E0 259 Data Analytics August 2024

Started on	Thursday, 28 November 2024, 2:10 PM
State	Finished
Completed on	Thursday, 28 November 2024, 3:56 PM
Time taken	1 hour 45 mins
Grade	**13.00** out of 20.00 (**65**%)

Quiz navigation

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Show one page at a time

Finish review

Question 1
Correct
Mark 1.00 out of 1.00
Flag question

Consider three adjacent grounds A, B, C hosting three matches. Teams 1A and 1B scored 250 off 50 overs. Team 1C scored 180 off 50 overs. All Teams 2 (2A, 2B, 2C) played 20 overs and lost 3 wickets, when it rained in all the three grounds (they are adjacent grounds). Team 2A is at 120/3, Team 2B is at 50/3, and Team 2C is at 50/3. Ten overs are lost to rain, and when play resumes, there are 20 overs to play. D/L stands for Duckworth-Lewis.

- A. Under the D/L method, the par scores for Teams 2A and 2C are the same.
- B. Under the D/L method, the par scores for Teams 2B and 2C are the same.
- C. Under the isoprobability criterion, the par scores for Teams 2A and 2C are the same. ✓
- D. Under the isoprobability criterion, the par scores for Teams 2B and 2C are the same.

The correct answer is:
Under the isoprobability criterion, the par scores for Teams 2A and 2C are the same.

Question 2
Correct
Mark 1.00 out of 1.00
Flag question

Given a random sample of numbers 3, 5, 4, 8 from an underlying distribution, find the best estimate of the variance of the underlying distribution, up to 2 decimals rounded upwards.

Answer: 4.67 ✓

The correct answer is: 4.67

Question 3
Correct
Mark 1.00 out of 1.00
Flag question

Let there be N locations for images. One of the locations has an odd ball image, while the others have identical distracter images. The distracter image is different from the odd ball image. The goal is to search for the odd ball image, as explained in class.

A policy is ε-admissible if, no matter what the ground truth for the odd ball location, the policy stops and identifies the correct odd ball location with probability at least $1-\varepsilon$.

Consider an ε-admissible policy. Let $q^{(i)}$ be the induced distribution of the decision on the odd ball location upon stoppage of the ε-admissible policy, when the correct odd ball location is i. Which of the following is correct?

- A. $D(q^{(1)} \| q^{(2)})$ is approximately $\log(1/\varepsilon)$ ✓
- B. $D(q^{(1)} \| q^{(2)})$ is approximately $\exp(\varepsilon)$
- C. Nothing can be said about $D(q^{(1)} \| q^{(2)})$'s value in terms of ε
- D. $D(q^{(1)} \| q^{(2)})$ is approximately $1/\varepsilon$

The correct answer is:
$D(q^{(1)} \| q^{(2)})$ is approximately $\log(1/\varepsilon)$

Question 4
Correct
Mark 1.00 out of 1.00
Flag question

In the context of Latent Dirichlet Allocation (LDA), suppose you have a corpus with a vocabulary of 5 words: $\{w_1, w_2, w_3, w_4, w_5\}$. You're using $K=2$ topics with symmetric Dirichlet priors where $\alpha = 0.5$ for document-topic distributions θ and $\beta = 0.5$ for topic-word distributions φ.

During collapsed Gibbs sampling, you need to compute the conditional probability $P(z_i = k \mid z_{-i}, w)$ for a specific word token $w_i = w_3$ in document d. Here, z_i represents the topic assignments for all other word tokens, and w is the entire set of word tokens.

Given the following counts excluding the current token w_i:

In document d:

- $n_{d,1} = 4$ (number of times topic 1 is assigned in document d)
- $n_{d,2} = 2$ (number of times topic 2 is assigned in document d)

Across the entire corpus:

- For topic 1:
  - $n_{1,w_3} = 3$ (number of times word w_3 is assigned to topic 1)
  - $n_{1,\cdot} = 15$ (total number of words assigned to topic 1)
- For topic 2:
  - $n_{2,w_3} = 1$ (number of times word w_3 is assigned to topic 2)
  - $n_{2,\cdot} = 10$ (total number of words assigned to topic 2).

Using the Gibbs sampling update formula:

Compute the unnormalized probability for $P(z_i = 1 | z_{-i}, w)$ and select the correct expression from the options below.

- A. Proportional to (9/13)*(7/34)
- B. Proportional to (10/13)*(8/35)
- C. Proportional to (9/13)*(7/35) ✓
- D. Proportional to (4/6)*(3/15)

The correct answer is:
Proportional to (9/13)*(7/35)

Question 5
Correct
Mark 1.00 out of 1.00
Flag question

In a topic modeling scenario using Latent Dirichlet Allocation (LDA), suppose you have two topics: Topic 1 and Topic 2. The prior probabilities for the topics are equal, so

P(Topic 1)=P(Topic 2)=0.5.

