Question Bank - Cognitive Analytics CSBA 3009 (AI&ML – 6th Semester)

Unit 1 (Introduction & Basic of Cognitive Computing)

Data Visualizations:

1. Can cognitive analytics visualizations be used to evaluate the effectiveness of machine learning algorithms?
2. What are the potential challenges and limitations of using data visualizations in cognitive analytics?
3. How can computer science contribute to the development of more effective and ethical data visualizations for cognitive research?
4. Compare and contrast different data visualization tools and techniques in the context of cognitive analytics.
5. Design your own data visualization to represent a specific cognitive process. Explain your design choices and how they address the cognitive principles mentioned above.
6. Analyze a real-world example of a data visualization used in cognitive analytics. Identify its strengths and weaknesses based on the questions above.
7. What ethical considerations are involved in using data visualizations in cognitive analytics?
8. What external factors or contextual information might influence the interpretation of the visualization?

Basics of Statistics:

1. How can you use statistical methods to identify and analyze patterns in large-scale cognitive datasets?
2. Participants in a cognitive experiment are asked to choose between two options in a decision-making task. How can you use a binomial distribution to model their choices and analyze the factors influencing their decisions?
3. EEG data often follows a normal distribution. How can this be used to identify potential outliers or anomalies in the data that might indicate cognitive dysfunction?
4. Reaction times in a cognitive task can be modeled using various distributions like exponential or Weibull. How can the choice of distribution impact the analysis of individual differences and learning effects?
5. In a cognitive aging study, researchers measure memory recall scores. How can you use a Poisson distribution to model the probability of recalling a specific number of items and compare memory performance across different age groups?
6. Explain the concept of Shannon entropy in the context of cognitive data. How can it be used to measure the uncertainty or complexity of cognitive states, like attention or emotional arousal?
7. Compare and contrast conditional entropy and mutual information in cognitive analytics. How can these measures be used to assess the relationship between different cognitive variables and their information content?

8. In a brain imaging study, how can you use entropy to analyze the spatial distribution of brain activity during different cognitive tasks? What does it tell you about the functional organization of the brain?
9. How can you use entropy to evaluate the effectiveness of a cognitive intervention or training program? Does a decrease in entropy necessarily indicate improvement in cognitive function?
10. Imagine you're analyzing EEG data from a group of participants performing a cognitive task. How can applying the Law of Large Numbers help you draw more reliable conclusions about the average brain activity patterns?
11. How can the Central Limit Theorem be used to justify the use of parametric statistical tests on cognitive data, even if the individual data points may not be normally distributed?
12. In a cognitive study, you want to measure the effectiveness of a new training program on memory recall. How can you use the Law of Large Numbers to determine the minimum sample size needed to obtain statistically significant results?
13. Researchers analyze reaction times from hundreds of participants in a visual attention task. How can they leverage the Central Limit Theorem to estimate the population mean and standard deviation of reaction times, even if the individual data follows a non-normal distribution?
14. A brain-computer interface (BCI) system aims to predict user intentions based on brain activity. How can you use the Law of Large Numbers and the Central Limit Theorem to improve the reliability and generalizability of the BCI predictions?
15. A study investigates the number of times participants press a button during a cognitive task. It follows a Poisson distribution with a mean of 5 presses. Calculate the probability of a participant pressing the button exactly 3 times.
16. Reaction times in a cognitive experiment are modelled by a normal distribution with a mean of 500 milliseconds and a standard deviation of 100 milliseconds. What is the probability of a participant reacting within the range of 400-600 milliseconds?
17. A machine learning classifier predicts cognitive states based on EEG data. It follows a multinomial distribution with 4 possible states (relaxed, focused, distracted, drowsy). Given prior probabilities of each state, calculate the posterior probability of being focused after observing specific EEG features.

Unit 2 (Predictive Modeling)

Big Data

1. Differentiate between Big Data Analytics (BDA) and traditional data analytics in the context of cognitive computing.
2. Explain the 5 V's of Big Data (Volume, Velocity, Variety, Veracity, and Value) and their significance for cognitive computing applications.
3. Describe two different Big Data frameworks (e.g., Hadoop, Spark) and explain their suitability for specific cognitive computing tasks.
4. Imagine you are developing a cognitive computing system that can diagnose mental health conditions. How would you leverage BDA to improve its accuracy and effectiveness?

