

20233_csci_544_30249:
Applied Natural
Language Processing

Announcements

Syllabus

Content

Assignments

Tools

USC Code of Ethics

Review Test Submission: Final Exam

User	Kayvan Shah
Course	20233_csci_544_30249: Applied Natural Language Processing
Test	Final Exam
Started	11/30/23 5:30 PM
Submitted	11/30/23 6:37 PM
Due Date	11/30/23 6:46 PM
Status	Completed
Attempt Score	452.99666 out of 600 points
Time Elapsed	1 hour, 6 minutes out of 1 hour and 10 minutes

Results Displayed All Answers, Submitted Answers, Correct Answers, Incorrectly Answered Questions

Question 1

10 out of 10 points



Consider the next-word prediction task with an RNN architecture. Assume that at time-step t, the model has predicted the following scores for a vocab of 3 words:

apple: -1.60

bag: -1.60

tree: -2.31

If the ground truth label is "tree", what would be the cross-entropy loss at this time step? (Use a calculator to write your answer in numerical form, rounded to 3 decimal places. When needed, use e is the base of logarithm)

Selected Answer: 1.63

Correct Answer: 1.622

Answer range +/- 0.111 (1.511 - 1.733)

Question 2

8 out of 10 points



Which are part of the Levenshtein edit distance for fixing spelling errors:

- Selected Answers: Transposition of two adjacent letters
 Deletion
 Substitution
 Insertion

- Answers:
separating vowels and consonants
 insertion of space or hyphen
 Transposition of two adjacent letters
 Deletion
 Substitution
 Insertion

Question 3

0 out of 10 points



In a trigram model, the probabilities are $P(\text{brown}|\text{the, quick}) = 0.5$, $P(\text{fox}|\text{the, brown}) = 0.4$, and $P(\text{fox}|\text{quick, brown}) = 0.3$. What is the probability of the sentence "the quick brown fox" according to this model?

Selected Answer: 0.06

Correct Answer: 0.15

Answer range +/- 0.001 (0.149 - 0.151)

Question 4

0 out of 10 points



Calculate the BLEU score given these sentences:

Reference Sentence: "Birds fly over mountains"

Predicted Sentence: "Birds fly over hills"

Selected Answer: 1.204

Correct Answer: 0

Answer range +/- 0 (0 - 0)

Question 5

10 out of 10 points



A standard Transformer architecture is sensitive to the words reordering.

Selected Answer: True
Answers: True
 False

Question 6

10 out of 10 points



Which type of spelling error best describes the following?

"bred" instead of "bread"
 "chocolate" instead of "chocolate"
 "Experienc" instead of "experience"

Selected Answer: Non-word errors
Answers: Non-word errors
 Typographical errors
 Semantic Error
 Information error

Question 7

0 out of 10 points



Consider a typical encoder-decoder Transformer architecture for the Machine Translation task. Select all that are correct.
Selected Answers:
 Each decoder layer benefits from a self-attention to the output of the previous decoder layer in addition to a cross-attention that attends to its counterpart encoder layer at the same level.
 During test time, we can avoid sequential generation by parallelizing the decoder to achieve speed boost.
Answers:
 Each decoder layer benefits from a self-attention to the output of the previous decoder layer in addition to a cross-attention that attends to its counterpart encoder layer at the same level.
 In contrast to the encoder, the self-attention in the decoder only attends to the left words rather than the whole sequence.
 Transformers can capture longer dependencies with a quadratic cost of computation.
 During test time, we can avoid sequential generation by parallelizing the decoder to achieve speed boost.

Question 8

0 out of 10 points



For every training sample (x,y) , given loss, L , between ground truth, y , and estimated output, y' , for every weight from hidden layer to output layer, we should update the weight using gradient descent by $\frac{d}{dw}L(f(x;w), y')$, where f is the model itself.

Selected Answer: True
Answers: True
 False

Question 9

5 out of 10 points



Which of the following **is NOT** a capability of TensorBoard? (Select all that applies)

Selected Answers:
 Conducting A/B testing on model variants for performance comparison.
 Image classification using deep learning
 Inspecting and analyzing training data input pipelines.
 Audio processing for neural networks
Answers:
 Visualization of machine learning model weights
 Conducting A/B testing on model variants for performance comparison.
 Image classification using deep learning
 Profiling and debugging of TensorFlow models
 Inspecting and analyzing training data input pipelines.
 Profiling model performance to identify bottlenecks.
 Audio processing for neural networks
 Monitoring of training loss

Question 10

10 out of 10 points



Which of the following post-processing actions is commonly performed in phrase-based models to improve the quality of the translated output?

