Final exam 2022

Deep Learning (Trường Đại học Bách khoa Hà Nội)

Exam Key Code: 112

(The exam sheet consists of 8 pages)

Time: 120'

Name: . . . . **Class:** . . . . **Student ID:** . . . .

SINGLE ANSWER MULTIPLE CHOICE QUESTIONS

Question 01. Which of the following is **FALSE** about Deep Learning and Machine Learning algorithms?
(A) Deep Learning algorithms are more interpretable as compared to Machine Learning algorithms
(B) None of the above
(C) Data augmentation can be done easily in Deep Learning as compared to Machine Learning
(D) Deep Learning algorithms efficiently solve computer vision problems

Question 02. Which of the following is a type of neural network?
(A) All of the above
(B) Autoencoders
(C) Capsule Neural Networks
(D) CNN (Convolutional Neural Network)

Question 03. Which of the following is **FALSE** about Neural Networks?
(A) During backward propagation, we update the weights using gradient descent algorithm
(B) We can use different activation functions in different layers
(C) We can use different gradient descent algorithms in different epochs
(D) None of the above

Question 04. Which of the following is **FALSE** about step activation function?
(A) It is linear in nature
(B) It either outputs 0 or 1
(C) None of the above
(D) It is also called Threshold activation function

Question 05. Which of the following is **FALSE** about activation functions?
(A) Activation functions help in achieving non-linearity in deep neural networks
(B) Commonly used activation functions are step, sigmoid, tanh, ReLU and softmax
(C) Activation functions help in reducing overfitting problem
(D) These are also called squashing functions as these squash the output under a certain range

Question 06. Output of step (threshold) activation function ranges from:
(A) 0 to 1
(B) -1 to 1
(C) Either 0 or 1
(D) Either -1 or 1

Question 07. Which of the following is **FALSE** about sigmoid and tanh activation function?
(A) None of the above
(B) Both are non-linear activation functions
(C) Output of both sigmoid and tanh is smooth, continuous and differentiable
(D) Output of sigmoid ranges from -1 to 1 while output of tanh ranges from 0 to 1

Question 08. Which of the following is FALSE about Dropout?
(A) Dropout can be used in input, hidden and output layers
(B) Dropout is a hyper-parameter
(C) None of the above
(D) Dropout is implemented per layer in a network

Question 09. Which of the following is FALSE about Dropout?
(A) Dropout is a learnable parameter in the network
(B) Dropout introduces sparsity in the network
(C) Dropout increases the accuracy and performance of the model
(D) Dropout makes training process noisy

Question 10. Which of the following is TRUE about Dropout?
(A) Dropout can be compared to boosting technique in machine learning
(B) Dropout should be implemented only during training phase, not in testing phase
(C) Dropout is computationally complex as compared to L1 and L2 regularization methods
(D) Dropout should be implemented during training phase as well as during testing phase

Question 11. Which of the following is **TRUE** about local and global minima?
(A) Hyper-parameter tuning plays a vital role in avoiding global minima
(B) All of the above
(C) Ideally, SGD should reach till local minima and should not stuck in global minima
(D) Sometimes local minimas are as good as global minimas

Question 12. Which of the following is a way to avoid local minima?
(A) All of the above
(B) Use momentum and adaptive learning
(C) Increase the learning rate
(D) Add some noise while updating weights

Question 13. Which of the following SGD variants is **NOT** based on adaptive learning?
(A) Adagrad
(B) Nesterov
(C) AdaDelta
(D) RMSprop

Question 14. Which of the following is **TRUE** about Weight Initialization?
(A) If weights are too high, it may lead to vanishing gradient
(B) If weights are too low, it may lead to exploding gradient
(C) All of the above
(D) Model may never converge due to wrong weight initialization

Question 15. Which of the following is **TRUE** about Momentum?
(A) All of the above
(B) It helps in faster convergence
(C) It helps in accelerating SGD in a relevant direction
(D) It helps SGD in avoiding local minima

Question 16. Which of the following is **FALSE** about Pooling Layer in CNN?
(A) It does down-sampling of an image which reduces dimensions by retaining vital information
(B) It does feature extraction and detects components of the image like edges, corners etc.
(C) Output of convolutional layer acts as an input to the pooling layer
(D) Pooling layer must be added after each convolutional layer

