

Chapter-wise Test Bank

for the book

Artificial Intelligence and Machine Learning

Pradeep Singh, Tapan K. Gandhi, Balasubramanian Raman

July 6, 2025



Chapter 1: Foundations of AI

Multiple Choice Questions (MCQs)

1. Which of the following best defines Artificial Intelligence (AI)?

- (a) Programming computers to mimic human mistakes.
- (b) Building systems that operate without any human input.
- (c) Developing intelligent agents that perceive and act rationally.
- (d) Creating machines that simply follow hard-coded rules.

Answer: (c)

2. Which of the following is NOT a type of AI system?

- (a) Narrow AI
- (b) General AI
- (c) Superintelligent AI
- (d) Conscious AI

Answer: (d)

3. Which period in AI history is known for reduced funding and skepticism about progress?

- (a) The Golden Age of AI
- (b) AI Spring
- (c) AI Winter
- (d) Deep Learning Revolution

Answer: (c)

4. What was the primary limitation of early symbolic AI systems?

- (a) Lack of rule-based frameworks
- (b) High performance on perception tasks
- (c) Inability to adapt to uncertainty or learn from data
- (d) Integration of deep neural networks

Answer: (c)

5. Which of the following is a key principle of AI development?

- (a) Fixed logic rules
- (b) Passive data storage
- (c) Adaptation and learning from data
- (d) Manual hardcoding of intelligent responses

Answer: (c)

Short Answer Questions**Q1. Differentiate between Strong AI and Weak AI. Provide one example of each.**

Answer:

Strong AI refers to systems with general cognitive abilities akin to human intelligence, including consciousness and self-awareness (e.g., a hypothetical AGI model).

Weak AI focuses on specific tasks and lacks generalization (e.g., Siri or Google Translate).

Q2. Explain the main causes and impact of the AI Winter.

Answer:

The AI Winter occurred due to unmet expectations, lack of computational power, and poor performance of symbolic systems. It led to decreased funding, skepticism, and stalled progress in AI research.

Q3. List any three applications of AI in modern technology.

Answer:

Examples include:

- Autonomous vehicles (e.g., Tesla Autopilot)
- Personalized recommendation systems (e.g., Netflix)

- Medical diagnosis tools (e.g., AI-assisted radiology)

Q4. What are the ethical concerns involved in AI development?**Answer:**

Concerns include bias in algorithms, lack of accountability, surveillance risks, data privacy violations, and the societal impact of automation and job displacement.

Q5. Define the term "rational agent" in the context of AI.**Answer:**

A rational agent is an entity that perceives its environment and takes actions that maximize its expected performance measure, based on its knowledge and goals.

Conceptual Discussion Question

Q: Discuss the transition from symbolic AI to machine learning and then to deep learning. What were the key motivations and shifts in methodology?**Answer:**

Symbolic AI emphasized rule-based logic systems, which failed in tasks involving uncertainty or perception. Machine learning introduced data-driven approaches capable of generalization and adaptation. Deep learning leveraged multi-layered neural networks and GPUs to achieve breakthroughs in speech, vision, and NLP, representing a shift toward hierarchical representation learning.

True/False

1. AI only includes machine learning. **False**
2. Deep learning is a subset of machine learning. **True**
3. Turing Test was proposed to evaluate machine learning algorithms. **False**
4. Strong AI systems do not yet exist in practice. **True**
5. AI has no application in creative fields. **False**

Chapter 2: Intelligent Agents and Environments

Multiple Choice Questions (MCQs)

1. Which of the following is a correct definition of an intelligent agent?

- (a) A system that blindly executes commands.
- (b) An entity that perceives its environment and takes actions to maximize some performance measure.
- (c) A rule-based machine with hardcoded logic.
- (d) Any software that uses data structures.

Answer: (b)

2. What does the PEAS framework stand for in the context of agent design?

- (a) Perception–Estimation–Action–Satisfaction
- (b) Performance measure–Environment–Actuators–Sensors
- (c) Policy–Environment–Actions–States
- (d) Plan–Execute–Assess–Simulate

Answer: (b)

3. Which type of environment is both fully observable and deterministic?

- (a) Self-driving car scenario
- (b) Game of chess
- (c) Online stock trading system
- (d) Chatbot in customer service

Answer: (b)

4. What distinguishes a utility-based agent from a goal-based agent?

- (a) It ignores percepts.
- (b) It can learn without rewards.
- (c) It evaluates multiple paths using a utility function.
- (d) It only chooses the shortest path to a goal.

Answer: (c)

5. Which of the following is not a characteristic of a rational agent?

- (a) Acts to maximize expected performance
- (b) Considers future percepts
- (c) Always chooses randomly
- (d) Acts based on available knowledge and percepts

Answer: (c)

Short Answer Questions**Q1. Define the concept of rationality in the context of intelligent agents.**

Answer:

Rationality refers to the ability of an agent to select actions that maximize its expected performance measure, given the percept sequence, available knowledge, and computational limits.

Q2. Differentiate between deterministic and stochastic environments with one example each.

Answer:

In deterministic environments, the outcome of each action is predictable (e.g., a maze-solving robot). In stochastic environments, outcomes are governed by probability (e.g., stock market simulation).

Q3. List and explain two components of the PEAS framework.

Answer:

- **Performance measure:** Criteria to evaluate agent success (e.g., delivery time for a logistics agent).

- **Sensors:** Mechanisms to perceive environment (e.g., cameras in robots).

Q4. What is the role of a learning agent?

Answer:

A learning agent improves its performance over time by learning from experience. It has components such as performance elements, learning elements, and critics to update its behavior.

Q5. What are some ethical concerns in agent behavior?

Answer:

Concerns include biased decision-making, safety in autonomous operation, privacy of user data, and unintended consequences of learned policies.

True/False

1. Goal-based agents make decisions by evaluating all possible outcomes of their actions.
True
2. Environments in AI are always deterministic. **False**
3. Rationality guarantees success in unknown environments. **False**
4. Reflex agents operate based on condition-action rules. **True**
5. Utility-based agents are inferior to learning agents. **False**

Conceptual Discussion Question

Q: Discuss the classification of environments and explain how it affects the design of agents. Provide real-world examples for each type.

Answer:

Environments can be classified as:

- **Fully vs. Partially Observable:** Fully observable (e.g., chess) allows full state access, while partially observable (e.g., driving) has hidden elements.
- **Deterministic vs. Stochastic:** Deterministic (e.g., robot arm in assembly) is predictable; stochastic (e.g., stock trading) is random.
- **Episodic vs. Sequential:** Episodic (e.g., image classification) has independent decisions; sequential (e.g., robotics navigation) requires planning across time.
- **Static vs. Dynamic:** Static (e.g., crossword solving) does not change during decision-making; dynamic (e.g., multiplayer games) evolves with time.
- **Discrete vs. Continuous:** Discrete (e.g., board games) has finite states; continuous (e.g., drone flight) involves real-valued input/output.

Each type requires a tailored agent design—reflex rules for simple settings and utility/learning modules for complex environments.

