20233_csci_544_30249: Applied Natural Language Processing

Announcements
Syllabus
Content
Assignments

Tools
USC Code of Ethics

Review Test Submission: Quiz 5

User	Kayvan Shah
Course	20233_csci_544_30249: Applied Natural Language Processing
Test	Quiz 5
Started	10/19/23 5:35 PM
Submitted	10/19/23 5:44 PM
Due Date	10/19/23 5:50 PM
Status	Completed
Attempt Score	90 out of 100 points
Time Elapsed	9 minutes out of 10 minutes
Results Displayed	All Answers, Submitted Answers, Correct Answers

Question 1

10 out of 10 points

the dog	0.2
the ball	0.3
dog chased	0.25
dog found	0.4
ball rolled	0.45
ball bounced	0.3
chased the	0.4
chased a	0.35
found the	0.3
found a	0.2

Given the partial sentence "The dog chased", which of the following is the most probable next word according to the bigram probabilities provided in the table?

Selected Answer: ✓ the

Answers:
ball
✓ the
a
dog

Question 2

10 out of 10 points

What will happen during a forward pass during the training of a neural network in Pytorch?(Select all that apply.)

Selected Answers:
✓ Input data is propagated through the layers of the neural network to generate predictions.
✓ Intermediate activations and results from each layer are stored.

Answers:
✓ Input data is propagated through the layers of the neural network to generate predictions.
Gradients for model parameters are computed.
✓ Intermediate activations and results from each layer are stored.
Model parameters are updated

Question 3

10 out of 10 points

In Pytorch, where is batching usually is coded to be performed?

Selected Answer: ✓ Dataloader

Answers:
Dataset
✓ Dataloader
Forward pass
Backward pass

Question 4

10 out of 10 points

Which of the following statements is true regarding Probabilistic Language Models (PLMs)?

Selected Answer: ✓ In the context of PLMs, $P(w_n|w_{1:n-1}) \times P(w_{n-1}|w_{1:n-2}) \times \ldots \times P(w_2|w_1) \times P(w_1)$ represents the

probability of the entire sentence.

Answers:

A PLM assigns a probability to a given sentence, and it does not utilize the sequence of words in the sentence.

For a given sentence "I like USC NLP", the probability can be calculated as P(NLP|I,like,USC).

A PLM aims to estimate the probability of an upcoming word without considering previous words.

In the context of PLMs, P(w_n|w_1,...,w_n−1) × P(w_n-1|w_1,...,w_n−2) × ... × P(w_2|w_1) × P(w_1) represents the probability of the entire sentence.

Question 5

10 out of 10 points

In PyTorch, to perform a backward pass for gradient computation during neural network training, which of the following components must be retained?

Selected Answer: Intermediate activations and computations from the forward pass.

Answers:

The loss value computed on the validation set.

Intermediate activations and computations from the forward pass.

The data used during the epoch.

The initial network's weights at the beginning of training.

Question 6

10 out of 10 points

Which TWO of the following tasks that N-gram models relatively CANNOT help to perform well?

Selected Answers: Article Summarization
Document retrieval

Answers:

Article Summarization

Document retrieval

Spell Correction

Auto-complete

Question 7

10 out of 10 points

Which of the following metric can be used for the evaluation of a typical NLP task like named entity erecognition, text classification, etc where we have a "gold standard" reference?

Selected Answer: F1 score

Answers:

F1 score

BLEU score

Perplexity

Mean squared error

Question 8

10 out of 10 points

When training a neural network with number of Epochs = 10, mini-batch size = 64, size of training dataset = 10000. How many instances**(repetitions also count)** of training samples will the model see during training?

Selected Answer: 100000

Answers:

100000

1570

10000

640000

Question 9

0 out of 10 points

Given the corpus:

USC is in Los Angeles.
Los Angeles is where USC is.
Many students choose USC because it's in Los Angeles.

Compute the bigram probability for P(Los|Angeles)

(Round to two decimal places, e.g.: 0.12)

Selected Answer: 0.5
Correct Answer: 0
Answer range +/- 0.01 (-0.01 - 0.01)

Question 10

10 out of 10 points

Which of the following statements is TRUE?

Selected Answer: Smoothing is needed when the test dataset is not similar to the training dataset

Answers:

Add-1 smoothing guarantees that every unseen word will be assigned a probability of 1.

Add-1 smoothing adds 1 count to unseen words only.

Sparse datasets always contain many repeated instances of the same word.

Smoothing is needed when the test dataset is not similar to the training dataset

Saturday, January 20, 2024 11:40:17 PM PST

← OK