Question 17. Which of the following is a valid reason for not using fully connected networks for image recognition?
(A) It creates a lot more parameters for computation as compared to CNN
(B) CNN is far efficient in terms of performance and accuracy for image recognition
(C) All of the above
(D) It may overfit easily as compared to CNN

Question 18. Which of the following is **FALSE** about Padding in CNN?
(A) There are two types of padding: Zero Padding and Valid Padding (no padding)
(B) In zero padding, we pad the image with zeros so that we do not lose any edge information
(C) Padding is used to prevent the loss of information about edges and corners during convolution
(D) There is no reduction in dimension when we use valid padding

Question 19. Which of the following is **FALSE** about Kernels in CNN?
(A) Kernels extract simple features in initial layers and complex features in deeper layers
(B) Kernels can be used in convolutional as well as in pooling layers
(C) Kernels keep sliding over an image to extract different components or patterns of an image
(D) None of the above

Question 20. Which of the following is **NOT** a hyper-parameter in CNN?
(A) Number of convolutional layers
(B) Number and size of kernels in a convolutional layer
(C) Padding in a convolutional layer (zero or valid padding)
(D) Code size for compression

Question 21. Which of the following is **FALSE** about LSTM?
(A) None of the above
(B) LSTM is an extension for RNA which extends its memory
(C) LSTM enables RNN to learn long-term dependencies
(D) LSTM solves the exploding gradients issue in RNN

Question 22. Which of the following is NOT an application of RNN?
(A) Algorithmic trading
(B) Image captioning
(C) Image compression
(D) Understanding DNA sequence

Question 23. Which of the following is NOT an application of RNN?
(A) Anomaly detection
(B) Weather prediction
(C) Stock market prediction
(D) Time series prediction

Question 24. How many parts can the GAN be divided into?
(A) 3
(B) 2
(C) 1
(D) 4

Question 25. Which of the following words is used to familiarize yourself with generative models and to explain how data is generated using probabilistic models.
(A) Generative
(B) Discriminator
(C) Adversarial
(D) Networks

Question 26. Which of the following is not an example of a generative model?
(A) Naive Bayes
(B) Discriminator models
(C) GAN models
(D) PixelRNN/PixelCNN

Question 27. What is the standard form of YOLO?
(A) None of the above
(B) You Once Look Only
(C) YOu Look Once
(D) You Only Look Once

Question 28. Which of the following are the components of object recognition system
(a) Model database,
(b) Hypothesizer,
(c) Feature detector,
(d) Hypothesis verifier
(A) (a) and (c)
(B) (a) and (b)
(C) None of the above
(D) (c) and (d)
(E) All (a), (b), (c) and (d)

Question 29. The face recognition system used in:
(a) Biometric identification
(b) Human and computer interface
(A) (a)
(B) None of the above
(C) Both (a) and (b)
(D) (b)

Question 30. Which of the following is TRUE about types of Vectorization in NLP?
(A) Count vectorization considers count and weightage of each word in a text
(B) All of the above
(C) N-gram vectorization considers context of the word depending upon the value of N
(D) TF-IDF vectorization considers both count and context of the words in the text

Question 31. Which of the following is an application of NLP?

(A) Google Assistant   (B) Chatbots   (C) Google Translate   (D) All of the above

Question 32. Which of the following is TRUE about NLP?

(A) NLP can be used for spam filtering, sentiment analysis and machine translation
(B) All of the above
(C) We must take care of Syntax, Semantics and Pragmatics in NLP
(D) Preprocessing tasks include Tokenization, Stemming, Lemmatization and Vectorization

Question 33. Which of the following techniques can be used to reduce model overfitting?

(a) Data augmentation

(b) Dropout

(c) Batch Normalization

(d) Using Adam instead of SGD

(A) All the choices (a),(b),(c) and (d) are correct   (B) (a) and (b)
(C) (a), (b) and (c)   (D) (a)
(E) (b)

Question 34. Which of the following is true about dropout?