Chapter 3: Search and Problem Solving

Multiple Choice Questions (MCQs)

1. Which of the following best describes a problem in AI search?

- (a) A set of equations with known variables
- (b) A tuple (S, A, T, s_0, G) where S is states, A is actions, T is the transition function, s_0 is initial state, and G is goal state(s)
- (c) A random function over time
- (d) An unsolvable task

Answer: (b)

2. Which search algorithm guarantees the shortest path if the cost function is uniform?

- (a) Depth-First Search
- (b) Breadth-First Search
- (c) Hill-Climbing
- (d) Random Walk

Answer: (b)

3. Which of the following is NOT a feature of the A* algorithm?

- (a) Uses both actual cost and estimated future cost
- (b) Is complete and optimal if $h(n)$ is admissible
- (c) Explores random successors first
- (d) Requires evaluation function $f(n) = g(n) + h(n)$

Answer: (c)

4. Which of the following problems is most suitable for adversarial search?

- (a) Route finding in a GPS system
- (b) Scheduling of train arrivals
- (c) Chess game playing
- (d) Diagnosing diseases

Answer: (c)

5. Which method is best suited for solving Constraint Satisfaction Problems (CSPs)?

- (a) Uniform Cost Search
- (b) Backtracking with constraint propagation
- (c) Simulated Annealing
- (d) Random Forest

Answer: (b)

Short Answer Questions**Q1. Define the components of a search problem.**

Answer:

A search problem is defined by:

- Initial state s_0
- Set of actions A
- Transition model $T(s, a)$
- Goal test $G(s)$
- Path cost function c

Q2. Differentiate between uninformed and informed search strategies with one example each.

Answer:

Uninformed (blind) search has no domain knowledge beyond the problem definition (e.g., BFS).

Informed (heuristic) search uses additional information to guide search (e.g., A* algorithm).

Q3. What is the role of a heuristic function in informed search?

Answer:

Heuristic function $h(n)$ estimates the cost from node n to a goal. It helps prioritize nodes closer to the goal in algorithms like A* and Greedy Best-First Search.

Q4. Describe the Minimax algorithm used in adversarial search.

Answer:

Minimax chooses actions that maximize the agent's minimum gain, assuming the opponent plays optimally. It constructs a game tree and recursively computes utility values for decision nodes.

Q5. What are constraint satisfaction problems? Give an example.

Answer:

CSPs involve variables with domains and constraints between them. Solutions satisfy all constraints.

Example: Map coloring where adjacent regions must have different colors.

True/False

1. Breadth-First Search always finds the optimal solution if all step costs are equal. **True**
2. Hill-Climbing can get stuck in local optima. **True**
3. A* search guarantees completeness only when the heuristic is inconsistent. **False**
4. In CSPs, arc-consistency means every value of one variable has a supporting value in the other. **True**
5. Alpha-Beta pruning reduces the size of the game tree without affecting the outcome. **True**

Conceptual Discussion Question

Q: Compare and contrast local search and complete search techniques in AI. Provide examples of where each is appropriate.

Answer:

Complete search explores the entire search space to guarantee finding a solution if one exists (e.g., BFS, DFS, A*). It is suitable for small to medium discrete spaces.

Local search operates in the space of complete states and uses iterative improvement (e.g., Hill-Climbing, Simulated Annealing). It is appropriate for optimization in large or infinite spaces like N-Queens or traveling salesperson problems.

Local search is memory-efficient but may converge to local optima, while complete search offers guarantees but is computationally expensive.

Chapter 4: Knowledge Representation and Logical Reasoning

Multiple Choice Questions (MCQs)

1. Which of the following is NOT a valid type of knowledge representation?

- (a) Semantic networks
- (b) Production rules
- (c) Propositional calculus
- (d) Pixel encoding

Answer: (d)

2. Which logic allows quantification over variables such as "for all" and "there exists"?

- (a) Propositional logic
- (b) First-order logic
- (c) Modal logic
- (d) Temporal logic

Answer: (b)

3. What is a drawback of using propositional logic for complex AI systems?

- (a) It requires too much memory
- (b) It cannot represent relationships between objects
- (c) It is probabilistic
- (d) It is not decidable

Answer: (b)

4. Which of the following is a sound inference technique used in rule-based systems?

- (a) Random sampling
- (b) Forward chaining
- (c) Tabulation
- (d) Residual propagation

Answer: (b)

5. What is the purpose of ontological engineering in AI?

- (a) Build game-playing strategies
- (b) Provide statistical estimation
- (c) Formalize the structure of domain-specific knowledge
- (d) Optimize path-finding in search problems

Answer: (c)

Short Answer Questions

Q1. What are the advantages of using First-Order Logic (FOL) over Propositional Logic in AI?

Answer:

FOL supports quantifiers (\forall, \exists), enabling representation of relations and generalizations. It can model complex domains more compactly and semantically richly than propositional logic.

Q2. Explain the difference between forward chaining and backward chaining.

Answer:

Forward chaining starts with facts and applies rules to infer conclusions, moving toward goals. Backward chaining starts from the goal and works backward to determine whether it can be derived from known facts.

Q3. Give one example each of declarative and procedural knowledge in AI.

Answer:

- Declarative: "Paris is the capital of France."
- Procedural: A sorting algorithm to arrange items in ascending order.

Q4. What is resolution in propositional logic?**Answer:**

Resolution is an inference rule that derives a new clause by eliminating a pair of complementary literals from two clauses. It is sound and complete for propositional logic.

Q5. What is the role of a knowledge base in a knowledge-based agent?**Answer:**

The knowledge base stores facts, rules, and heuristics in a formal representation. It is queried by the inference engine to derive conclusions or make decisions.

True/False

1. Ontologies represent structured knowledge using concepts, relationships, and hierarchies. **True**
2. Propositional logic can express facts but not relations between objects. **True**
3. Resolution is only applicable to probabilistic models. **False**
4. Inference engines apply rules to known facts to generate new knowledge. **True**
5. First-order logic cannot model recursive relationships. **False**

Conceptual Discussion Question

Q: Discuss the significance of logical reasoning in AI. How does it enable intelligent behavior, and what are the practical limitations of logic-based systems?

Answer:

Logical reasoning enables AI systems to deduce new facts from known ones, verify consistency, and make decisions. It provides soundness and completeness in reasoning tasks. Logic is especially powerful in rule-based expert systems, planning, and verification. However, its limitations include lack of support for uncertainty, poor scalability in large domains, and difficulty in learning rules from raw data. Integration with probabilistic reasoning or neural-symbolic systems helps overcome these issues.

Chapter 5: Probabilistic Reasoning and Uncertainty

Multiple Choice Questions (MCQs)

1. Which of the following is a key motivation for using probability in AI?

- (a) All AI problems are deterministic
- (b) Logic systems are faster
- (c) Real-world data often involve noise and uncertainty
- (d) Symbolic representations are more scalable

Answer: (c)

2. In Bayesian networks, what does each edge represent?

- (a) A random noise channel
- (b) A logical dependency
- (c) A conditional dependence between parent and child
- (d) A distance function

Answer: (c)

3. Which of the following statements is true regarding Naïve Bayes classifiers?

- (a) They assume conditional independence among features given the class.
- (b) They require training of deep neural networks.
- (c) They rely on joint maximum likelihood of all features.
- (d) They are always more accurate than decision trees.

Answer: (a)

4. Which inference method is exact in probabilistic graphical models?

- (a) Sampling via MCMC
- (b) Belief propagation in trees
- (c) Loopy belief propagation
- (d) Rejection sampling

Answer: (b)

5. Which concept is central to decision-making under uncertainty?

- (a) Forward chaining
- (b) Utility theory
- (c) Symbolic inference
- (d) Structural recursion

Answer: (b)

Short Answer Questions

Q1. Define conditional independence in probabilistic terms. Provide an example.

