Introduction to Deep Learning

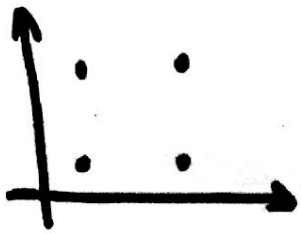
Mid-Term Examination 2019/4/18 8:45~10:00

Written Test 50%

Make the correct choice in each case. 3 points for each, totally 54 points, of which 4 points is bonus.

1) Is the data linearly separable?

A



A) Yes
B) No
C) May be

2) Question Context:

A

Statement 1: It is *possible* to train a network well by initializing all the weights as 0
Statement 2: It is *possible* to train a network well by initializing biases as 0

Which of the statements given above is true?

A) Statement 1 is true, while Statement 2 is false
B) Statement 2 is true, while statement 1 is false
C) Both statements are true
D) Both statements are false

C **3) The number of nodes in the input layer is 8 and the hidden layer is 6. The maximum number of connections from the input layer to the hidden layer are**

A) 48

B) Less than 48

C) More than 48

D) It is an arbitrary value

C **4) In a simple Multi-Layer Perceptron (MLP) model with 6 neurons in the input layer, 4 neurons in the hidden layer and 1 neuron in the output layer. What is the size of the weight matrices between hidden & output layers and between input & hidden layers?**

A) [1 X 4] , [4 X 6]

B) [6 X 4] , [1 X 4]

C) [6 X 4] , [4 X 1]

D) [4 x 1] , [6 X 4]

C ***5) Which of the following functions can be used as an activation function in the output layer if we wish to predict the probabilities of n classes (p1, p2..pk) such that sum of p over all n equals to 1?***

A) Sigmoid

B) ReLu

C) Softmax

D) Any of the above functions

C **6) Assume a simple MLP model with 3 neurons and inputs = 1, 2, 3. The weights to the input neurons are 4, 5, and 6, respectively. Assume the activation function is a linear function f(z) = 3z. What will be the output?**

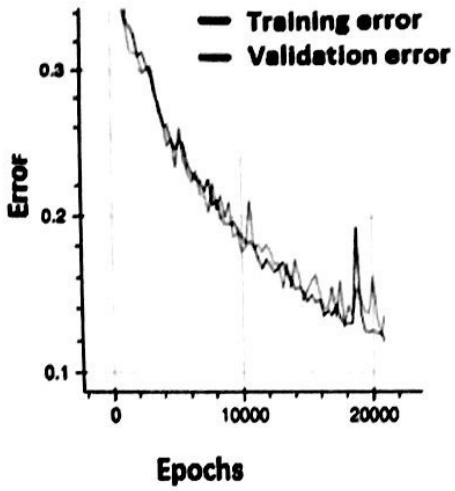
A) 32

B) 643

C) 96

D) 48

B **7) In the graph below, we observe that the error has many "ups and downs"**



Should we be worried?

A. Yes, because this means there is a problem with the learning rate of neural network.
B. No, as long as there is a cumulative decrease in both training and validation error, we don't need to worry.

A **8) In the neural network, every parameter "can" have its own learning rate, though it is not done often.**

A) TRUE
B) FALSE

9) Dropout can be applied at the input layer of a Neural Network model.

B

A) TRUE
B) FALSE

A **10) I am working with a fully connected NN architecture having one hidden layer with 3 neurons and one output neuron to solve a binary classification challenge. Below is the structure of input and output:**

Input dataset: [[1,0,1,0] , [1,0,1,1] , [0,1,0,1]]

Output: [[1] , [1] , [0]]

To train the model, I have initialized all weights for hidden and output layers as 1.

Will the model be able to learn the pattern in the data?

