Code No: 156BN **R18**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech III Year II Semester Examinations, August - 2022

MACHINE LEARNING

(Computer Science and Engineering)

Time: 3 Hours **Max.Marks:75**

Answer any five questions

All questions carry equal marks

- - -

1. a) Which disciplines have their influence on machine learning? Explain with examples.
b) What are the different types of Machine Learning models? [8+7]

2. a) List the problems that can be solved using machine learning.
b) Discuss the issues in the decision tree learning algorithm in detail. [8+7]

3. a) Explain back-propagation algorithm in detail.
b) Explain the following:
i) General consistent hypothesis.
ii) Closed concepts in the path through the hypothesis. [7+8]

4. a) Discuss the issues related to neural network learning.
b) Write a detailed note on sampling theory. [8+7]

5. a) Describe the Naive Bayesian method of classification. What assumptions does this method make about the ribues and the classification? Give an example where this assumption is justified.
b) What is the Laplacian correction and why it is necessary? [10+5]

6.a) Write the differences between the Eager Learning and Lazy Learning approaches.
b) State Bayes theorem. Illustrate Bayes theorem with an example. [7+8]

7. a) Write the basic algorithm for learning sets of First-Order Rules.
b) Apply inverse resolution in propositional form to the clauses C=A B, C1=A B G. Give at least two possible results for C2. [7+8]

8. What are the differences between inductive learning and analytical learning problems and explain the same with an example. [15]

---ooOoo---

ANSWER KEY

1.a) Which disciplines have their influence on machine learning? Explain with examples.

1. Statistics: Provides foundational methods for data analysis and inference, crucial for building and evaluating models (e.g., regression analysis).
2. Computer Science: Offers algorithms, computational complexity theory, and data structures, essential for designing efficient learning algorithms.
3. Mathematics: Underpins ML with linear algebra, calculus, and probability theory, used in model formulation and optimization.
4. Artificial Intelligence: AI's quest for replicating human intelligence has driven many ML applications and algorithms (e.g., neural networks).
5. Neuroscience: Influences ML through understanding of the brain and neural networks, inspiring models like artificial neural networks.
6. Psychology: Provides insights into cognitive processes, informing algorithms that mimic human learning patterns.
7. Linguistics: Contributes to natural language processing, an ML application for understanding and generating human language.
8. Philosophy: Offers perspectives on logic and reasoning, influencing ML in areas like knowledge representation.
9. Physics: Has inspired techniques in ML, such as simulated annealing, derived from thermodynamics.
10. Economics: Offers models of decision-making and risk, which are integral in reinforcement learning.

1. b) What are the different types of Machine Learning models?

1. Supervised Learning: Models are trained on labeled data (e.g., regression, classification).
2. Unsupervised Learning: Involves learning patterns from unlabeled data (e.g., clustering, association).
3. Semi-supervised Learning: Uses both labeled and unlabeled data for training, often used with limited labeled data.
4. Reinforcement Learning: Learns to make decisions by receiving rewards or penalties (e.g., Q-learning).
5. Deep Learning: Uses neural networks with many layers, effective for high-dimensional data like images and audio.
6. Probabilistic Models: Include Bayesian networks, which model relationships between variables with probabilities.
7. Ensemble Methods: Combine predictions from multiple models to improve accuracy (e.g., Random Forest, Boosting).
8. Dimensionality Reduction: Techniques like PCA reduce the number of variables in data while preserving important information.

9. Evolutionary Algorithms: Use evolutionary processes like mutation and selection to optimize models.
10. Instance-based Learning: Make predictions based on similarities with known instances (e.g., k-Nearest Neighbors).

2.a) List the problems that can be solved using machine learning.

1. Image Recognition: Identifying objects, people, or actions in images.
2. Speech Recognition: Transcribing spoken language into text.
3. Natural Language Processing: Includes tasks like language translation, and sentiment analysis.
4. Medical Diagnosis: Analyzing medical data to assist in diagnosis.
5. Stock Market Analysis: Predicting stock prices or market trends.
6. Fraud Detection: Identifying fraudulent activities in finance.
7. Recommendation Systems: Suggesting products or content to users.
8. Self-driving Cars: Autonomous vehicles using ML for navigation.
9. Game Playing: AI learning to play and win games (e.g., Chess, Go).
10. Predictive Maintenance: Forecasting equipment failures in industries.

2.b) Discuss the issues in the decision tree learning algorithm in detail.

1. Overfitting: Trees can become too complex, fitting noise in the data.
2. Handling Continuous Variables: Requires discretization, which can lose information.
3. Handling Missing Data: Decision trees struggle with incomplete datasets.
4. Scalability: Large datasets can lead to very large trees, increasing computational complexity.
5. Decision Boundary Limitations: Trees create perpendicular decision boundaries, which may not always capture the true data distribution.
6. Non-Robustness: Small changes in data can lead to very different trees.
7. Bias in Tree Construction: Tendency to prefer features with more levels.
8. Pruning Strategies: Determining when to stop tree growth is not straightforward.
9. Interpretability Issues: While generally interpretable, very large trees can be hard to understand.
10. Algorithmic Bias: Can inherit biases present in the training data.