Answer:

Two variables X and Y are conditionally independent given Z if:

$$P(X, Y | Z) = P(X | Z)P(Y | Z)$$

Example: Given the weather (Z), whether I carry an umbrella (X) and whether I get wet (Y) may be conditionally independent.

Q2. What is the difference between prior and posterior probabilities in Bayesian inference?

Answer:

The prior $P(H)$ is the belief in a hypothesis before observing data. The posterior $P(H | D)$ is the updated belief after observing data D , computed using Bayes' theorem.

Q3. List three applications of probabilistic reasoning in AI.

Answer:

- Spam filtering using Naïve Bayes
- Medical diagnosis using Bayesian networks
- Robot localization using particle filters

Q4. Explain the working principle of rejection sampling.**Answer:**

Rejection sampling draws candidate samples from a proposal distribution and accepts them with probability proportional to the target density. It is inefficient in high dimensions due to many rejected samples.

Q5. What is a causal network? How does it differ from a standard Bayesian network?**Answer:**

A causal network encodes not just statistical dependencies but also causal relationships. Interventions on variables (do-calculus) make sense in causal networks but not in standard Bayesian networks.

True/False

1. In a Bayesian network, cycles are allowed. **False**
2. Expectation in probability theory refers to the average value over many trials. **True**
3. Naïve Bayes assumes all features are independent. **False**
(Correction: Naïve Bayes assumes features are conditionally independent given the class.)
4. The Junction Tree algorithm is used for approximate inference. **False**
5. MCMC methods are useful when exact inference is intractable. **True**

Conceptual Discussion Question

Q: Explain how Bayesian networks model uncertainty. Describe the inference mechanism, and discuss challenges in using them for real-world applications.

Answer:

Bayesian networks represent probabilistic dependencies between variables using a DAG where each node is a random variable and edges encode direct influences. Each node has a conditional probability table (CPT) based on its parents. Inference involves computing marginal or conditional probabilities using algorithms such as variable elimination or belief propagation. Challenges include:

- Exponential growth of CPTs with many parents
- Computational intractability for dense or large graphs
- Need for accurate structure and parameter learning from limited data

They remain valuable tools due to their interpretability and strong foundation in probability theory.

Chapter 5: Probabilistic Reasoning and Uncertainty

Multiple Choice Questions (MCQs)

1. **What is the primary motivation for using probabilistic models in AI?**

- (a) To increase computational cost
- (b) To make decisions under deterministic assumptions
- (c) To handle uncertainty and incomplete information
- (d) To optimize symbolic reasoning

Answer: (c)

2. **Which of the following best describes conditional independence?**

- (a) Variables that are always independent
- (b) Variables that are dependent only when conditioned on another
- (c) Variables that are independent given a third variable
- (d) Variables with joint Gaussian distributions

Answer: (c)

3. **What is the key assumption in a Naïve Bayes model?**

- (a) All features are dependent on each other
- (b) Features are conditionally independent given the class
- (c) Classes are continuous variables
- (d) Data follows uniform distribution

Answer: (b)

4. Which inference technique approximates posterior distributions using random sampling?

- (a) Variable elimination
- (b) Junction tree algorithm
- (c) MCMC
- (d) Belief propagation

Answer: (c)

5. What is the key structure of a Bayesian Network?

- (a) A fully connected undirected graph
- (b) A sequence of temporal nodes
- (c) A directed acyclic graph (DAG) representing dependencies
- (d) A matrix of joint probabilities

Answer: (c)

Short Answer Questions

Q1. Define marginal and conditional probabilities.

Answer:

Marginal probability refers to the probability of a single event irrespective of other variables.

Conditional probability $P(A|B)$ refers to the probability of event A occurring given that B has occurred.

Q2. What are the advantages of using Bayesian networks in AI systems?

Answer:

Bayesian networks model joint distributions compactly, represent conditional independencies, and support both exact and approximate inference. They are interpretable and useful in diagnostic reasoning.

Q3. Differentiate between exact inference and approximate inference in probabilistic models.

Answer:

Exact inference computes the precise posterior distribution (e.g., via variable elimination), while approximate inference estimates it using sampling (e.g., MCMC) or heuristics when exact computation is intractable.

Q4. Describe the Naïve Bayes classification pipeline.

Answer:

Given a class variable and feature vector, it computes posterior probabilities using Bayes' Theorem under the independence assumption. The class with the highest posterior is assigned to the instance.

Q5. What are the key components of a causal network?

Answer:

Nodes representing variables, directed edges indicating causal influence, and conditional probability tables (CPTs) to define relationships.

True/False

1. Probabilistic models assume complete knowledge of all variables. **False**
2. Belief propagation works only on tree-structured graphs. **True**
3. In Naïve Bayes, features are assumed to be marginally independent. **False**
4. MCMC techniques are useful when exact inference is computationally hard. **True**
5. Bayesian inference updates prior beliefs using observed evidence. **True**

Conceptual Discussion Question

Q: Why is probabilistic reasoning essential in AI applications such as medical diagnosis and robotics? Illustrate with examples.

Answer:

Real-world environments are uncertain and noisy. In medical diagnosis, symptoms may be caused by multiple diseases with overlapping manifestations. Probabilistic models such as Bayesian networks can infer the most probable cause based on observed evidence. In robotics, sensor data is often incomplete or noisy—probabilistic filters (e.g., Kalman filters, particle filters) are essential to estimate position and control. These models allow AI systems to reason under uncertainty and update beliefs dynamically.

Chapter 6: Machine Learning Basics and Techniques

Multiple Choice Questions (MCQs)

1. Which of the following is NOT a typical phase in the machine learning pipeline?

- (a) Data preprocessing
- (b) Model evaluation
- (c) Neural coding
- (d) Feature selection

Answer: (c)

2. What does cross-validation aim to prevent?

- (a) Dataset duplication
- (b) Overfitting
- (c) Underflow errors
- (d) Data corruption

Answer: (b)

3. Which concept captures the tension between bias and variance in model performance?

- (a) Learning rate schedule
- (b) Decision surface complexity
- (c) Bias–variance tradeoff
- (d) K-means clustering

Answer: (c)

4. Which of the following best describes unsupervised learning?

- (a) Learning from labeled output
- (b) Learning without using input features
- (c) Learning from unlabelled data
- (d) Learning from a reinforcement signal

Answer: (c)

5. What is an inductive bias in machine learning?

- (a) A technique for removing class imbalance
- (b) An assumption about the hypothesis space to generalize from finite data
- (c) A measure of model fairness
- (d) A corruption in training labels

Answer: (b)

Short Answer Questions

Q1. Define supervised, unsupervised, and semi-supervised learning with one example each.

Answer:

- **Supervised:** Learning from labeled data (e.g., spam classification).
- **Unsupervised:** Learning patterns from unlabeled data (e.g., clustering customers).
- **Semi-supervised:** Learning from small labeled and large unlabeled data (e.g., image labeling).

Q2. What is overfitting? How can it be mitigated?

Answer:

Overfitting occurs when a model memorizes training data and fails to generalize. It can be mitigated via regularization, cross-validation, pruning, dropout (for neural nets), or early stopping.

Q3. Explain the difference between training error and test error.

Answer:

Training error is the error the model makes on the data it was trained on. Test error evaluates performance on unseen data and reflects the model's generalization ability.

Q4. What is the role of the hypothesis space in machine learning?

Answer:

The hypothesis space defines the set of all models or functions the learning algorithm can choose from. A good hypothesis space balances expressivity and generalization.