Selected Answer:
 Applying a language model to reorder phrases into a more natural-sounding target language sentence structure.
Answers:
 Re-running the translation model to refine translation choices based on a monolingual target language model.
 Applying a language model to reorder phrases into a more natural-sounding target language sentence structure.
 Utilizing a bilingual dictionary to replace out-of-vocabulary words with their most frequent translations.

Question 11

10 out of 10 points



In a phrase-based translation model, what is the primary purpose of the translation phase?

- Selected Answer: To generate the target language sentence from the aligned source language phrases
 Answers: To perform word alignment between the source and target languages
 To generate the target language sentence from the aligned source language phrases
 To handle post-processing tasks like capitalization and punctuation
 To tokenize the input text into individual words

Question 12

0 out of 10 points

During the training of a deep neural network, you observe that the loss is slowly dropping, and the trend continues through the end of training.
What are possible solutions to this issue?

- Selected Answers: Reduce learning rate
 Increase the number of epochs
 Answers: Increase learning rate
 Reduce learning rate
 Increase the number of epochs
 Reduce the number of epochs

Question 13

10 out of 10 points



Which of the following statements is true about Autograd in PyTorch?

- Selected Answer: It's a powerful feature that automatically calculates the gradients of tensors, making gradient-based optimization possible.
 Answers: It's a deep learning framework for reinforcement learning.
 It's a module for generating random numbers.
 It's a powerful feature that automatically calculates the gradients of tensors, making gradient-based optimization possible.
 It's a method for plotting tensors directly using Matplotlib.
 It's a function that enables saving tensors directly to disk in the JPEG image format.

Question 14

5 out of 10 points



What are the pros of vanilla transformer?

- Selected Answers: Easy to capture dependencies
 Easy to parallelize
 Answers: Computational efficiency at all time
 Easy to capture dependencies
 Easy to model positional information
 Easy to parallelize

Question 15

10 out of 10 points



What is the primary goal of the Expectation-Maximization (EM) algorithm?

- Selected Answer: To maximize the likelihood of observed data given the model parameters.
 Answers: To maximize the likelihood of observed data given the model parameters.
 To minimize the number of iterations required for training.
 To eliminate the need for labeled training data.
 To perform dimensionality reduction on text data.

Question 16

10 out of 10 points



What is the role of an 'epoch' in PyTorch model training?

- Selected Answers: A single pass through the entire training dataset.
 Answers: The number of times the learning rate is adjusted.
 The total number of training steps.
 A single pass through the entire training dataset.
 The process of saving model state.

Question 17

10 out of 10 points



Which of the following statements are true about the variant of RNN?

- Selected Answers: Bidirectional RNN applies two RNN that process the input in opposite direction.
 RNN can process natural language in unidirectional, but natural language itself has the property of bi-directionality.
 Gated RNN is good at remembering long-term dependency than vanilla RNN.
- Answers: Bidirectional RNN applies two RNN that process the input in opposite direction.
 RNN can process natural language in unidirectional, but natural language itself has the property of bi-directionality.
In gated RNN, reset and update gate are always close to 0 or 1 due to the use of tanh activation function.
 Gated RNN is good at remembering long-term dependency than vanilla RNN.

Question 18

5 out of 10 points



Select the correct answers regarding the Automatic Metrics for evaluating MT text generation tasks:

- Selected Answers: Metrics consider the relevance of words.
 Metrics operate on local levels.
 Scores are meaningless unless compared in similar contexts and testing datasets
- Answers: Metrics consider the relevance of words.
Human translators score high on BLEU.
 Metrics operate on local levels.
 Scores are meaningless unless compared in similar contexts and testing datasets

Question 19

5 out of 10 points



Which of the following operations **cannot** be performed on PyTorch tensors?
Select all that apply:

- Selected Answers: Using tensors as input to machine learning algorithms in scikit-learn
 Training neural networks on GPUs without CUDA support
- Answers: Backpropagation
 Directly applying traditional Python functions like a filter to tensors
Arithmetic operations (addition, subtraction, multiplication, division)
 Using tensors as input to machine learning algorithms in scikit-learn
Element-wise operations (max, min, mean, sum)
Building complex neural network architectures
Linear algebra operations (matrix multiplication, matrix inversion)
 Training neural networks on GPUs without CUDA support

Question 20

10 out of 10 points



What is the primary advantage of IBM Model 2 over IBM Model 1 in the context of machine translation?

