Assignment 1: Understanding data : Part 3

INFO 7390: Advance Data Science ¶

Vinay Jogani, NEU ID: 002839145

Data Cleaning Techniques:

Question: What is the primary goal of data cleaning in the context of data pre-processing?

- A. Enhancing data visualization
- B. Improving model performance
- C. Reducing dataset size
- D. Adding noise to the data

Data Transformation Methods:

Question: Which of the following is a data transformation method used to handle skewed distributions?

- A. Standardization
- B. Min-Max Scaling
- C. Log Transformation
- D. Imputation

Data Normalization Approaches:

Question: In data normalization, what does Z-score normalization (standardization) involve?

- A. Scaling data to a specific range [0, 1]
- B. Converting data into binary format
- C. Scaling data to have a mean of 0 and standard deviation of 1
- D. Removing outliers from the dataset

Handling Missing Data and Imputation Techniques:

Question: What is a common imputation technique for missing numerical data?

- A. Mean imputation
- B. Mode imputation
- C. Random imputation
- D. One-hot encoding

Data Integration Techniques:

Question: What does data integration involve in the context of data pre-processing?

- A. Combining multiple datasets into a single, cohesive dataset
- B. Removing outliers from a dataset
- C. Transforming categorical variables into numerical format
- D. Imputing missing values in a dataset

Data Visualization Types and Techniques:

Question: Which type of data visualization is suitable for displaying the distribution of a single numerical variable?

- A. Bar chart
- B. Scatter plot
- C. Histogram
- D. Pie chart

Exploratory Data Analysis:

Question: What is the primary goal of exploratory data analysis (EDA)?

- A. Building predictive models
- B. Understanding the structure of the dataset
- C. Imputing missing values
- D. Applying machine learning algorithms

Importance of Understanding Data:

Question: Why is understanding the context and domain of the data crucial in data pre-processing?

- A. It helps in hiding the complexities of the data
- B. It facilitates faster computation
- C. It enables better decision-making in data processing
- D. It ensures data security

Data Reduction Techniques:

Question: What is the purpose of data reduction techniques in data pre-processing?

- A. Increasing the dataset size
- B. Reducing noise in the data
- C. Introducing missing values
- D. Encoding categorical variables

Feature Engineering:

Question: What does feature engineering involve in data pre-processing?

- A. Reducing the dimensionality of the dataset
- B. Creating new features or modifying existing features
- C. Applying data visualization techniques
- D. Eliminating outliers from the dataset

Answers:

- B, 2. C, 3. C, 4. A, 5. A, 6. C, 7. B, 8. C, 9. B, 10. B