Q5. What is the purpose of model evaluation metrics? Name two examples.

Answer:

Model evaluation metrics quantify the performance of a learning algorithm. Examples include accuracy (for classification) and RMSE (for regression).

True/False

1. A high training accuracy always implies a good generalization performance. **False**
2. Bias–variance tradeoff guides the choice of model complexity. **True**
3. Underfitting occurs when a model is too simple to capture patterns. **True**
4. Semi-supervised learning assumes no labeled data is available. **False**
5. k -fold cross-validation divides the data into k subsets for reliable evaluation. **True**

Conceptual Discussion Question

Q: Discuss the bias–variance tradeoff in the context of model complexity. How does it influence underfitting and overfitting?

Answer:

Bias–variance tradeoff describes how increasing model complexity reduces bias but increases variance. A model with high bias oversimplifies the problem and underfits. A model with high variance captures noise and overfits. Optimal performance is achieved by balancing both—using regularization, cross-validation, and model selection techniques.

Chapter 7: Supervised Learning Techniques

Multiple Choice Questions (MCQs)

1. Which loss function is commonly used for linear regression?

- (a) Cross-entropy loss
- (b) Hinge loss
- (c) Mean squared error (MSE)
- (d) Kullback–Leibler divergence

Answer: (c)

2. What is the primary advantage of Support Vector Machines (SVMs)?

- (a) Efficient with missing values
- (b) Works well with small datasets and high-dimensional space
- (c) Requires no regularization
- (d) Always gives probabilistic outputs

Answer: (b)

3. Which of the following classifiers is based on Bayes' Theorem?

- (a) Decision Trees
- (b) Naïve Bayes
- (c) K-Nearest Neighbors
- (d) Support Vector Machines

Answer: (b)

4. Which ensemble technique combines weak learners sequentially to reduce bias?

- (a) Bagging
- (b) Stacking
- (c) Boosting
- (d) Dropout

Answer: (c)

5. Which method finds a linear combination of features that best separates multiple classes?

- (a) PCA
- (b) LDA
- (c) Logistic Regression
- (d) Ridge Regression

Answer: (b)

Short Answer Questions

Q1. What is the geometric interpretation of the SVM margin?

Answer:

The SVM margin is the perpendicular distance between the separating hyperplane and the closest data points (support vectors) from either class. Maximizing this margin leads to better generalization.

Q2. Differentiate between generative and discriminative models with examples.

Answer:

Generative models learn joint distribution $P(X, Y)$ and can generate data (e.g., Naïve Bayes). Discriminative models learn conditional distribution $P(Y|X)$ or decision boundaries (e.g., logistic regression, SVM).

Q3. How does pruning help in decision trees?

Answer:

Pruning removes branches that do not improve generalization, reducing overfitting and making the tree simpler and more robust.

Q4. Describe the k-NN algorithm briefly.

Answer:

k-NN classifies a new point by majority voting among its k nearest neighbors in the training data, based on a chosen distance metric.

Q5. What are the key differences between bagging and boosting?

Answer:

- Bagging trains models independently on bootstrap samples to reduce variance.
- Boosting trains models sequentially to correct previous errors, aiming to reduce bias.

True/False

1. Logistic regression is a linear model for binary classification. **True**
2. LDA maximizes class overlap for better separability. **False**
3. Decision trees are sensitive to small perturbations in data. **True**
4. SVMs require kernel functions to handle linear data. **False**
5. Ensemble methods often outperform individual weak learners. **True**

Conceptual Discussion Question

Q: Compare logistic regression and support vector machines in terms of decision boundaries, loss functions, and generalization.

Answer:

Logistic regression uses a sigmoid function and log-loss to estimate class probabilities. It minimizes a convex loss function for probabilistic interpretation. SVM constructs a maximum-margin hyperplane using hinge loss, focusing on support vectors. While logistic regression provides smooth probabilistic boundaries, SVMs offer better margins and can generalize well with fewer support points, especially in high-dimensional or linearly separable problems.

Chapter 8: Unsupervised Learning Techniques

Multiple Choice Questions (MCQs)

1. Which of the following algorithms assumes spherical clusters of equal variance?

- (a) DBSCAN
- (b) Hierarchical Clustering
- (c) K-Means
- (d) Gaussian Mixture Models

Answer: (c)

2. Which method is particularly effective at discovering arbitrarily shaped clusters?

- (a) K-Means
- (b) Principal Component Analysis
- (c) DBSCAN
- (d) t-SNE

Answer: (c)

3. Which of the following is a linear dimensionality reduction technique?

- (a) PCA
- (b) t-SNE
- (c) Autoencoder
- (d) UMAP

Answer: (a)

4. What does the Expectation-Maximization (EM) algorithm optimize?

- (a) Minimum spanning trees
- (b) Likelihood of observed data under latent variables
- (c) Posterior probability of outliers
- (d) Reconstruction error in autoencoders

Answer: (b)

5. In Hidden Markov Models (HMMs), what does the Viterbi algorithm compute?

- (a) Emission probabilities
- (b) Most probable sequence of hidden states
- (c) Transition matrix
- (d) Stationary distribution

Answer: (b)

Short Answer Questions

Q1. Describe the key limitation of K-Means clustering.

Answer:

K-Means assumes spherical, equally sized clusters and is sensitive to outliers and initialization. It may fail to detect non-convex shapes or varying density clusters.

Q2. What is the role of principal components in PCA?

Answer:

Principal components are orthogonal directions capturing maximum variance in the data. They form a reduced-dimensional basis that preserves the most significant structure.

Q3. How does DBSCAN differ from K-Means in its clustering approach?

Answer:

DBSCAN identifies dense regions separated by low-density space and can discover arbitrarily shaped clusters. It doesn't require the number of clusters as input and handles noise explicitly.

Q4. What are the two main steps of the EM algorithm?

Answer:

- **E-Step:** Estimate the expected value of latent variables given current parameters.
- **M-Step:** Maximize the expected log-likelihood to update parameters.

Q5. Explain the difference between hierarchical agglomerative and divisive clustering.

Answer:

Agglomerative clustering starts with each point as a singleton cluster and merges them iteratively. Divisive clustering starts with all points in one cluster and recursively splits them.

True/False

1. PCA maximizes the variance captured in the lower-dimensional space. **True**
2. DBSCAN requires the number of clusters as a hyperparameter. **False**
3. EM algorithm can be used to estimate parameters in Gaussian Mixture Models. **True**
4. HMMs assume observable states and hidden emissions. **False**
5. t-SNE is mainly used for visualization, not clustering. **True**

Conceptual Discussion Question

Q: Compare K-Means, DBSCAN, and Gaussian Mixture Models (GMMs) in terms of assumptions, strengths, and weaknesses.

Answer:

- **K-Means** assumes spherical clusters and minimizes within-cluster variance. It is efficient but sensitive to outliers and requires k .
- **DBSCAN** groups points based on density and handles outliers well. It can find non-convex clusters but struggles with varying densities.
- **GMMs** are probabilistic and model data as mixtures of Gaussians. They offer soft assignments but assume parametric distributions and require EM optimization.