The word "algorithm" appears in a document, and the conditional probabilities of the word given each topic are:

- P("algorithm" | Topic 1)=0.8
- P("algorithm" | Topic 2)=0.4

What is the probability that the word "algorithm" in the document belongs to Topic 1? Give your answer rounded up to two decimal places.

Answer: 0.67 ✓

The correct answer is: 0.67

Question 6
Incorrect
Mark 0.00 out of 1.00
Flag question

Consider the following undirected graph:

- Vertices: A, B, C, D
- Edges:
  - (A-B)
  - (A-C)
  - (B-D)
  - (C-D)

Calculate the edge betweenness centrality for the edge (A-B). Which of the following is the correct value? Enter your answer rounded to the nearest integer.

Answer: 4 ✗

The correct answer is: 2

Question 7
Correct
Mark 1.00 out of 1.00
Flag question

If the geocentric longitude of Mars is 30 degrees then what is the best approximation for the heliocentric longitude of Mars at this instant? Assume Mars is at 1.5 AUs from the Sun.

- A. 60-arcsin(sin(150 deg)/1.5)
- B. 30-arcsin(sin(150 deg)/1.5) ✓
- C. 30+arcsin(sin(150 deg)/1.5)
- D. 30-arcsin(2 * sin(150 deg)/1.5)

The correct answer is: 30-arcsin(sin(150 deg)/1.5)

Question 8
Correct
Mark 1.00 out of 1.00
Flag question

The BWT of the following string ACACACG is given by which of the following choices? Assume $>A,C,G,T.

- A. $CACAACG
- B. $CCAAACG ✓
- C. $CAAACCG
- D. $CCCAAAG

The correct answer is:
$CCAAACG

Question 9
Incorrect
Mark 0.00 out of 1.00
Flag question

In a recommendation system that combines TF-IDF vectorization with collaborative filtering, consider the following scenario:

- Each item (e.g., document, product) is represented as a TF-IDF vector based on its content.
- User profiles are constructed by aggregating the TF-IDF vectors of the items they have interacted with.
- The system computes similarities using cosine similarity for both item-item and user-user relationships.

Which of the following statements is TRUE regarding the use of TF-IDF in this hybrid recommendation system?

- A. By representing users and items in the same TF-IDF vector space, the system can directly compute user-item affinity without any interaction data. ✗
- B. The inverse document frequency (IDF) component of TF-IDF reduces the influence of terms that are rare across all items, focusing on more common terms.
- C. The TF-IDF weighting scheme helps in emphasizing commonly occurring terms across all items, enhancing collaborative signals.
- D. This approach can alleviate the cold-start problem for new items by leveraging content features in the absence of sufficient user interaction data.

The correct answer is:
This approach can alleviate the cold-start problem for new items by leveraging content features in the absence of sufficient user interaction data.

Question 10
Correct
Mark 1.00 out of 1.00
Flag question

When performing community detection in networks, various metrics such as modularity, edge betweenness, and cut metrics are used to identify community structures. Which of the following statements best describes how these metrics are applied and their effectiveness in different contexts?

- A. Edge betweenness centrality identifies edges that, when removed, disconnect the network, making it effective for detecting community boundaries; cut metrics focus on minimizing the total weight of edges between communities to partition the network efficiently.
- B. Cut metrics are used to maximize the number of edges within communities by removing edges with high modularity scores, whereas modularity aims to minimize the number of edges between communities.
- C. Modularity optimization seeks to maximize the difference between the actual number of intra-community edges and the expected number in a random graph, effectively identifying densely connected communities; edge betweenness is useful for detecting community boundaries by targeting edges that bridge different communities. ✓
- D. Modularity optimization is most effective for detecting small communities in networks with uniform degree distributions, while edge betweenness excels in identifying large communities in scale-free networks.

The correct answer is:
Modularity optimization seeks to maximize the difference between the actual number of intra-community edges and the expected number in a random graph, effectively identifying densely connected communities; edge betweenness is useful for detecting community boundaries by targeting edges that bridge different communities.

Question 11
Correct
Mark 1.00 out of 1.00
Flag question

After which exon of the red and green genes should the lopsided crossover breakpoints be to create a configuration where the distinction between green and red is completely lost? (Enter the exon number as an integer.)

Answer: 1 ✓

The correct answer is: 1

Question 12
Incorrect
Mark 0.00 out of 1.00
Flag question

In a recommendation system using user-based collaborative filtering, you have the following user-item rating matrix:

	Item 1	Item 2	Item 3	Item 4	Item 5
User A	5	3	4	4	?
User B	3	1	2	3	3
User C	4	3	4	3	5
User D	3	3	2	5	4
User E	1	5	5	2	1

Predict the rating of User A for item 5, by using the 2 most similar users to User 1. Give your answer rounded up to one decimal place.

