Question Bank

Exploratory Data Analysis and Feature Engineering

Unit-1: Fundamentals of Exploratory Data Analysis

1. Explain Data Science Process.
2. What is significance of Exploratory Data Analysis (EDA)? Explain steps in EDA.
3. Explain steps in Exploratory Data Analysis (EDA).
4. What are the broad groups of dataset?
5. Explain different types of measurement scales in statistics with suitable example.
6. Compare Exploratory Data Analysis (EDA) with classical and Bayesian analysis.
7. Explain visual aids for Exploratory Data Analysis (EDA). How to choose the best chart to visualize data?
8. Explain visual aids Line chart, Bar chart, Scatter plot, Area plot, Pie chart, Table chart for Exploratory Data Analysis (EDA).
9. Explain visual aids Pie chart, Table chart, Polar chart, Histogram, Lollipop chart. What are the guidelines for choosing best chart to visualize data?

Unit -2: Hypothesis Testing and Analysis of Variance

1. Explain Descriptive Statistics and Inferential Statistics.
2. Explain fundamentals of Kernel Density Estimation (KDE) and its use to estimate kernel density.
3. Explain Cumulative Distribution Function (CDF).
4. What is Hypothesis testing? Explain steps in Hypothesis testing.
5. Explain One Sample T-test Hypothesis testing technique with suitable example.
6. Explain One Independent Samples Hypothesis testing technique with suitable example.
7. Explain Fisher's Hypothesis testing technique with suitable example.
8. Explain Chi-Square Test of Independence with suitable example.
9. Explain one way ANOVA (Analysis of Variance).
10. Explain two way ANOVA (Analysis of Variance).

Unit-3: Exploratory Data Analysis

1. What is Exploratory Data Analysis (EDA)? What are objectives of EDA? Why do we need to perform EDA?
2. Explain Univariate non-graphical Exploratory Data Analysis (EDA).
3. What is Skewness and kurtosis? How Skewness and kurtosis can be used in Exploratory Data Analysis (EDA)? How Skewness can be used to decide distribution of data?
4. Explain Univariate Graphical Exploratory Data Analysis (EDA).
5. What is histogram? How histogram can be used for examining the relationship between a numerical predictor and the target?
6. **Explain use of Stem-and-leaf plots, Boxplots, Quantile-normal plots in Univariate graphical EDA.**
7. Explain Multivariate non-graphical Exploratory Data Analysis (EDA).,
8. Explain Multivariate Graphical Exploratory Data Analysis (EDA).
9. How histogram can be used to examine relationship between a numerical predictor and the target variable?
10. Explain steps for Exploratory Data Analysis for Text data.

Unit 4: Feature Construction and Feature Selection

1. Explain methods for Imputing categorical features.
2. Explain methods for Encoding categorical variables.
3. Explain methods for Bucketing continuous features into categories.
4. Explain methods for Extending numerical features.
5. Explain Text-specific feature construction methods.
6. What is Feature Selection? What are goals of Feature Selection?
7. Feature selection mythologies intrinsic (or implicit) methods, filter methods, and wrapper methods
8. What is the effect of Irrelevant Features in the data set?
9. Explain Simple univariate filters for feature selection.
10. Explain Recursive Feature Elimination (RFE) method for feature selection.

Unit-5 : Feature Transformations

1. What is feature transformation? Explain machine learning pipeline.
2. Compare feature transformations, feature selection, and feature construction.
3. Explain dimensionality reduction technique Principal Component Analysis (PCA).
4. What is Principal Component Analysis (PCA)? How PCA can be used to reduce dimensionality?
5. What are Principal Components? How Principal Components are used for reducing the dimensions of the data set?
6. **How centering and scaling data affects** Principal Component Analysis (PCA)?
7. Explain Linear Discriminant Analysis (LDA) dimensionality reduction technique.
8. Compare Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

Unit 6: Feature Learning

1. What is parametric assumptions of data? What are assumptions for Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).
2. What are differences between feature learning and transformation?
3. Explain Restricted Boltzmann Machines (RBM) architecture.
4. What is Restricted Boltzmann Machines (RBM)? How data is rreconstructed in RBM?
5. What is Word embedding? What are methods for Word embedding?
6. Explain Word2Vec is a method to construct Word embedding.
7. Explain Skip Gram method of Word2Vec.
8. Explain Common Bag of Words (CBOW) method of Word2Vec.
9. What are applications of **of** Word embedding?