Chapter 9: Reinforcement Learning

Multiple Choice Questions (MCQs)

1. Which of the following is a key characteristic of reinforcement learning (RL)?

- (a) Learning from labeled data
- (b) Supervised optimization
- (c) Learning through interaction with the environment
- (d) Predefined rule-based learning

Answer: (c)

2. What does the Bellman equation relate in an MDP?

- (a) Probabilities and priors
- (b) Expected return and immediate reward with future value
- (c) Discount factor and learning rate
- (d) Value iteration and stochastic gradient descent

Answer: (b)

3. Which of the following is a model-free RL algorithm?

- (a) Value Iteration
- (b) Policy Iteration
- (c) Q-Learning
- (d) Linear Programming

Answer: (c)

4. What is the key difference between SARSA and Q-learning?

- (a) SARSA is policy-based; Q-learning is value-based
- (b) SARSA updates based on current policy; Q-learning uses greedy future estimate
- (c) SARSA does not use rewards
- (d) Q-learning updates only after the full episode

Answer: (b)

5. What is the role of the discount factor γ in RL?

- (a) Controls learning rate
- (b) Penalizes incorrect actions
- (c) Balances immediate vs. future rewards
- (d) Modifies the reward signal

Answer: (c)

Short Answer Questions

Q1. Define a Markov Decision Process (MDP).

Answer:

An MDP is defined by a tuple (S, A, P, R, γ) , where S is the set of states, A is the set of actions, P is the transition probability function, R is the reward function, and γ is the discount factor.

Q2. Differentiate between value-based and policy-based methods.

Answer:

Value-based methods estimate the value function to derive the optimal policy (e.g., Q-learning). Policy-based methods learn the policy directly by optimizing expected return (e.g., REINFORCE).

Q3. Explain the exploration–exploitation tradeoff in RL.

Answer:

Exploration involves trying new actions to gather information, while exploitation uses known actions with high expected rewards. Balancing both is crucial for efficient learning.

Q4. What is temporal-difference (TD) learning?

Answer:

TD learning updates value estimates based on the difference between predicted and

observed rewards, using bootstrapping. It combines advantages of Monte Carlo and dynamic programming.

Q5. How do Deep Q-Networks (DQN) enhance traditional Q-learning?

Answer:

DQN uses neural networks to approximate Q-values and includes mechanisms like experience replay and target networks to stabilize training and handle large state spaces.

True/False

1. RL requires labeled supervision for training. **False**
2. Q-learning is an off-policy algorithm. **True**
3. SARSA always converges to the same policy as Q-learning. **False**
4. The discount factor γ close to 1 encourages long-term planning. **True**
5. TD learning can update values after every time step. **True**

Conceptual Discussion Question

Q: Discuss the advantages and challenges of deep reinforcement learning in real-world applications like robotics and game playing.

Answer:

Deep RL combines the representational power of neural networks with sequential decision-making. It enables agents to handle high-dimensional inputs (e.g., images in games or sensor data in robotics). However, it faces challenges like high sample complexity, instability in training, sensitivity to hyperparameters, and safety concerns in deployment. Solutions include reward shaping, exploration strategies, and transfer learning.

Chapter 10: Neural Networks and Deep Learning

Multiple Choice Questions (MCQs)

1. **What does the activation function in a neural network introduce?**

- (a) Linearity
- (b) Differentiability
- (c) Non-linearity
- (d) Orthogonality

Answer: (c)

2. **Which of the following is true about backpropagation?**

- (a) It trains weights layer-wise using SVD
- (b) It uses dynamic programming to update activation functions
- (c) It computes gradients using the chain rule
- (d) It only applies to output layers

Answer: (c)

3. **Which optimization technique helps escape saddle points in non-convex loss surfaces?**

- (a) Gradient clipping
- (b) Batch normalization
- (c) Momentum
- (d) Weight decay

Answer: (c)

4. What is the role of dropout in training deep neural networks?

- (a) Normalize inputs
- (b) Reduce overfitting by randomly deactivating neurons
- (c) Accelerate convergence
- (d) Increase weight sharing

Answer: (b)

5. Which architecture is primarily used for sequential data modeling?

- (a) Convolutional Neural Network (CNN)
- (b) Recurrent Neural Network (RNN)
- (c) Fully Connected Network
- (d) Support Vector Machine

Answer: (b)

Short Answer Questions**Q1. What is the vanishing gradient problem and in which networks is it most severe?**

Answer:

Vanishing gradients occur when gradients become too small for weight updates in early layers, hindering learning. It is most severe in deep and recurrent networks, especially with sigmoid or tanh activations.

Q2. Describe the mathematical form of the convolution operation in CNNs.

Answer:

The convolution operation for input x and kernel w is given by:

$$(y * w)[i] = \sum_j x[i + j] \cdot w[j]$$

In 2D, it extends to matrix convolution with strides and padding.

Q3. How does ReLU activation function help in deep learning?

Answer:

ReLU introduces non-linearity, avoids saturation in positive region, and accelerates convergence by mitigating vanishing gradients.

Q4. What is batch normalization and why is it used?**Answer:**

Batch normalization normalizes layer inputs to reduce internal covariate shift, stabilizing learning and enabling higher learning rates.

Q5. Compare CNNs and RNNs in terms of input structure and usage.**Answer:**

CNNs process grid-like data (e.g., images), using shared weights and local filters. RNNs process sequential data (e.g., text, time series) using recurrent connections to retain temporal information.

True/False

1. Neural networks with a single hidden layer can approximate any continuous function.
True
2. ReLU activation is bounded above and below. **False**
3. CNNs exploit translation invariance through local connectivity and weight sharing.
True
4. Gradient descent guarantees finding the global minimum for non-convex problems.
False
5. LSTMs help mitigate both vanishing and exploding gradients. **True**

Conceptual Discussion Question

Q: Discuss the training process of a deep neural network, highlighting the roles of forward pass, loss function, backpropagation, and optimization.

Answer:

The forward pass computes outputs using weights and activations layer by layer. The loss function quantifies the error between predicted and true labels. Backpropagation calculates gradients using the chain rule and propagates them backward. An optimizer (e.g., SGD, Adam) updates weights based on gradients to minimize the loss. Together, this iterative process trains the network.

Chapter 11: Advanced Machine Learning Techniques

Multiple Choice Questions (MCQs)

1. Which of the following techniques combines the predictions of multiple models to improve performance?

- (a) Feature selection
- (b) Dimensionality reduction
- (c) Ensemble learning
- (d) Transfer learning

Answer: (c)

2. Which ensemble method reduces variance by training models on random subsets of data?

- (a) Boosting
- (b) Bagging
- (c) Blending
- (d) Stacking

Answer: (b)

3. What is Bayesian optimization primarily used for in machine learning?

- (a) Model training
- (b) Hyperparameter tuning
- (c) Overfitting reduction
- (d) Data visualization

Answer: (b)

4. Which of the following is a key assumption in Bayesian learning?

- (a) Independence of errors
- (b) Frequentist hypothesis testing
- (c) Prior distribution over parameters
- (d) Infinite training data

Answer: (c)

5. Which technique is most appropriate for model compression?

- (a) Ensemble learning
- (b) Regularization
- (c) Pruning
- (d) Batch normalization

Answer: (c)

Short Answer Questions**Q1. What is the difference between bagging and boosting?**

Answer:

Bagging trains multiple models independently on bootstrapped subsets to reduce variance. Boosting trains models sequentially, correcting previous errors to reduce bias.

Q2. What is model stacking in ensemble learning?

Answer:

Stacking involves training a meta-model on the predictions of several base models. The base models learn the data, and the meta-model learns how to combine their outputs.

Q3. What is Bayesian inference and how is it used in ML?

Answer:

Bayesian inference updates a prior distribution over parameters with data (likelihood) to form a posterior. It allows uncertainty modeling and probabilistic prediction.