3. a) Explain the back-propagation algorithm in detail.

1. Definition: Back-propagation is a supervised learning algorithm used for training artificial neural networks.
2. Function: It adjusts the weights of the neural network through a process that minimizes the difference between actual output and desired output.
3. Forward Pass: Initially, inputs are fed into the network to produce an output.
4. Error Calculation: The error is calculated by comparing the network's output with the actual expected output.

5. Backward Pass: The error is propagated back through the network, allowing the algorithm to adjust the weights.
6. Gradient Descent: Uses gradient descent to find the minimum of the error function.
7. Learning Rate: This involves a parameter called learning rate, which controls how much the weights are adjusted during training.
8. Iterative Process: The process is repeated for many iterations or until the error is reduced to an acceptable level.
9. Multi-layer Networks: Particularly effective for multi-layer networks (more than one hidden layer).
10. Applications: Widely used in applications like image and speech recognition, and other predictive models.

3.b) .i) Explain the General consistent hypothesis.
1. Definition: A hypothesis is considered generally consistent if it agrees with all the examples in the training set.
2. Flexibility: It is not overly specific to a few examples but is general enough to apply to the entire dataset.
3. Inclusiveness: Includes the widest range of instances that are consistent with the training data.
4. Generalization: Aims to generalize from the training data to unseen instances.
5. Robustness: Such hypotheses are often more robust to noise and exceptions in the data.
6. Learning Algorithms: Often used in algorithms that aim to find a balance between overfitting and underfitting.
7. Simplicity: Prefers simpler hypotheses to complex ones (Occam's Razor principle).
8. Predictive Performance: Generally consistent hypotheses are expected to perform well on unseen data.
9. Inductive Bias: Reflects the inductive bias of the learning algorithm.
10. Use in Machine Learning: Essential in machine learning for creating models that generalize well.

3. b).ii) Closed concepts in the path through the hypothesis.
1. Definition: Closed concepts represent a set of hypotheses that are complete and consistent with the observed data.
2. Path through Hypothesis Space: Represents a trajectory in the hypothesis space that progressively refines the hypothesis.
3. Concept Learning: Closed concepts are integral in concept learning, where the goal is to identify the underlying rule or concept.
4. Completeness: Ensures that the hypotheses cover all the possible instances that fit the concept.
5. Consistency: Maintains consistency with the examples encountered.

6. Closure Property: These concepts adhere to the closure property under the operations of the learning algorithm.
7. Refinement Operators: Utilizes refinement operators to move from general to specific hypotheses or vice versa.
8. Search Strategy: The path is often determined by the search strategy of the learning algorithm.
9. Convergence: Aims at converging to the most accurate hypothesis that describes the data.
10. Handling Ambiguity: Helps in handling ambiguities and uncertainties inherent in the learning process.

4. a) Discuss the issues related to neural network learning.

1. Overfitting: Neural networks can overfit the training data, failing to generalize to new data.
2. Data Requirements: Require large amounts of data for training to achieve high performance.
3. Computational Cost: Training, especially deep neural networks, is computationally intensive.
4. Local Minima: The learning process can get stuck in local minima, leading to suboptimal solutions.
5. Interpretability: Neural networks, particularly deep networks, are often seen as 'black boxes' with low interpretability.
6. Parameter Tuning: Involves a lot of hyperparameter tuning (like learning rate, number of layers, etc.).
7.Data Preprocessing: Requires careful preprocessing of data (normalization, handling missing values, etc.).
8. Vanishing/Exploding Gradients: Problems can occur during training due to vanishing or exploding gradients.
9. Dependency on Architecture: The performance heavily depends on the choice of network architecture.
10. Generalization vs. Memorization: Balancing between learning the training data patterns and memorizing the data.

4. b) Write a detailed note on sampling theory.

1. Definition: Sampling theory is a study that explains how to select a subset (sample) of individuals from a population to estimate the characteristics of the whole population.
2. Representativeness: Stresses on the representativeness of the sample for unbiased estimation.
3. Sampling Error: Discusses the concept of sampling error and how to minimize it
4. Sample Size: Addresses how to determine the appropriate sample size for reliable results.

5. Random Sampling: Emphasizes the importance of random sampling to avoid bias.
6. Statistical Inference: Forms the basis for statistical inference, allowing for predictions about the population from the sample.
7. Types of Sampling: Covers different types of sampling methods (e.g., stratified, cluster, systematic).
8. Central Limit Theorem: Relies on the Central Limit Theorem for making inferences about population parameters.
9. Data Quality: Highlights the impact of sample quality on the reliability of the conclusions.
10. Applications: Essential in fields like statistics, research methodology, and data analysis for designing experiments and surveys.