A) Yes
B) No

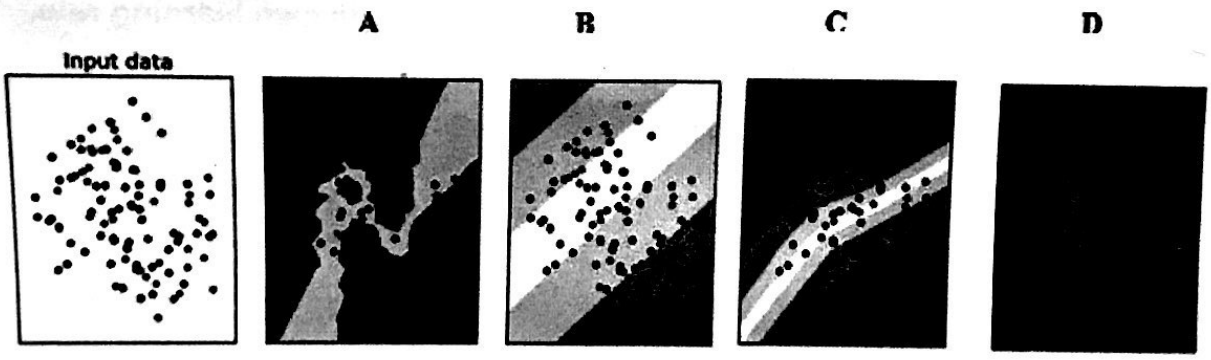
11) Which of the following neural network training challenge can be solved using batch normalization?

A

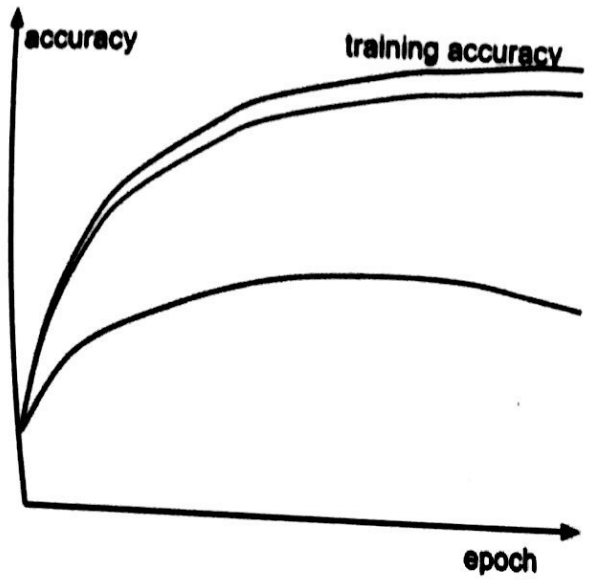
A) Overfitting
B) Restrict activations to become too high or low
C) Training is too slow
D) All of the above

12) Which of the following is a decision boundary of Neural Network?

D



A) B
B) A
C) D
D) C
E) All of the above



B 13) **The red curve above denotes training accuracy with respect to each epoch in a deep learning algorithm. Both the green and blue curves denote validation accuracy.**

Which of these indicate overfitting?

A) Green Curve
B) Blue Curve

B **14) Which of the following statement is true regarding dropout?**

1. Dropout is a way to approximate by combining several different architectures
2: Dropout demands high learning rates
3: Dropout can help prevent overfitting

A) Both 1 and 2
B) Both 1 and 3
C) Both 2 and 3
D) All 1, 2 and 3

A **15) Sentiment analysis using Deep Learning is a many-to one prediction task**

A) TRUE
B) FALSE

F **16) What steps can we take to prevent overfitting in a Neural Network?**

A) Data Augmentation
B) Weight Sharing
C) Early Stopping
D) Dropout
E) All of the above

B **17) Which gradient technique is more advantageous when the data is too big to handle in RAM simultaneously?**

A. Full Batch Gradient Descent
B. Stochastic Gradient Descent
C. All of the above
D. None of the above

B ***18) What if we use a learning rate that's too large?***

A. Network will converge
B. Network will not converge
C. Can't Say