Q4. What is the objective of model pruning in neural networks?

Answer:

Pruning removes redundant or less important weights or neurons, reducing model size and inference cost while preserving accuracy.

Q5. Explain the concept of uncertainty in Bayesian deep learning.**Answer:**

Bayesian deep learning incorporates uncertainty in weights or outputs using posterior distributions. It improves robustness, especially in noisy or high-risk settings.

True/False

1. Boosting reduces bias more effectively than bagging. **True**
2. Stacking is more prone to overfitting than bagging. **True**
3. Bayesian optimization uses gradient descent for hyperparameter tuning. **False**
4. Dropout is a model compression technique. **False**
5. Bayesian models require a likelihood function and prior. **True**

Conceptual Discussion Question

Q: Discuss the advantages and challenges of ensemble learning. When should techniques like bagging, boosting, and stacking be applied?

Answer:

Ensemble learning improves accuracy, stability, and generalization. Bagging reduces variance and is effective with high-variance models (e.g., decision trees). Boosting reduces bias and is useful when weak learners can be improved incrementally. Stacking leverages diverse models but risks overfitting and requires careful validation. Challenges include increased complexity, interpretability loss, and computational cost.

Chapter 12: Natural Language Processing (NLP)

Multiple Choice Questions (MCQs)

1. Which of the following tasks is most directly associated with syntactic analysis in NLP?

- (a) Part-of-speech tagging
- (b) Named entity recognition
- (c) Parsing
- (d) Text classification

Answer: (c)

2. Which representation technique converts text into dense, real-valued vectors?

- (a) Bag-of-words
- (b) TF-IDF
- (c) Word embeddings (e.g., Word2Vec)
- (d) Regular expressions

Answer: (c)

3. Which technique is used to reduce words to their root form while preserving meaning?

- (a) Tokenization
- (b) Lemmatization
- (c) Stemming
- (d) Vectorization

Answer: (b)

4. What is the role of attention mechanisms in language models?

- (a) Discard irrelevant tokens
- (b) Encode word frequency
- (c) Weigh the importance of input tokens relative to each other
- (d) Compress sequences using pooling

Answer: (c)

5. Which of the following models is best suited for sequence-to-sequence tasks like machine translation?

- (a) CNN
- (b) SVM
- (c) Transformer
- (d) PCA

Answer: (c)

Short Answer Questions**Q1. Differentiate between stemming and lemmatization.**

Answer:

Stemming reduces words to their base form using heuristics, often truncating the word. Lemmatization uses linguistic knowledge to return the dictionary form of a word, preserving grammar and meaning.

Q2. What is tokenization in NLP?

Answer:

Tokenization is the process of splitting text into units such as words, phrases, or subwords (tokens). It is the first step in text preprocessing for most NLP pipelines.

Q3. Explain the role of positional encoding in transformers.

Answer:

Since transformers lack recurrence, positional encoding injects information about token positions in a sequence using sine and cosine functions or learnable embeddings.

Q4. What is named entity recognition (NER)? Provide examples.

Answer:

NER identifies and classifies named entities in text into categories like persons (e.g., "Einstein"), organizations (e.g., "UNESCO"), and locations (e.g., "Paris").

Q5. How is TF-IDF computed and what does it capture?

Answer:

TF-IDF (Term Frequency–Inverse Document Frequency) weights terms based on their importance. It increases with frequency in a document but decreases with occurrence across documents. It helps in distinguishing informative words.

True/False

1. Part-of-speech tagging assigns grammatical labels to tokens. **True**
2. Word2Vec produces sparse and interpretable vectors. **False**
3. Transformers use recurrence to capture long-range dependencies. **False**
4. BERT is a unidirectional language model. **False**
5. Text classification is a supervised learning task in NLP. **True**

Conceptual Discussion Question

Q: Discuss the advantages of transformer-based architectures over RNNs in NLP tasks. What are some limitations of transformers?

Answer:

Transformers offer parallel processing, improved long-range dependency capture, and attention mechanisms that weigh relevance across tokens. Unlike RNNs, they avoid sequential bottlenecks, enabling faster training. However, transformers require large memory, massive training data, and are sensitive to hyperparameters. Interpretability and computational efficiency remain open challenges.

Chapter 13: Data Mining and Predictive Analysis

Multiple Choice Questions (MCQs)

1. Which of the following best describes the goal of data mining?

- (a) Store large volumes of raw data
- (b) Extract patterns and knowledge from large datasets
- (c) Perform ETL operations
- (d) Train convolutional neural networks

Answer: (b)

2. Which task would be considered descriptive analysis?

- (a) Forecasting sales next quarter
- (b) Predicting customer churn
- (c) Identifying customer segments
- (d) Estimating future stock prices

Answer: (c)

3. Which algorithm is typically used for association rule mining?

- (a) Apriori
- (b) SVM
- (c) K-Means
- (d) Q-Learning

Answer: (a)

4. Which metric evaluates the strength of an association rule?

- (a) Recall
- (b) Confidence
- (c) Silhouette score
- (d) MSE

Answer: (b)

5. Which technique is commonly used in predictive modeling of time series data?

- (a) PCA
- (b) Apriori
- (c) ARIMA
- (d) TF-IDF

Answer: (c)

Short Answer Questions

Q1. Differentiate between predictive and descriptive data analysis.

Answer:

Predictive analysis forecasts future outcomes (e.g., sales forecasting), while descriptive analysis summarizes historical data to understand patterns (e.g., clustering customer behavior).

Q2. What is a frequent itemset in association rule mining?

Answer:

A frequent itemset is a set of items that appears together in a dataset with frequency above a specified support threshold.

Q3. Explain the role of confidence in association rules.

Answer:

Confidence measures the likelihood that the consequent of a rule is present given the antecedent. It is defined as:

$$\text{confidence}(A \Rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)}$$

Q4. List and briefly describe two types of clustering techniques.

Answer:

- **K-Means:** Partitions data into k clusters by minimizing within-cluster variance.
- **Hierarchical Clustering:** Builds nested clusters using a bottom-up (agglomerative) or top-down (divisive) approach.

Q5. How is time series analysis different from standard regression?

Answer:

Time series analysis models temporal dependencies and autocorrelation in sequential data, while standard regression assumes independent observations.

True/False

1. Predictive modeling uses historical data to forecast future outcomes. **True**
2. The Apriori algorithm is suitable for high-dimensional time series prediction. **False**
3. Descriptive analytics includes clustering and association rule mining. **True**
4. Support is always greater than confidence in association rules. **False**
5. Time series models assume temporal ordering of data. **True**

Conceptual Discussion Question

Q: Discuss the process of association rule mining using the Apriori algorithm. Include its strengths and limitations.

Answer:

Apriori identifies frequent itemsets using a level-wise search, pruning those below a support threshold. Rules are generated from frequent sets with confidence exceeding a threshold. It is interpretable and efficient on small datasets but suffers from high computational cost in dense or large datasets due to candidate explosion. FP-Growth addresses this by avoiding explicit candidate generation.