- Selected Answer: Model 2 can handle linguistic phenomena like word reordering more effectively.
- Answers: Model 2 is faster in terms of translation speed.
 Model 2 can handle linguistic phenomena like word reordering more effectively.
Model 2 requires fewer computational resources for training.
Model 2 eliminates the need for human-annotated alignment data.

Question 21

6.66 out of 10 points



In PyTorch, what are key practices for model evaluation and debugging?

- Selected Answers: Using `torch.no_grad()` during model evaluation to reduce memory usage.
 Applying `model.eval()` mode during evaluation for correct behavior of certain layers like dropout.
- Answers: Using `torch.no_grad()` during model evaluation to reduce memory usage.
 Enable random shuffle of test dataloader.
Augment the test data in the same way the training data is augmented
 Applying `model.eval()` mode during evaluation for correct behavior of certain layers like dropout.

Question 22

10 out of 10 points



Which of the following statements accurately do not describe the limitations or challenges associated with the EM algorithm?

- Selected Answers: The EM algorithm is limited to linear models and cannot handle the complexity of models with non-linear relationships between variables.
 The EM algorithm is limited to linear models and cannot handle the complexity of models with non-linear relationships between variables.
- Answers: The EM algorithm is limited to linear models and cannot handle the complexity of models with non-linear relationships between variables.

EM can get stuck in local optima, leading to suboptimal results.
The algorithm's performance heavily depends on the choice of initial parameter values.
EM assumes that the data follows a specific parametric distribution.

Question 23

10 out of 10 points



In language modeling, what is a primary application of 3-grams?

Selected Answer: Next word prediction.

Answers: Next word prediction.

Text summarization.

Speech-to-text conversion.

Text-to-Speech conversion.

Question 24

10 out of 10 points



Which of the following statements are false regarding the application of the EM algorithm in NLP?

Selected Answers: The EM algorithm is particularly useful for supervised learning tasks where all labels are known.

Answers: The EM algorithm assumes that the underlying data distribution is non-parametric and does not require initial parameter estimates.

The EM algorithm is used to estimate the parameters of probabilistic models like Hidden Markov Models, which are applied in tasks such as part-of-speech tagging.

In text clustering, the EM algorithm can be applied to soft-clustering methods where data points can belong to multiple clusters

The EM algorithm is particularly useful for supervised learning tasks where all labels are known.

The EM algorithm assumes that the underlying data distribution is non-parametric and does not require initial parameter estimates.

Question 25

10 out of 10 points



Which of the following(s) is/are true about stochastic gradient descent?

Selected Answers: Gradient descent is performed for every batch.

It can help to reduce computational load compared to vanilla gradient descent

The random batches are created from the training data

Answers: Gradient descent is performed for every batch.

It can help to reduce computational load compared to vanilla gradient descent

The random batches are created from the training data

It is guaranteed to converge to a global solution.

Question 26

10 out of 10 points



Consider the following code written in PyTorch for NER tagging (similar to HW4):

```
class NeuralNet(nn.Module):
    def __init__(self, vocab_size, embedding_dim, hidden_dim, num_classes):
        super(NeuralNet, self).__init__()

        self.fc2 = nn.Linear(d, e)
        self.fc1 = nn.Linear(hidden_dim, hidden_dim)
        self.lstm = nn.LSTM(b, c, batch_first=True)
        self.embedding = nn.Embedding(a, embedding_dim)

    def forward(self, x):
        x = self.embedding(x)
        x, _ = self.lstm(x)
        x = x[:, -1, :]
        x = F.relu(self.fc1(x))
        x = self.fc2(x)
        return x
```

Determine the values of a, b, c, d, and e.

Question Correct Match

a 3. vocab_size

Selected Match

3. vocab_size

b > embedding dim

> embedding dim

- c 4. hidden_dim 4. hidden_dim
- d 4. hidden_dim 4. hidden_dim
- e 1. num_classes 1. num_classes

All Answer Choices

- 1. num_classes
- 2. embedding_dim
- 3. vocab_size
- 4. hidden_dim

Question 27

10 out of 10 points



What is/are the pre-training task(s) for BERT?

