neuron is enough.

---

7. What should be the loss function used to train the model?
    a. Multi-Class Cross-Entropy Loss
    b. Mean Squared Error
    c. Binary Cross-Entropy Loss

    Correct Answer: b
    Detailed Solution: Mean Squared Error finds the average squared difference between the predicted value and the true value. Since there are no classes involved as in case of classification tasks, Cross-Entropy Loss of any type doesn't qualify to be a loss function.

---

8. A Convolutional Neural Network (CNN) is a Deep Neural Network that can extract various abstract features from an input required for a given task. Given the operations performed by a CNN on an input:
    1) Max Pooling
    2) Convolution Operation
    3) Flatten
    4) Forward propagation by Fully Connected Network
    Identify the correct sequence from the options below:
    a. 4,3,2,1
    b. 2,1,3,4
    c. 3,1,2,4
    d. 4,2,1,3

    Correct Answer: b
    Detailed Solution:
    Follow the lecture slides.

---

9. An autoencoder is a Neural Network architecture used to create lower dimensional input representation. Which of the following statements are true about it?

    a. It is an unsupervised algorithm similar to PCA
    b. It can generate new data by learning the probability distribution
    c. Its target output is the input
    d. Autoencoders have linear encoder and decoder

    Correct Answer: a, c

Detailed Solution: Autoencoders perform dimensionality reduction and are unsupervised similar to PCA. The second option is true for Variational Auto Encoder which is a generative model, unlike conventional autoencoders. Autoencoders can have any form of encoders and decoders.

---

10. In a simple MLP model with 8 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer. What is the size of the weight matrices between hidden to output layer and input to hidden layer?

    a. [5 X 1], [8 X 5]
    b. [8 X 5], [1 X 5]
    c. [3 X 1], [3 X 3]
    d. [3 X 3], [3 X 1]

    Correct Answer. a

    Explanation:

    The weight matrix between the hidden layer (5 neurons) and the output layer (1 neuron) will be of size [5 X 1].
    The weight matrix between the input layer (8 neurons) and the hidden layer (5 neurons) will be of size [8 X 5].

---

11. If you increase the number of hidden layers in a Multi-Layer Perceptron, the classification error of test data always decreases. True or False?
    a. True
    b. False

    Correct Answer: b

Explanation: Increasing the number of hidden layers in a Multi-Layer Perceptron (MLP) doesn't guarantee a decrease in classification error for test data. While adding more hidden layers can potentially help the network learn more complex representations, it can also lead to overfitting, where the model performs well on the training data but poorly on the test data. The optimal number of hidden layers depends on the complexity of the problem, the amount of available data, and careful tuning of various hyperparameters.

---

12. Which of the following represents the range of output values for a sigmoid function?
    a. -1 to 1
    b. $-\infty$to ∞
    c. 0 to 1
    d. 0 to ∞

Correct answer: c

Explanation: A sigmoid function, such as the logistic sigmoid function, maps input values to an output range between 0 and 1. As the input values become larger, the output of the sigmoid function approaches 1, and as the input values become more negative, the output approaches 0. This property makes sigmoid

functions useful for tasks that involve binary classification or when you want to squash values into a limited range.

---

13. A single perceptron can compute the XOR function
    a. True
    b. False

Correct answer: b

Explanation: A single perceptron is not capable of directly computing the XOR function. The XOR function is not linearly separable, which means that a single perceptron, which uses a linear decision boundary, cannot accurately represent it. The XOR function's output is 1 when the number of input 1s is odd, and the output is 0 when the number of input 1s is even. This behavior cannot be achieved with just a single linear threshold. However, XOR can be computed using a multi-layer perceptron (a neural network with at least one hidden layer), which can model more complex decision boundaries and accurately represent non-linear relationships like the XOR function.

---

14. What are the steps for using a gradient descent algorithm?
    1. Calculate error between the actual value and the predicted value
    2. Repeat until you find the best weights of network
    3. Pass an input through the network and get values from output layer
    4. Initialize random values for weight and bias
    5. Go to each neuron which contributes to the error and change its respective values to reduce the error
    a. 4,3,1,5,2
    b. 1,2,3,4,5
    c. 3,4,5,2,1
    d. 2,3,4,5,1

Correct answer: a

Explanation:

Initialize random values for weight and bias: The process begins by initializing random values for the weights and biases in the neural network. This is necessary to start the optimization process.

Pass an input through the network and get values from the output layer: The input data is propagated through the network to obtain the predicted values at the output layer. This step is the forward pass and helps to calculate the predicted output.

Calculate the error between the actual value and the predicted value: The calculated predicted values are compared with the actual target values to compute the error or loss. This quantifies how far off the predictions are from the true values.

Go to each neuron that contributes to the error and change its respective values to reduce the error: This step involves backpropagation, where the gradients of the loss with respect to the network's parameters (weights and biases) are computed. The weights are adjusted in a way that minimizes the error by using gradient information.

Repeat until you find the best weights of the network: Steps 2 through 4 are repeated iteratively for a certain number of epochs or until convergence criteria are met. The goal is to find the weights that minimize the error and optimize the network's performance.

So, the correct sequence of steps is 4, 3, 1, 5, 2.

---

15. The back-propagation learning algorithm applied to a two-layer neural network
    a. always finds the globally optimal solution.
    b. finds a locally optimal solution which may be globally optimal.
    c. never finds the globally optimal solution.
    d. finds a locally optimal solution which is never globally optimal

Correct answer: b

Explanation:

The backpropagation learning algorithm applied to a two-layer neural network, or any neural network, does not guarantee to find the globally optimal solution but rather tends to find a locally optimal solution.

a. always finds the globally optimal solution: This is incorrect. Neural networks can have complex loss surfaces with many local minima, making it challenging for backpropagation to guarantee the globally optimal solution.

b. finds a locally optimal solution which may be globally optimal: This is the most accurate description. Backpropagation seeks to minimize the loss function by iteratively updating weights using gradient descent. It converges to a local minimum that represents a good solution, but it may also coincide with the globally optimal solution, especially in simpler cases.

c. never finds the globally optimal solution: This is not entirely accurate. While it's challenging to guarantee finding the global optimum due to the complex nature of neural network loss landscapes, it's still possible for the found local optimum to also be the global optimum, especially in simpler settings.

d. finds a locally optimal solution which is never globally optimal: This is too strong a statement. While a locally optimal solution may not always be globally optimal, it's not accurate to state that it's "never" globally optimal.

In summary, backpropagation in a two-layer neural network often converges to a locally optimal solution. This local optimum might also be the global optimum, depending on the specific problem and network architecture.

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