Chapter 14: Data Warehousing and Business Intelligence

Multiple Choice Questions (MCQs)

1. Which of the following best defines a data warehouse?

- (a) A distributed file storage system
- (b) A system for real-time data processing
- (c) A subject-oriented, integrated, time-variant, and non-volatile collection of data
- (d) A temporary data staging server

Answer: (c)

2. Which process extracts, transforms, and loads data into a warehouse?

- (a) OLAP
- (b) ETL
- (c) SQL
- (d) OLTP

Answer: (b)

3. What is the main use of OLAP in business intelligence?

- (a) Online data entry
- (b) Operational transaction processing
- (c) Multidimensional analytical queries
- (d) Flat file storage

Answer: (c)

4. Which schema uses one central fact table connected to dimension tables?

- (a) Snowflake schema
- (b) Star schema
- (c) ER model
- (d) Fact constellation

Answer: (b)

5. Which of the following allows analysis at multiple levels of aggregation in a data cube?

- (a) Slicing and Dicing
- (b) Drill-down and Roll-up
- (c) Normalization
- (d) Pivoting

Answer: (b)

Short Answer Questions**Q1. Define OLAP and differentiate it from OLTP.**

Answer:

OLAP (Online Analytical Processing) supports complex analytical queries for decision-making, while OLTP (Online Transaction Processing) handles high-volume, transactional updates. OLAP is read-intensive and multidimensional; OLTP is write-intensive and relational.

Q2. What are the advantages of star schema in data warehouse design?

Answer:

Star schema offers simplicity, faster query performance, and efficient joins. It organizes data into a central fact table linked to denormalized dimension tables.

Q3. Explain the difference between a data lake and a data warehouse.

Answer:

A data warehouse stores structured, curated, and integrated data for analytical queries, whereas a data lake stores raw, unstructured, semi-structured, and structured data at scale, often for later processing.

Q4. What is drill-down in OLAP operations?**Answer:**

Drill-down is an OLAP operation that moves from higher-level summaries to more detailed levels of data, such as going from yearly to monthly sales.

Q5. What is the role of ETL in business intelligence?**Answer:**

ETL (Extract, Transform, Load) processes raw data from diverse sources, cleans and integrates it, and loads it into a data warehouse for querying and analysis.

True/False

1. Data warehouses are optimized for high-frequency transactional processing. **False**
2. OLAP enables multidimensional data analysis. **True**
3. Fact tables contain descriptive attributes only. **False**
4. A snowflake schema normalizes dimension tables. **True**
5. Real-time BI systems update warehouse contents continuously. **True**

Conceptual Discussion Question

Q: Discuss the components and architecture of a modern business intelligence (BI) system. How does it support strategic decision-making?

Answer:

A BI system consists of ETL pipelines, data warehouses or lakes, OLAP engines, dashboards, and reporting tools. It ingests data from transactional systems, processes it into structured formats, and enables interactive analysis. Real-time BI integrates with streaming sources. These components support KPIs, trends, and forecasts that inform strategic, tactical, and operational decisions across an enterprise.

Chapter 15: Ethics in AI and Future Trends

Multiple Choice Questions (MCQs)

1. Which of the following is a major ethical concern in AI systems?

- (a) Data compression
- (b) Hardware acceleration
- (c) Algorithmic bias
- (d) High-dimensional vectors

Answer: (c)

2. Which concept refers to making AI decisions understandable to humans?

- (a) Differential privacy
- (b) Black-box modeling
- (c) Explainability
- (d) Hyperparameter tuning

Answer: (c)

3. Which domain requires special attention to fairness in AI due to societal impact?

- (a) Image compression
- (b) Game simulation
- (c) Criminal justice
- (d) Web scraping

Answer: (c)

4. What is the goal of General Artificial Intelligence (AGI)?

- (a) Solve only domain-specific tasks
- (b) Imitate sensory input
- (c) Exhibit human-level cognitive ability across domains
- (d) Compress language models efficiently

Answer: (c)

5. **Which emerging technology is expected to accelerate AI algorithms through quantum parallelism?**

- (a) Graph analytics
- (b) Quantum computing
- (c) Blockchain
- (d) Federated learning

Answer: (b)

Short Answer Questions

Q1. Define algorithmic fairness and give one example.

Answer:

Algorithmic fairness ensures that AI systems do not systematically disadvantage individuals or groups. For example, a loan approval algorithm should not reject applicants based on race or gender if irrelevant to creditworthiness.

Q2. What is transparency in the context of AI systems?

Answer:

Transparency refers to the degree to which the internal mechanisms of an AI system are open to inspection, interpretation, and understanding by humans, including stakeholders and regulators.

Q3. What is meant by “bias” in AI, and how does it arise?

Answer:

Bias in AI refers to systematic errors or unfairness in predictions due to imbalanced or prejudiced training data, flawed model design, or skewed evaluation criteria.

Q4. How can AI contribute to sustainability and climate action?

Answer:

AI can optimize energy usage, forecast climate patterns, manage smart grids, design efficient logistics, and support scientific research in sustainability domains.

Q5. Mention two concerns about the development of AGI.

Answer:

- Loss of human control over autonomous systems.
- Risk of misalignment between AI goals and human values.

True/False

1. Explainable AI aims to improve user trust. **True**
2. Federated learning increases centralization of training data. **False**
3. Bias in AI systems is only a result of biased training data. **False**
4. Quantum computing can improve the efficiency of AI search and optimization. **True**
5. Fairness metrics are universally applicable across all domains. **False**

Conceptual Discussion Question

Q: What ethical frameworks guide the responsible development of AI? Discuss at least three principles with examples.

Answer:

Ethical AI frameworks are based on principles such as:

- **Fairness:** AI should not discriminate unfairly. For example, ensuring hiring algorithms don't prefer certain genders.
- **Accountability:** Developers and institutions must be answerable for AI behavior. For instance, facial recognition errors must be traceable.
- **Transparency:** Systems should be explainable. Black-box models in healthcare must justify decisions to practitioners and patients.

Additional principles include privacy, safety, inclusiveness, and human oversight. These form the foundation for governance and public trust in AI.

Appendix A: Linear Algebra

Multiple Choice Questions (MCQs)

1. Which of the following properties must a set satisfy to be considered a vector space?

- (a) Closed under matrix inversion
- (b) Closed under scalar multiplication and vector addition
- (c) Contains orthogonal vectors
- (d) Has unit norm elements

Answer: (b)

2. What does the determinant of a matrix represent geometrically (in 2D or 3D)?

- (a) Maximum eigenvalue
- (b) Volume scaling factor of linear transformation
- (c) Matrix trace
- (d) Angle between row vectors

Answer: (b)

3. What condition indicates that a square matrix is invertible?

- (a) It has zero determinant
- (b) It is symmetric
- (c) It has full rank
- (d) Its trace is positive

Answer: (c)

4. Which decomposition expresses a matrix $A \in \mathbb{R}^{m \times n}$ as $U\Sigma V^T$?

- (a) QR Decomposition
- (b) Eigenvalue Decomposition
- (c) LU Decomposition
- (d) Singular Value Decomposition (SVD)

Answer: (d)

5. Which of the following is always true for an orthogonal matrix Q ?

- (a) $QQ^T = I$
- (b) Q is symmetric
- (c) Q is lower triangular
- (d) $Q^{-1} = -Q$

Answer: (a)

Short Answer Questions

Q1. Define the rank of a matrix and explain its significance.

Answer:

The rank of a matrix is the maximum number of linearly independent rows or columns. It indicates the dimension of the row space (or column space) and determines whether a system of equations has a unique solution.

Q2. State and explain the geometric interpretation of eigenvectors and eigenvalues.

Answer:

An eigenvector of a matrix is a nonzero vector whose direction remains unchanged under the linear transformation, and the eigenvalue is the scalar by which it is stretched or compressed. Geometrically, they represent principal axes of transformation.