- Selected Answer: Masked Language Modeling, Next Sentence Prediction
- Answers:
- Masked Langauge Modeling, Next Word Prediction
 - Masked Lanuage Modeling, Token Classification
 - Token Classification, Next Sentence Prediction
 - Masked Language Modeling, Next Sentence Prediction

Question 28

0 out of 10 points



Select all that is correct:

- Selected Answers: Both RNN and Transformer architectures can theoretically process arbitrary-length input sequences.
- Answers:
- In a Seq2Seq RNN architecture, Bi-LSTM architecture can be used in the decoder to produce higher-quality outputs.
 - Both RNN and Transformer architectures can theoretically process arbitrary-length input sequences.
 - It is computationally harder to model positional information in RNNs compared to Transformers.
 - In a Seq2Seq RNN architecture, Bi-LSTM architecture can be used in the decoder to produce higher-quality outputs.
 - In an encoder-decoderTransformer architecture for translation, the decoder takes the target language sentence as input during training.

Question 29

10 out of 10 points



In the context of deep learning model training, which of the following techniques are commonly used to improve model robustness and generalization?

- Selected Answers: Randomly initializing network weights.
 Applying random dropout during training.
 Implementing data augmentation techniques.
- Answers:
- Randomly initializing network weights.
 - Applying random dropout during training.
 - Implementing data augmentation techniques.
 - Using a fixed learning rate throughout training.

Question 30

10 out of 10 points



In machine learning, which techniques are effective in preventing overfitting in a model?

- Selected Answers: Using cross-validation.
 Gathering more training data.
 Applying L1 or L2 regularization.
- Answers:
- Using cross-validation.
 - Gathering more training data.
 - Applying L1 or L2 regularization.
 - Increasing the model complexity.

Question 31

0 out of 10 points



What is the Levenshtein edit distance between the words "dexperience" and "experience" when using the 4 possible operators in Levenshtein edit operators? Select the correct option:

- Selected Answer: 3
- Answers:
- 3
 - 1
 - 2

Question 32

10 out of 10 points



What is the main disadvantage of statistical MT compared to the rule base method?

- Selected Answer: Dependency on extensive parallel data
- Answers:
- Elevated evaluation costs
 - Diminished quality of outcomes
 - Dependency on extensive parallel data
 - Substantial need for direct human supervision

Question 33

10 out of 10 points



Consider a batch of B sentences where each sentence has a length of T, and each word is represented by an embedding vector with size D. We use the following piece of code to define a BiLSTM model in PyTorch and apply it to the input data:

```
lstm_model = torch.nn.LSTM(input_size=D, hidden_size=H, batch_first=True, bidirectional=True)
output, _ = lstm_model(input)
```

Which of the following choices correctly shows the expected size of **input** and **output** tensors?

- Selected Answer: input = [B, T, D]
 output = [B, T, 2H]
- Answers:
- input = [T, B, D]
 output = [T, B, H]
 - input = [B, T, 2D]
 output = [B, T, H]
 - input = [T, B, H]
 output = [T, B, D]
 - input = [B, T, D]
 output = [B, T, 2H]

Question 34

10 out of 10 points



In a corpus with 100,000 words, a bigram "new york" occurs 300 times, "new" occurs 1,000 times, and "york" occurs 800 times. What is the conditional probability P(york|new) in this bigram model?

- Selected Answer: 0.3
 Correct Answer: 0.3
 Answer range +/- 0.01 (0.29 - 0.31)

Question 35

10 out of 10 points



What is $\frac{dReLU(z)}{dz}$ when z = 10?

- Selected Answer: 1
 Answers:
- 10
 - 0
 - 10
 - 1

Question 36

0 out of 10 points



In pytorch, which processes have to be explicitly defined?

- Selected Answers: Model architecture
 Forward pass
 Backward pass
 Gradient calculation
 Answers:
- Model architecture
 - Forward pass
 - Backward pass
 - Gradient calculation

Question 37

10 out of 10 points



In the context of LSTM architecture, what role does the New Gate play?

- Selected Answer: It computes the new candidate cell state based on the current input and the previous hidden state.
 Answers: It computes the new candidate cell state based on the current input and the previous hidden state.

It decides whether to update the cell state with new information.