Answer: 4.9 ✗

The correct answer is: 4.5

Question 13
Incorrect
Mark 0.00 out of 1.00
Flag question

In a 50 over ODI, Team 1 scores 80/0 in 10 overs when rain reduces the match to 10 overs for each side. Take $R_1 = 0.1$, $R_2 = 0.34$, and $G(50) = 250$. What is the par score in the D/L framework?

Answer: 119 ✗

The correct answer is: 140

Question 14
Incorrect
Mark 0.00 out of 1.00
Flag question

I have a bit array of size 10 billion. I want to create a rank data structure that along with the bit array should fit into 2 Gigabytes. What is the minimum step size (integer) that I must choose to maximize speed assuming each precomputed answer is 4 bytes?

Answer: 20 ✗

The correct answer is: 54

Question 15
Incorrect
Mark 0.00 out of 1.00
Flag question

Suppose that a neuron fires at an average rate of 1 spike/second when shown the image "1" and at an average rate of 2 spikes/second when shown the image "2". Assume that the spike trains are Poisson point processes with the indicated rates. One of these two images is shown and the number of spikes recorded in one second is 5 spikes. The likelihood ratio of the image being "2" with respect to the image being "1" is Le^{-1}. What is the value of L rounded to one decimal point.

Answer: 6.9 ✗

The correct answer is: 32

Question 16
Correct
Mark 1.00 out of 1.00
Flag question

If the geocentric latitude of Mars at an opposition (Earth between Sun and Mars in one line) is 6 degrees, then what is the heliocentric latitude of Mars at this instant best approximated by? Assume Mars is at 1.5 AUs from the Sun.

- A. arctan(3 tan(6 deg))
- B. arctan(tan(6 deg)/3) ✓
- C. There are multiple possibilities
- D. arctan(2 * tan(6 deg)/3)

The correct answer is:
arctan(tan(6 deg)/3)

Question 17
Correct
Mark 1.00 out of 1.00
Flag question

Consider the Sripati-Olson visual search experiment. Which of the following comes closest to the comparison made by Sripati and Olson in their work?

- A. The time taken by the human subjects on the search task was correlated with the inverse L_2 neuronal distance between the objects.
- B. The time taken by the human subjects on the search task was correlated with the L_1 neuronal distance between the objects.
- C. The time taken by the human subjects on the search task was correlated with the inverse L_1 neuronal distance between the objects. ✓
- D. The time taken by the human subjects on the search task was correlated with the L_2 neuronal distance between the objects.

The correct answer is:
The time taken by the human subjects on the search task was correlated with the inverse L_1 neuronal distance between the objects.

Question 18
Correct
Mark 1.00 out of 1.00
Flag question

In the Probabilistic Latent Semantic Analysis (PLSA) model, which of the following best describes how the probability of a word w in a document d is calculated?

- A. By multiplying the probability of the word $P(w)$ by the probability of the document $P(d)$, assuming independence.
- B. By calculating the joint probability $P(w, d)$ without considering latent topics.
- C. By summing over all topics z the product of the probability of the topic given the document $P(z \mid d)$ and the probability of the word given the topic $P(w \mid z)$. ✓
- D. By summing over all topics z the product of the probability of the word given the topic $P(w \mid z)$ and the probability of the topic $P(z)$.

The correct answer is:
By summing over all topics z the product of the probability of the topic given the document $P(z \mid d)$ and the probability of the word given the topic $P(w \mid z)$.

Question 19
Incorrect
Mark 0.00 out of 1.00
Flag question

You sample weights of people from India and from China with a null hypothesis that the two distributions are identical and normal. The samples are 40, 50, 60, 70 in one group and 30, 40, 50, 60, in the other. What is the F-statistic value, up to 2 decimals rounded upwards?

Answer: 0.39 ✗

The correct answer is: 1.2

Question 20
Correct
Mark 1.00 out of 1.00
Flag question

In a recommendation system utilising latent factor decomposition, each user and each item is represented by a latent feature vector in a low-dimensional space. The predicted rating r^{ui} that user u would give to item i is computed as the dot product of their respective latent feature vectors.

Consider the following latent feature vectors in a 2-dimensional space:

Users: $P_1 = (3, -1), P_2 = (0, 2)$

Items: $Q_A = (1, 4), Q_B = (-2, 1)$

Using the latent factor model, compute the predicted ratings for each user-item pair. Based on your calculations, which of the following statements is true?

- A. Both users prefer Item B over Item A.
- B. Both users prefer Item A over Item B. ✓
- C. User 1 prefers Item A over Item B; User 2 prefers Item B over Item A.
- D. User 1 prefers Item B over Item A; User 2 prefers Item A over Item B.

The correct answer is:
Both users prefer Item A over Item B.

Finish review