Q3. What is the Moore–Penrose pseudo-inverse and when is it used?

Answer:

The pseudo-inverse generalizes the matrix inverse for non-square or singular matrices and is used to compute least-squares solutions to linear systems when a unique solution does not exist.

Q4. Explain the role of projections in linear algebra.

Answer:

A projection maps a vector onto a subspace such that the error (difference) is orthogonal to the subspace. It is used in least-squares approximation and dimensionality reduction.

Q5. What are the conditions for diagonalizability of a matrix?

Answer:

A matrix is diagonalizable if it has enough linearly independent eigenvectors to form a basis, i.e., it can be written as $A = PDP^{-1}$, where D is diagonal.

True/False

1. Every symmetric matrix has real eigenvalues. **True**
2. If a matrix is orthogonal, its determinant is always 0. **False**
3. The nullity of a matrix is the dimension of its null space. **True**
4. All square matrices are diagonalizable. **False**
5. The Frobenius norm is invariant under orthogonal transformations. **True**

Conceptual Discussion Question

Q: Describe the purpose and applications of Singular Value Decomposition (SVD) in machine learning and numerical linear algebra.

Answer:

SVD decomposes any matrix into orthogonal matrices U , V and a diagonal matrix Σ . It enables dimensionality reduction (e.g., PCA), noise filtering, image compression, and solving ill-posed systems. In ML, it underlies techniques like latent semantic analysis and collaborative filtering.

Appendix B: Probability and Statistics

Multiple Choice Questions (MCQs)

1. What is the expected value of a discrete random variable?

- (a) Its mode
- (b) Its standard deviation
- (c) The average outcome weighted by probability
- (d) The maximum observed value

Answer: (c)

2. Which of the following distributions is discrete?

- (a) Normal distribution
- (b) Exponential distribution
- (c) Bernoulli distribution
- (d) Beta distribution

Answer: (c)

3. What is the variance of a random variable X ?

- (a) $\mathbb{E}[X]^2$
- (b) $\mathbb{E}[X - \mu]$
- (c) $\mathbb{E}[(X - \mu)^2]$
- (d) $\sqrt{\mathbb{E}[X]}$

Answer: (c)

4. What is the covariance of two independent random variables?

- (a) 1
- (b) 0
- (c) Undefined
- (d) Equal to their product

Answer: (b)

5. Which theorem justifies using the normal distribution for averages of i.i.d. variables?

- (a) Bayes' Theorem
- (b) Law of Large Numbers
- (c) Central Limit Theorem
- (d) Chebyshev's Inequality

Answer: (c)

Short Answer Questions

Q1. Define conditional probability and provide an example.

Answer:

Conditional probability is the probability of an event A given another event B :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (\text{assuming } P(B) > 0)$$

Example: Given it is raining, the chance someone carries an umbrella.

Q2. State Bayes' Theorem. What is its significance?

Answer:

Bayes' Theorem relates conditional probabilities:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

It allows updating beliefs based on new evidence and is foundational in Bayesian inference.

Q3. What is the difference between population parameters and sample statistics?

Answer:

Population parameters describe entire populations (e.g., true mean μ), while sample statistics (e.g., \bar{x}) are computed from a subset and used to estimate parameters.

Q4. Explain the concept of hypothesis testing.**Answer:**

Hypothesis testing evaluates a null hypothesis H_0 against an alternative H_1 using sample data and a test statistic. Based on a p-value or critical region, we reject or fail to reject H_0 .

Q5. What does correlation measure? How is it different from causation?**Answer:**

Correlation quantifies the linear relationship between variables. Causation implies one variable directly affects another, which correlation alone does not confirm.

True/False

1. The probability of a union of mutually exclusive events is the sum of their individual probabilities. **True**
2. The expectation of the sum of random variables is equal to the sum of expectations. **True**
3. Two events with zero covariance are always independent. **False**
4. The standard deviation is the square of the variance. **False**
5. Confidence intervals give a range of plausible values for a population parameter. **True**

Conceptual Discussion Question

Q: Discuss the Central Limit Theorem (CLT) and its role in inferential statistics. How does it enable confidence interval construction and hypothesis testing?

Answer:

The CLT states that the sampling distribution of the sample mean approaches a normal distribution as the sample size grows, regardless of the population distribution, provided variance is finite. This underpins many statistical procedures: it justifies approximating sampling errors using the normal curve, enabling the construction of confidence intervals and hypothesis testing based on z- or t-statistics, even when the underlying population is not normal.

Appendix C: Geometry and Multivariate Calculus

Multiple Choice Questions (MCQs)

1. Which of the following is true about Euclidean geometry?

- (a) Parallel lines always intersect
- (b) The sum of angles in a triangle is less than 180°
- (c) It assumes flat space with the parallel postulate
- (d) It is defined on curved manifolds

Answer: (c)

2. The gradient of a scalar field points in the direction of:

- (a) Maximum decrease
- (b) Random motion
- (c) Level curves
- (d) Maximum increase

Answer: (d)

3. Which of the following statements is true about convex sets?

- (a) Every point in a convex set has a unique tangent
- (b) The intersection of two convex sets is not convex
- (c) Any line segment connecting two points in the set lies entirely within the set
- (d) Convex sets must be closed and bounded

Answer: (c)

4. Which term describes a space that locally resembles Euclidean space but may be curved globally?

- (a) Vector space
- (b) Affine subspace
- (c) Manifold
- (d) Hyperplane

Answer: (c)

5. Which operation gives the total volume under a multivariate surface?

- (a) Gradient
- (b) Curl
- (c) Divergence
- (d) Multiple integral

Answer: (d)

Short Answer Questions

Q1. What is a tangent space in the context of differentiable manifolds?

Answer:

A tangent space at a point on a differentiable manifold is a vector space that linearly approximates the manifold near that point. It consists of all possible directions in which one can tangentially pass through the point.

Q2. Define a convex function and give one property that characterizes it.

Answer:

A function f is convex if the line segment between any two points on its graph lies above or on the graph. A key property:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \text{for } \lambda \in [0, 1]$$

Q3. What is the Hessian matrix and what does it represent?

Answer:

The Hessian is a square matrix of second-order partial derivatives. It provides local curvature information and helps classify critical points of a multivariate function.

Q4. Explain the difference between partial and total derivatives.**Answer:**

Partial derivatives take derivatives with respect to one variable, holding others constant. Total derivatives consider all dependencies via the chain rule.

Q5. How is the Jacobian matrix used in transformations?**Answer:**

The Jacobian represents the best linear approximation to a differentiable vector-valued function. It is used to change variables in multiple integrals and understand how volume scales under transformation.

True/False

1. A surface with zero Gaussian curvature is locally flat. **True**
2. A function with a positive definite Hessian at a point has a local maximum there. **False**
3. The divergence of a vector field measures its outward flux. **True**
4. All manifolds are Euclidean spaces. **False**
5. Convex geometry is crucial in optimization theory. **True**

Conceptual Discussion Question

Q: Explain how multivariate calculus and differential geometry support modern machine learning techniques such as optimization, backpropagation, and geometric deep learning.

Answer:

Multivariate calculus enables gradient-based optimization methods (e.g., SGD) by providing tools like gradients and Hessians. Backpropagation is an application of the chain rule in vector calculus. Differential geometry contributes to understanding curved data manifolds, leading to geometric deep learning approaches that generalize neural networks to non-Euclidean domains like graphs and manifolds.