It determines the amount of information to be discarded from the cell state.

It calculates the new hidden state based on the current input and the previous hidden state.

Question 38

6.67 out of 10 points



What is/are the advantage of ViT over CNN?

- Selected Answers: interpretability of the model
 It can process images with different sizes
 Answers: interpretability of the model
 Computational efficiency during training time
 It can process images with different sizes
 short inference time

Question 39

6.66666 out of 10 points



Select all that are part of spelling errors

- Selected Answers: Non-word Errors
 Typographical errors
Answers: Non-word Errors
 Typographical errors
 Cognitive Errors
Semantic Error
Information error

Question 40

10 out of 10 points



Consider the following reference sentence for a machine translation task:

"The Sky Is Not Blue"

Which of the following hypothesis sentences has the lowest word error rate (WER)?

Selected Answer: The Sky Is Blue

Answers: The Sky

The Sky Is Blue

No Sky Is Blue

Blue Not Is Sky The

Question 41

0 out of 10 points



Assume a batch of k sentences with the length of T words where each word is a vector with size h . The computational cost of applying a single self-attention layer on this batch would be $O(k^x T^y h^z)$. Find x, y, z and calculate $(x+y) \times z$.

Selected Answer: 8

Correct Answer: 3

Answer range +/- 0 (3 - 3)

Question 42

10 out of 10 points



Select all that are incorrect about Vanilla RNNs:

- Selected
Answers: Unlike Fully Connected networks, they are invulnerable to optimization challenges such as gradient explosion or vanishing.
 RNNs can be parallelized in the sequence length dimension but are sequential in the batch direction.
Answers: Unlike Fully Connected networks, they are invulnerable to optimization challenges such as gradient explosion or vanishing.
These models are sensitive to the reordering of the words in the input sentence.
Theoretically, they can process arbitrary-length input sentences by design.
 RNNs can be parallelized in the sequence length dimension but are sequential in the batch direction.

Question 43

10 out of 10 points



Which of the following is not the way to improve the efficiency of transformer for long sequence input?

Selected Answer: context the current word from both sides of the sequence.

Answers: process the input in segments and then aggregate.

decouple the attention matrix

context the current word from both sides of the sequence.

Question 44

10 out of 10 points



Which of the following statements is true regarding IBM Model 1?

Selected Answer: It uses the expectation-maximization (EM) algorithm to estimate the word alignment probabilities.

Answers: It is a neural machine translation model.

It can handle bidirectional translation from source to target language and vice versa.

Selected Answer: It uses the expectation-maximization (EM) algorithm to estimate the word alignment probabilities.

Answers: It focuses on generating high-quality translations of entire sentences in one step.

Question 45

10 out of 10 points



What is(are) the advantage(s) of GPU training in Pytorch?

Selected Answers: Faster matrix operations Efficient parallel processing

Answers: Use less memory

 Faster matrix operations Efficient parallel processing

Larger available system memory

Question 46

10 out of 10 points



GPT-2 explicitly shows the capacity to do zero-shot learning.

Selected Answer: TrueAnswers: True

False

Question 47

10 out of 10 points



In a PLM, what does perplexity measure?

Selected Answer: The complexity and uncertainty of the model's predictions

Answers: Accuracy of the model's predictions according to grammar

The model's ability in lexical disambiguation

Selected Answer: The complexity and uncertainty of the model's predictions

The speed of the model's computations

Question 48

10 out of 10 points



Inflexibility to new phrases, lack of context understanding, and flexibility to variable length sequences are three attributes of IBM MT models.

Selected Answer: False

Answers: True

 False**Question 49**

5 out of 10 points



Which of the following challenges does the EM algorithm help address?

Selected Answers: Dealing with partially observed data.

Answers: Handling high-dimensional feature spaces in text data.

 Dealing with partially observed data. Solving for a local maximum likelihood parameters of a statistical model

Reducing computational complexity in neural network training.

Solving sequence labeling tasks such as part-of-speech tagging.

Question 50

10 out of 10 points



In PyTorch, which of the following is(are) CORRECT about 'batch size'?

Selected Answers: It determines the number of samples processed before updating model weights. It determines how many samples are loaded in a forward pass.Answers: It determines the number of samples processed before updating model weights. It determines how many samples are loaded in a forward pass.