(a) Dropout leads to sparsity in the trained weights

(b) At test time, dropout is applied with inverted keep probability

(c) The larger the keep probability of a layer, the stronger the regularization of the weights in that layer

(A) (c)   (B) All the choices (a),(b) and (c) are correct
(C) (b)   (D) (a)
(E) None of the above

Question 35. You are training a Generative Adversarial Network to generate images of reptiles. But, you think your Generator might be showing mode collapse. Which of the options below could be indicators of this problem?

(a) The generator is only producing images of komodo dragons

(b) The generator loss is oscillating.

(c) The generator loss remains low whereas the discriminator loss is high

(d) The discriminator has high accuracy on real images but low accuracy on fake ones

(A) (a) and (b)   (B) (c)
(C) All the choices (a),(b),(c) and (d) are correct   (D) (c) and (d)
(E) (d)

Question 36. Which one of the following is not a pre-processing technique in NLP?

(A) converting to lowercase   (B) Stemming and Lemmatization
(C) removal of stop words   (D) Sentiment analysis
(E) removing punctuations

Question 37. You are training a GANs to generate images of reptiles. But, you think your Generator might be showing mode collapse. Which of the options below could be indicators of this problem?

(a) The generator is only producing images of komodo dragons.

(b) The generator loss is not stable.

(c) The generator loss remains low whereas the discriminator loss is high.

(d) The discriminator has high accuracy on real images but low accuracy on fake ones.

Question 38. Which of the following propositions are TRUE about a CONV layer? (Check all that apply)

(a) The number of weights depends on the depth of the input volume.

(b) The number of biases is equal to the number of filters.

(c) The total number of parameters depends on the stride.

(d) The total number of parameters depends on the padding.

A (b) and (c)   **B** (c) and (d)   **C** (a) and (b)

D All the choices are correct   **E** (a) and (d)

II. SHORT ANSWER QUESTIONS

Question 39. You have been tasked to build a classifier that takes in an image of a **movie poster** and classifies it into one of four genres: comedy, horror, action, and romance. You have been provided with a large dataset of movie posters where each movie poster corresponds to a move with exactly one of these genres.

Your model now has 100% accuracy on the training set, and 96% accuracy on the validation set! You now decide to expand the model to posters of movies belonging to multiple genres. Now, each poster can have multiple genres associated with it; for example, the poster of a movie like "Lật mặt: Nhà có khách" falls under both "comedy" and "action". Propose a way to label new posters, where each example can simultaneously belong to multiple classes?

To avoid extra work, you decide to retrain a new model with the same architecture (softmax output activation with cross-entropy loss). Explain why this is problematic?

Question 40. Explain the difference between gradient descent, stochastic gradient descent, and mini-batch gradient descent

Question 41. Consider the convolutional neural network defined by the layers in the left column below.

Fill in the shape of the output volume and the number of parameters at each layer. You can write the activation shapes in the format (H, W, C), where H, W, C are the height, width and channel dimensions, respectively. Unless specified, assume padding 1, stride 1 where appropriate.

Notation:

- **CONV***x***-N** denotes a convolutional layer with N filters with height and width equal to x.
- **POOL-***n* denotes a $n \times n$ max-pooling layer with stride of n and 0 padding.
- **FLATTEN** flattens its inputs, identical to `torch.nn.flatten` / `tf.layers.flatten`
- **FC-***N* denotes a fully-connected layer with N neurons

Layer	Activation Volume Dimension	Number of Parameters
Input	$32 \times 32 \times 3$	0
CONV3-8		
Leaky ReLU		
POOL-2		
BATCHNORM		
CONV3-16		
Leaky ReLU		
POOL-2		
FLATTEN		
FC-10		

Question 42. Give a method to fight vanishing gradient in fully-connected neural networks. Assume we are using a network with Sigmoid activations trained using SGD.

Question 43. How do we train the deep network?

Question 44. Explain the difference between the sigmoid and tanh activation function.

Question 45. What is the Jacobian Matrix?

Question 46. Explain Generative Adversarial Network.

Question 47. How LSTM differ from the RNN?

Question 48. What is the difference between the same padding and valid padding?

Question 49. What is IoU?

Question 50. In NLP, how word embedding techniques help to establish the distance between 2 tokens?