CTRL - V

Chapter 2 – Machine Learning and Algorithms
Study Guide Question Sheet

Section A: Multiple Choice (20 Questions)
Choose the correct answer (A, B, C, or D).

What is Machine Learning primarily concerned with?
A. Preprogramming fixed rules
B. Enabling machines to learn from data
C. Hardcoding expert knowledge
D. Storing information efficiently
Answer: B

Which of the following is NOT a type of Machine Learning?
A. Supervised Learning
B. Unsupervised Learning
C. Reinforcement Learning
D. Manual Programming
Answer: D

In supervised learning, training data consists of:
A. Unlabeled examples
B. Partially labeled data
C. Input-output pairs
D. Only outputs
Answer: C

Which of the following algorithms is used for classification tasks?
A. Linear Regression
B. K-Means Clustering
C. Decision Trees
D. Apriori
Answer: C

The process of dividing data into training and testing sets is called:
A. Model Tuning
B. Data Validation
C. Data Splitting
D. Feature Engineering
Answer: C

In a linear regression model, the goal is to:
A. Find clusters of similar points
B. Predict continuous outcomes
C. Predict categorical labels
D. Reduce dimensionality
Answer: B

The algorithm used for grouping unlabeled data is:
A. K-Means
B. Naïve Bayes
C. Logistic Regression
D. Decision Tree
Answer: A

Overfitting occurs when a model:
A. Performs poorly on both training and test data
B. Fits training data too well but fails to generalize
C. Is undertrained
D. Has too few parameters
Answer: B

The evaluation metric used for regression is typically:
A. Accuracy
B. Precision
C. Mean Squared Error (MSE)
D. Recall
Answer: C

A confusion matrix is used for evaluating:
A. Regression models
B. Clustering models
C. Classification models
D. Reinforcement models
Answer: C

Gradient Descent is used to:
A. Maximize error
B. Minimize the cost function
C. Randomize model parameters
D. Generate synthetic data
Answer: B

What is the main purpose of Feature Scaling?
A. To normalize data for better convergence
B. To add more features
C. To remove noise completely
D. To overfit the model
Answer: A

The K in KNearest Neighbors represents:
A. The number of features
B. The number of neighbors considered
C. The number of clusters
D. The number of labels
Answer: B

Which Machine Learning type uses reward and punishment signals?
A. Supervised Learning
B. Unsupervised Learning
C. Reinforcement Learning
D. Self-Supervised Learning
Answer: C
Dimensionality reduction helps by:
A. Increasing model complexity
B. Reducing overfitting and improving visualization
C. Expanding dataset size
D. Creating redundant features
Answer: B

The algorithm that assumes feature independence is:
A. Naïve Bayes
B. Logistic Regression
C. SVM
D. Decision Tree
Answer: A

Cross-validation is used to:
A. Validate test data
B. Assess model performance stability

C. Overfit intentionally
D. Shuffle labels randomly
Answer: B

The Sigmoid function is mainly used in:
A. Linear Regression
B. Logistic Regression
C. Clustering
D. Reinforcement Learning
Answer: B

Which of the following is a non-parametric algorithm?
A. Linear Regression
B. Decision Trees
C. Logistic Regression
D. Naïve Bayes
Answer: B

A model that performs poorly on both training and test data is said to be:
A. Overfitted
B. Underfitted
C. Regularized
D. Tuned
Answer: B

Section B: True or False (20 Statements)

Machine Learning allows computers to learn patterns without explicit programming. — **True**
Supervised Learning uses unlabeled data. — **False**
K-Means is a type of unsupervised learning. — **True**
Overfitting is desirable in predictive models. — **False**
Linear Regression can be used for predicting prices. — **True**
Accuracy is the best metric for all models. — **False**
Reinforcement Learning uses rewards and penalties. — **True**
Logistic Regression is used for classification problems. **True**
Neural Networks are a subset of Machine Learning. — **True**
PCA (Principal Component Analysis) increases feature dimensions. **False**
Regularization helps reduce overfitting. — **True**
Decision Trees can handle both categorical and numerical data. **True**
Gradient Descent updates weights to minimize loss. — **True**
The loss function measures how well a model fits data. **True**
Deep Learning is a type of Reinforcement Learning. — **False**
Clustering requires labeled training data. **False**
Random Forests are based on multiple decision trees. — **True**
Bias and variance are key components of model error. **True**
Feature Scaling is unnecessary for distance-based models. — **False**
AI and Machine Learning are the same concept. **False**

Section C: Open-Ended Essay Questions (10)

Define Machine Learning and explain its relationship with Artificial Intelligence.
Describe the three main types of Machine Learning with examples.
Explain the difference between overfitting and underfitting.
How does Gradient Descent optimize learning in algorithms?
Discuss the importance of data preprocessing and feature scaling.