Larger batch size reduces memory usage

A smaller batch size makes optimization smoother

Question 51

10 out of 10 points



In IBM Model 1 for machine translation, which one does the model aim to learn from the training data?

- Selected Answer: Word alignments between source and target language words.
- Answers: Word alignments between source and target language words.
Linguistic syntax and grammar rules.
Sentence-level translation probabilities.
Semantic representations of words.

Question 52

10 out of 10 points



In IBM Model 1 for machine translation, which of the following accurately describes key aspects of the model?

- Selected Answer: The model is trained using parallel bilingual corpora.
- Answers: It assumes a one-to-one alignment between words in the source and target languages.
 The model is trained using parallel bilingual corpora.
It takes into account the word order of the source and target sentences during the alignment process.
It was developed primarily for speech recognition tasks.

Question 53

10 out of 10 points



Given the task to reconstruct an image of bird with noise back to the un-noised version, which of the following models is most preferred?

- Selected Answer: BART
- Answers: GPT-2
RoBERTa
BERT
 BART

Question 54

10 out of 10 points



In the context of spell-checking, what is the primary purpose of the Noisy Channel Model? Select the correct option:

- Selected Answer: To identify potentially misspelled words and suggest corrections based on the probability of generating the observed misspelling from the intended word.
- Answers: To predict the likelihood of a word being misspelled based on its frequency in the language.
To measure the accuracy of spell-checking algorithms in a controlled laboratory environment.
To compare the effectiveness of various spell-checking software programs.
 To identify potentially misspelled words and suggest corrections based on the probability of generating the observed misspelling from the intended word.

Question 55

10 out of 10 points



Pay attention to the translation of the Spanish word "realizar" in these two translations below. Which one of the challenges in Machine Translation (MT) best describes the issue with the word?

- Spanish: "Quiero realizar mi sueño."
English Translation 1: "I want to carry out my dream."
English Translation 2: "I want to realize my dream."

- Selected Answer: Lexical ambiguity
- Answers: Word order mismatch
Different Syntactic Structure
 Lexical ambiguity
Pronoun Resolution
Syntactic ambiguity

Question 56

10 out of 10 points



In IBM Model 2 for machine translation, which of the following statements accurately describes a key improvement over IBM Model 1?

- Selected Answer: Model 2 introduces bidirectional word alignment maps, allowing for more flexible alignments.

- Answers:
- Model 2 introduces bidirectional word alignment maps, allowing for more flexible alignments.
 - Model 2 employs deep neural networks for translation, enhancing accuracy.
 - Model 2 eliminates the need for parallel bilingual corpora during training.
 - Model 2 focuses exclusively on handling long-distance syntactic transformations.

Question 57

10 out of 10 points



What is a key advantage of using a phrase-based translation model over a traditional statistical machine translation (SMT) model?

- Selected Answer: Phrase-based models can better handle word reordering and long-distance dependencies.
- Answers:
- Phrase-based models require less computational resources.
 - Phrase-based models are more interpretable and explainable.
 - Phrase-based models can better handle word reordering and long-distance dependencies.
 - Phrase-based models are less affected by the quality of the training data.

Question 58

0 out of 10 points



Select all that are **NOT** part of the forward pass in designing neural networks with Pytorch.

- Selected Answers: Compute the accuracy of the model
 Compute loss value
 Computer output of the network
- Answers:
- Compute the accuracy of the model
 - Computing gradients
 - Compute loss value
 - Updating parameters
 - Computer output of the network

Question 59

10 out of 10 points



Which of the following statements are true about RNN?

- Selected Answers: Every path from weight, W, to the loss, L, is one dependency for differentiation during backpropagation.
 For step t, it takes both input feature at time t, x_t , and state at time t, h_t , as input.
 The weight matrices of RNN are shared cross the time.
- Answers:
- RNN doesn't suffer from gradient vanishing and exploding.
 - Every path from weight, W, to the loss, L, is one dependency for differentiation during backpropagation.
 - For step t, it takes both input feature at time t, x_t , and state at time t, h_t , as input.
 - The weight matrices of RNN are shared cross the time.

Question 60

0 out of 10 points



For Universal Approximation Theorem, it states that MLPs can represent a wide range of functions given proper values for the weights, and the value for the weights can be automatically provided.

- Selected Answer: True
Answers:
 True
 False

Saturday, January 20, 2024 11:45:04 PM PST

← OK