

20233_csci_544_30249: Applied Natural Language Processing

Announcements

Syllabus

Content

Assignments

Tools

USC Code of Ethics

Review Test Submission: Quiz 3

User	Kayvan Shah
Course	20233_csci_544_30249: Applied Natural Language Processing
Test	Quiz 3
Started	9/12/23 5:35 PM
Submitted	9/12/23 5:45 PM
Due Date	9/12/23 5:50 PM
Status	Completed
Attempt Score	26.667 out of 100 points
Time Elapsed	9 minutes out of 10 minutes
Results Displayed All Answers, Submitted Answers, Correct Answers	

Question 1

0 out of 10 points

Select the correct answer(s) regarding the machine learning algorithms:

- Selected Answers: ☒ CRF is a generative approach to sequence labeling.
☒ Logistic regression is a discriminative approach to sequence labeling.
☒ Naive Bayes, is a generative approach to classification.
☒ MEMM is a discriminative approach to sequence labeling.
- Answers: ☐ CRF is a generative approach to sequence labeling.
☐ Logistic regression is a discriminative approach to sequence labeling.
☒ Naive Bayes, is a generative approach to classification.
☒ MEMM is a discriminative approach to sequence labeling.

Question 2

0 out of 10 points

Consider this SVD decomposition for the matrix $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$ as:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & \frac{-1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & \frac{-2}{3} & \frac{-1}{3} \end{bmatrix}$$

What would be the new representation of A after keeping only the most important singular value?

Selected Answer: ☒ $\begin{bmatrix} 10 \\ 3 \end{bmatrix}$

- Answers: ☒ $\begin{bmatrix} 10 \\ 3 \end{bmatrix}$
☐ $\begin{bmatrix} \frac{5}{2} & \frac{-3}{2} & 0 \\ \frac{-3}{2} & \frac{5}{2} & 0 \end{bmatrix}$
☒ $\begin{bmatrix} \frac{5}{2} & \frac{5}{2} & 0 \\ \frac{5}{2} & \frac{5}{2} & 0 \end{bmatrix}$
☐ $\begin{bmatrix} \frac{3}{2} & \frac{3}{2} & 0 \\ \frac{3}{2} & \frac{3}{2} & 0 \end{bmatrix}$

Question 3

10 out of 10 points

Select the correct answer(s):

Selected Answers: ☒ LSA is a **term weighting-based** model and considers occurrences of terms at the **document level**.☒ GloVe is a **count-based** model that considers occurrences of terms at the **context level**.Answers: ☒ Word2Vec is a **prediction-based** model that considers occurrences of terms at the **document level**.☒ LSA is a **term weighting-based** model and considers occurrences of terms at the **document level**.☒ GloVe is a **count-based** model that considers occurrences of terms at the **context level**.

Question 4

6.667 out of 10 points

Select all that apply to the Word2Vec:

Selected Answers: ☒ Word2Vec uses a two-layer neural network architecture.☒ To train a Word2Vec model, explicit supervision from a human annotator is needed to label the dataset.Answers: ☒ Word2Vec uses a two-layer neural network architecture.☐ To train a Word2Vec model, explicit supervision from a human annotator is needed to label the dataset.☐ Word2Vec can capture only syntactic information, not semantic meaning.☐ Similar to the Bag-of-words method, Word2Vec does not involve neighboring words in extracting word embeddings.

Question 5

0 out of 10 points

Consider an HMM with two states S_1 and S_2 with two observations X_1 and X_2 . In this case, $p(X_2 | S_1)$ can be written as:Selected Answer: ☒ $\sum_{S_2} p(S_1 | S_2) p(X_2 | S_2)$ Answers: ☐ $\sum_{S_2} p(S_1 | S_2) p(X_2 | S_2)$ ☐ $p(S_1 | X_2) p(X_2)$ ☒ $\sum_{S_2} p(X_2 | S_2) p(S_2 | S_1)$ ☐ $\sum_{X_1} p(X_1 | S_1) p(X_2 | X_1)$

Question 6

0 out of 10 points

Which of the following is incorrect about SVD decomposition?

Selected Answer: ☒ Computational complexity of SVD is $O(mn^2)$ for an $n \times m$ matrix.Answers: ☐ Many singular values are going to be zero or negligible and can be dropped.☒ SVD is flexible to adding new words.☐ Rows of the U matrix can be used as word embeddings.☐ Computational complexity of SVD is $O(mn^2)$ for an $n \times m$ matrix.

Question 7

0 out of 10 points

Which of the following equations is not necessarily correct in the SVD decomposition?

Selected Answer: ☒ $A = U\Sigma V^{-1}$ Answers: ☐ $VV^T = V^T V$ ☐ $U^T A = \Sigma V^T$ ☐ $A = U\Sigma V^{-1}$ ☒ $\Sigma = \Sigma^T$

Question 8

0 out of 10 points

Which of the following is not true when comparing HMMs with MEMMs?

Selected Answer: ☒ In contrast to HMM, MEMM assumes that every state depends on all words.

Answers:



Markov assumption on the states holds for both algorithms.

In contrast to HMM, MEMM assumes that every state depends on all words.



Similar to HMM, MEMM uses a generative probabilistic approach.

In MEMM, context is considered to determine POS.

Question 9

0 out of 10 points

Given a sentence of length n words and k possible tags, which of the following answers correctly shows the time complexity of the three decoding algorithms?

Selected Answer:



Viterbi: $O(n^2k)$, greedy: $O(nk)$, brute-force: $O(k^n)$

Answers:

Viterbi: $O(nk^2)$, greedy: $O(nk)$, brute-force: $O(n^k)$

Viterbi: $O(n^2k)$, greedy: $O(nk)$, brute-force: $O(k^n)$

Viterbi: $O(n^2k)$, greedy: $O(n)$, brute-force: $O(n^k)$



Viterbi: $O(nk^2)$, greedy: $O(nk)$, brute-force: $O(k^n)$

Question 10

10 out of 10 points

In the one-hot word representation, the pairwise distance between all pairs of words is proportional to the similarity of their meaning in the context.

Selected Answer: False

Answers: True

False

Saturday, January 20, 2024 11:37:05 PM PST

← OK