Data Mining

1. Differentiate between data mining and traditional data analysis. How does data mining specifically benefit cognitive computing applications?
2. What are the different types of data mining techniques used in cognitive computing? Briefly describe their purpose and potential applications.
3. Explain the concept of feature selection in data mining and its importance in building effective cognitive computing models.
4. Describe how data mining can be used to identify patterns in brain activity that correlate with specific cognitive states (e.g., attention, memory).
5. Explain how data mining can be used to analyze large-scale text data to understand human emotions and sentiment.
6. How can data mining be used to improve the performance of machine learning models used in cognitive computing tasks (e.g., natural language processing, image recognition)?
7. Describe two different data mining algorithms (e.g., decision trees, clustering) and explain their suitability for specific cognitive computing tasks.
8. How can visualization techniques be used to explore and understand the results of data mining in cognitive computing?
9. A study measures EEG data of 50 participants in three different cognitive states (relaxed, focused, distracted). You apply K-means clustering with K=3. Calculate the within-cluster and between-cluster sum of squares to evaluate the clustering quality.
10. Imagine you have a dataset of brain activity maps from fMRI scans. You use hierarchical clustering to identify groups of regions with similar activation patterns. How many clusters would you choose based on a dendrogram and silhouette analysis?
11. Compare the performance of two different data mining techniques (e.g., K-means vs. DBSCAN) on a specific cognitive analysis task based on evaluation metrics. Discuss the factors influencing the choice of technique.
12. You analyze a dataset of student response times in an online learning platform. Use logistic regression to predict whether a student is likely to answer a question correctly based on their response time history. Evaluate the model performance using metrics like F1 score and Accuracy.

LAB Related Questions

Regression Analysis

1. Differentiate between linear and non-linear regression in the context of cognitive computing. When would you choose each type of model?
2. Imagine a study measuring reaction times in a visual search task. You use linear regression to predict reaction time based on the number of distractor items present. Interpret the coefficients of your model and explain their meaning.
3. Explain how polynomial regression can be used to model the relationship between cognitive performance and age in a large-scale study.
4. How can regression analysis be used to analyze the relationship between brain activity (e.g., fMRI data) and specific cognitive variables (e.g., memory recall scores)?
5. What are the challenges of dealing with missing data or outliers in regression analysis for cognitive data? How can you handle these issues?

6. Explain the concept of residual analysis and its importance in evaluating the fit and assumptions of a regression model.

Numerical (Regression Analysis)

Numerical 1: Predicting Memory Performance

Imagine a study collects data on memory performance (number of items recalled) and brain activity (measured by EEG) from 30 participants. You want to use linear regression to predict memory performance based on specific EEG features (e.g., average amplitude in a specific frequency band).

Answer the following questions:

i. Analyze the data and build a linear regression model relating memory performance to the EEG feature.
ii. Calculate the coefficient of determination (R-squared) and explain what it tells you about the model's fit.
iii. Perform a hypothesis test to determine if the relationship between the EEG feature and memory performance is statistically significant.
iv. Interpret the regression coefficients and explain how changes in the EEG feature are associated with changes in memory performance.

Numerical 2: Evaluating Training Intervention

A study investigates the effect of a cognitive training program on reaction times in a visual attention task. 20 participants perform the task before and after the training program, and their reaction times are recorded. You want to use linear regression to quantify the effect of the training program.

Answer the following questions:

i. Build a linear regression model where the dependent variable is the change in reaction time (post-training minus pre-training) and the independent variable is a binary indicator of whether the participant received the training program.
ii. Interpret the regression coefficient for the training program indicator and explain how much the program is predicted to improve reaction times on average.
iii. Evaluate the model's assumptions and discuss any potential limitations of using regression in this scenario.

Numerical 3: Modeling Individual Differences

A study assesses cognitive flexibility in a group of 50 participants. Their performance is measured, and personality traits (e.g., openness to experience, neuroticism) are also collected. You want to use regression analysis to understand how personality traits influence cognitive flexibility.

Answer the following Questions:

i. Build a multiple regression model where cognitive flexibility is the dependent variable and the independent variables are the personality traits.
ii. Interpret the regression coefficients for each personality trait and explain how they relate to cognitive flexibility.
iii. Assess multicollinearity among the independent variables and discuss its potential impact on the model's interpretation.

Numerical 4: Predicting Emotion from Facial Expressions:

Imagine a dataset with 1000 images labeled with six different emotions (happy, sad, angry, etc.) and corresponding facial feature measurements (e.g., eyebrow position, mouth curvature). You want to compare Random Forest and Logistic Regression for predicting emotion based on the features.

Answer the following questions:

i. Split the data into training and testing sets. Train a Random Forest and a Logistic Regression model on the training set to predict emotion based on facial features.
ii. Calculate accuracy, precision, recall, and F1-score for both models on the testing set. Compare their performance and discuss which model is better suited for this task.
iii. Analyze the feature importance scores in the Random Forest model. Which features are most important for predicting emotion?

Numerical 5: Personalizing Learning Experiences:

An online learning platform collects data on student performance, learning styles, and engagement. You want to use Random Forest and Logistic Regression to predict which learning materials are most effective for each student.

Answer the following questions:

i. Train both models to predict student performance on specific learning materials based on their data. Evaluate their performance metrics like Root Mean Squared Error (RMSE).
ii. Compare the interpretability of both models. How easy is it to understand why a student is predicted to perform well or poorly on specific material?
iii. Discuss the advantages and limitations of using Random Forest and Logistic Regression for personalized learning applications.