Basic Knowledge and Quiz on Machine Learning Topics

Descriptive and Inferential Statistics

Descriptive statistics summarize or describe the main features of a dataset using measures such as mean, median, mode, and standard deviation. Inferential statistics use a random sample of data taken from a population to make inferences or predictions about that population.

Quiz

Which of the following is an example of descriptive statistics?

A) Hypothesis testing

B) Mean calculation

C) Regression analysis

D) ANOVA

Correct Answer: B

Correlation

Correlation measures the strength and direction of the linear relationship between two variables. It is represented by a correlation coefficient (r), where values range from -1 to 1.

Quiz

What does a correlation coefficient of 0 indicate?

A) Strong positive relationship

B) No relationship

C) Strong negative relationship

D) Weak positive relationship

Correct Answer: B

Causality

Causality implies that a change in one variable directly causes a change in another. Correlation does not imply causation, as other factors may be involved in the relationship.

Quiz

Which of the following best describes causality?

A) One event leads directly to another

B) Two events happen together

C) Events occur by chance

D) Events are independent

Correct Answer: A

Linear Regression

Linear regression is used to model the relationship between a dependent variable and one or more independent variables. It predicts the dependent variable as a linear function of the independent variables.

Quiz

What is the primary goal of linear regression?

A) Classifying data

B) Predicting a continuous value

C) Clustering data

D) Reducing dimensionality

Correct Answer: B

Logistic Regression

Logistic regression is used to model binary outcomes by estimating the probability that an observation belongs to a certain category. It is based on the logistic function.

Quiz

What type of outcome does logistic regression predict?

A) Continuous

B) Categorical

C) Ordinal

D) Dimensional

Correct Answer: B

Underfitting

Underfitting occurs when a model is too simple and fails to capture the underlying patterns in the data, leading to poor performance on both training and test data.

Quiz

Underfitting typically happens when:

A) The model is too complex

B) There is too much noise in the data

C) The model is too simple

D) The data is insufficient

Correct Answer: C

Overfitting

Overfitting occurs when a model learns the training data too well, including its noise, and performs poorly on new data. It fits the training data too closely and lacks generalization ability.

Quiz

Overfitting happens when:

A) The model performs well on training data but poorly on new data

B) The model performs poorly on both training and test data

C) The model performs well on both training and test data

D) The model is too simple

Correct Answer: A

Gradient Descent

Gradient Descent is an optimization algorithm used to minimize a loss function in machine learning models by iteratively updating parameters in the direction of the steepest decrease of the loss.

Quiz

What is the goal of gradient descent?

A) To maximize the loss function

B) To minimize the loss function

C) To increase model complexity

D) To optimize training time

Correct Answer: B

Drift Detection

Drift detection identifies when a machine learning model's performance degrades due to changes in data patterns over time. Drift occurs when the distribution of the data changes, affecting the model's accuracy.

Quiz

What does drift detection help with?

A) Monitoring performance over time

B) Increasing dataset size

C) Tuning hyperparameters

D) Improving cross-validation

Correct Answer: A

Data Visualization

Data visualization is the graphical representation of data to help users understand complex patterns, trends, and relationships within the data using tools like bar charts, histograms, and scatter plots.

Quiz

Which of the following is an example of a data visualization technique?

A) Scatter plot

B) Mean calculation

C) Data cleaning

D) Hypothesis testing

Correct Answer: A

Filling Missing Values

Filling missing values is a data preprocessing technique used to handle incomplete data. Techniques include mean/mode imputation, forward/backward filling, and using algorithms like KNN to fill gaps.

Quiz

Which technique is commonly used for filling missing values?

A) Mean imputation

B) Data augmentation

C) Normalization

D) Feature scaling

Correct Answer: A

Treating Outliers

Treating outliers involves identifying and handling extreme values that deviate significantly from the rest of the data. Techniques include removal, transformation, or capping the outlier values.

Quiz

Which method is used to treat outliers?

A) Normalization

B) Transformation

C) Data expansion

D) Sampling

Correct Answer: B