KATHMANDU UNIVERSITY
End Semester Examination
January 2024

Marks Scored:

Level : B.E./B.Sc.
Year : IV

26 JAN 2024

Course : COMP 488
Semester : I

Exam Roll No. :

Time: 30 mins.

F. M. : 10

Registration No.:

Date :

SECTION "A"
[20Q. × 0.5 = 10 marks]

Choose and encircle the most appropriate option from each set of choices.

1. Let's say you are working with categorical features, and you have not looked at the distribution of the categorical variable in the test data. You want to apply one-hot encoding on the categorical features. What challenges may you face if you have applied one-hot encoding on a categorical variable of the training dataset?
   a. All categories of the categorical variables are not present in the test dataset.
   b. The frequency distribution of categories is different in the train compared to the test dataset.
   c. Train and Test always have the same distribution.
   d. Both "a" and "b"

2. Imagine you are solving a classification problem with a highly imbalanced class. The majority class is observed in the training data 99% of the time. Your model has 99% accuracy after taking the predictions on the test set. Which of the following is true in such a case?
   I. The accuracy metric is not a good idea for imbalanced class problems.
   II. The accuracy metric is a good idea for imbalanced class problems.
   III. Precision and recall metrics are good for imbalanced class problems.
   IV. Precision and recall metrics aren't good for imbalanced class problems.
   a. I and III b. I and IV c. II and III d. II and IV

3. In linear regression, we try to _____ the least square errors of the model to identify the line of best fit.
   a. Change b. Maximize c. Minimize d. None of the mentioned

4. Which of the following is **TRUE** about unsupervised Machine Learning?
   a. A semi-autonomous Machine Learning where researchers control some parts of the modeling process
   b. Unsupervised learning comprises algorithms with no pre-existing outcomes
   c. A fully autonomous Machine Learning with no human interference
   d. Learning algorithms with no control over the quality of their predictions

5. What is the primary purpose of an activation function in an artificial neural network?
   a. To initialize the weights of the neurons
   b. To compute the gradient during backpropagation
   c. To introduce non-linearity in the model
   d. To determine the learning rate

6. In a feedforward neural network, what is the role of the input layer?
   a. Performs mathematical operations on input data
   b. Passes the input data to the subsequent layers
   c. Contains the neurons responsible for making predictions
   d. Normalizes the input data

7. Let's say that you are using activation function X in hidden layers of a neural network. At a particular neuron for any given input, you get the output as "-0.0001". Which of the following activation function could X represent?
   a. ReLU b. tanh c. SIGMOID d. None of mentioned

8. Which type of regularization technique penalizes the magnitude of weights in a neural network to prevent large weight values?
   a. L_1 regularization b. L_2 regularization
   c. Dropout regularization d. Batch normalization

9. Which regularization technique combines both L_1 and L_2 penalties to the loss function?
   a. Dropout regularization b. Elastic Net regularization
   c. Batch normalization d. Gradient Clipping

10. What is a common technique used to prevent overfitting in deep neural networks?
    a. Increasing the learning rate
    b. Reducing the number of neurons in each layer
    c. Adding more hidden layers
    d. Using dropout layers during training

11. In a convolutional neural network (CNN), what is the primary advantage of using convolutional layers?
    a. They reduce the number of parameters in the network
    b. They perform element-wise matrix multiplication
    c. They enable the network to learn spatial hierarchies of features
    d. They replace the need for fully connected layers

12. What is the primary advantage of using pre-trained CNN models?
    a. Faster training times b. Smaller model size
    c. Improved performance on new tasks d. Reduced memory usage

13. Which type of RNN is ideal for tasks like machine translation and named entity recognition, where sequential inputs are dependent on each other and context is crucial?
    a. One to One b. One to Many c. Many to One d. Many to Many

14. In the architecture of an LSTM cell, which component is responsible for selectively updating its memory and controlling the flow of information?
    a. Memory cell b. Input Gate c. Output Gate d. Forget Gate

15. In the context of reinforcement learning, what is the "reward signal"?
    a. A signal indicating the start of the episode
    b. A signal indicating the end of the episode
    c. A signal indicating the quality of an agent's action in an environment
    d. A signal indicating the learning rate

16. Which type of neural network architecture is commonly used for generative tasks, such as image generation or text generation?
    a. Convolutional Neural Network (CNN)
    b. Recurrent Neural Network (RNN)
    c. Autoencoder
    d. Feedforward Neural Network (FNN)

17. In the context of transformers, what does self-attention refer to?
    a. Paying attention to oneself
    b. Focusing on the most important words in a sentence
    c. Learning context in sequential data
    d. Avoid overfitting

18. What is the purpose of the multi-head attention in the transformer architecture?
    a. Increasing the number of layers
    b. Reducing computational complexity
    c. Focusing on different aspects of the input sentence
    d. Eliminating positional encoding

19. What is one of the best practices for optimizing transformer model performance?
    a. Using a very small batch size
    b. Ignoring positional encoding
    c. Avoiding dropout layers
    d. Employing early stopping

20. Which transformer-based model was developed by Google and is popular for general question-answering tasks?
    a. GPT-3
    b. BERT
    c. BART
    d. RoBERTa

KATHMANDU UNIVERSITY
End Semester Examination
January 2024

Level : B.E./B.Sc.
Year : IV
Time : 2 hrs. 30mins.

26 JAN 2024

Course : COMP 488
Semester : I
F. M. : 40

SECTION "B"
[6 Q. × 4 = 24 marks]

Attempt *ANY SIX* questions.

1. Compare and contrast "Machine Learning" and "Deep Learning".
2. Explain the Bayes theorem and the associated parameters in the theorem.
3. Explain the Gradient Descent algorithm and the "Local minima" and "Global Minima" concepts.
4. Define the terms "Capacity", "Overfitting" and "Underfitting" for a Machine Learning model.
5. What is regularization? Briefly discuss "early stopping" and "dropout" as regularization strategies.
6. What are Convolutional Neural Networks? Briefly explain the three stages of a typical convolutional layer.
7. Discuss the three design patterns of Recurrent Neural Networks (RNNs).

SECTION "C"
[2Q. × 8 =16 marks]

Attempt *ANY TWO* questions.

8. Consider the following dataset.

S.No.	Age	Competition	Type	Profit
1	Old	Yes	Software	Down
2	Old	No	Software	Down
3	Old	No	Hardware	Down
4	Mid	Yes	Software	Down
5	Mid	Yes	Hardware	Down
6	Mid	No	Hardware	Up
7	Mid	No	Software	Up
8	New	Yes	Software	Up
9	New	No	Hardware	Up
10	New	No	Software	Up

a. Calculate the entropy of the target variable "Profit". [2]
b. Calculate the information gain of the attributes – "Age", "Competition" and "Type". [4]
c. Construct a decision tree based on the calculations of the information gain above. [2]

9. Refer to the sample weather dataset of Table 1:

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Table 1. Sample Weather Dataset

Compute the probabilities of Play Golf (Yes or No) using the Naïve Bayes rule for the weather conditions – "Outlook = Rainy", "Temp = Cool", "Humidity=High" and "Windy= True".

10. Discuss the Transformer model and the components involved in the architecture. What makes it more robust to handle different Natural Language Processing tasks more efficiently compared to the sequence-to-sequence models? [4+4]