5. a) Describe the Naive Bayesian method of classification. What assumptions does this method make about the ribues and the classification? Give an example where this assumption is justified.

1. Definition: Naive Bayes is a probabilistic machine learning model used for classification tasks. It applies Bayes' theorem with the 'naive' assumption of conditional independence between every pair of features.
2. Assumptions: The primary assumption is that the attributes are conditionally independent given the class. This means the effect of an attribute value on a given class is independent of the values of other attributes.
3. Advantage of Simplicity: This assumption simplifies the computation, making Naive Bayes a highly efficient and easy-to-implement method.
4. Probability Calculation: It calculates the posterior probability of each class, based on the input features.
5. Use in Real-World Applications: Despite its simplicity, Naive Bayes can outperform more complex models when the data set isn’t large enough to train them.
6. Text Classification Example: A classic example is spam filtering in emails, where the presence or absence of certain words (features) are used to classify an email as 'spam' or 'not spam'.
7. Assumption Justification in Spam Filtering: In spam filtering, the assumption that words (attributes) occur independently in emails is reasonably justified, since most spam keywords don't depend on each other.
8. Performance Dependence: Its performance is dependent on the representation of the input features and the assumption of independence.
9. Limitation: The independence assumption can be a limitation when attributes are not actually independent, potentially leading to less accurate results.
10. Widespread Usage: Despite its simplicity, it's widely used due to its effectiveness in large datasets with many features, such as text classification.

5. b) What is the Laplacian correction and why it is necessary?

1. Problem Addressed: Laplacian correction (or Laplace smoothing) addresses the issue of zero probability in Naive Bayes classifiers.
2. Zero Probability Issue: Without correction, if a categorical variable has a category in the test data set that was not present in the training data set, it would assign a 0 probability and hence, an incorrect prediction.
3. Method: It adds a small number (usually 1) to each count to avoid zero probabilities.
4. Normalization: Adjusts the resultant probabilities to ensure they sum up to 1.
5. Usefulness: This is crucial for dealing with new features not seen in training, ensuring the model remains flexible and robust.

6. a) Write the differences between the Eager Learning and Lazy Learning approaches.

1. Model Construction: Eager learning constructs a model before seeing the test data, while lazy learning doesn't construct a model until it sees the test data.
2. Examples: Decision trees and neural networks are eager learners; k-NN and case-based reasoning are lazy learners.
3. Computation Time: Eager learners spend more time in training and less in prediction, while lazy learners are the opposite.
4. Memory Usage: Lazy learners typically require more memory as they store the entire dataset.
5. Adaptability: Lazy learners can adapt to changes quickly as they don't have a predefined model.
6. Data Requirements: Eager learners may perform better when the amount of data is limited.
7. Real-Time Data: Lazy learning is more suitable for applications where data is continuously changing.

6. b) State Bayes theorem. Illustrate Bayes theorem with an example.

1. Bayes Theorem Formula: P(A|B) = [P(B|A) * P(A)] / P(B).
2. Explanation: It describes the probability of an event, based on prior knowledge of conditions that might be related to the event.
3. Components:
   P(A|B) is the posterior probability.
   P(B|A) is the likelihood.
   P(A) is the prior probability.
   P(B) is the marginal probability.
4. Application Example: Medical Diagnosis.
5. Example Scenario: Determining the probability of a disease (A) given a symptom (B).
6. Calculation in Example: Use known probabilities (like the prevalence of the disease and the likelihood of the symptom given the disease) to calculate the probability of having the disease given that the symptom is present.

7. Importance in Machine Learning: It's a foundational theorem used in a variety of machine learning algorithms, especially in probabilistic models.
8. Versatility: Useful in many fields including statistics, economics, and engineering.

7. a) Write the basic algorithm for learning sets of First-Order Rules.
1. Initialization: Start with an empty set of rules.
2. Rule Generation: Generate candidate rules from the given data. This involves identifying potential relationships among attributes.
3. Hypothesis Space Search: Navigate through the space of possible rules, often using heuristic methods to guide the search.
4. Refinement: Refine each rule by adding conditions to reduce errors like overfitting or underfitting.
5. Evaluation: Assess the quality of the rules using metrics like accuracy, precision, recall, or F1 score.
6. Pruning: Remove rules that do not contribute significantly to the model's performance or that are redundant.
7. Stopping Criteria: Determine when to stop the learning process, which could be based on a maximum number of rules, convergence of accuracy, or computational limits.
8. Generalization: Ensure the rules are general enough to apply to new, unseen data, not just the training set.
9. Optimization: Optimize the rule set for efficiency and effectiveness, possibly using techniques like genetic algorithms.
10. Output: Finalize and output the set of learned first-order rules.

7. b) Apply inverse resolution in propositional form to the clauses C=A B, C1=A B G. Give at least two possible results for C2.
Given Clauses:
$C = A \wedge B$
$C1 = A \wedge B \wedge G$
Applying inverse resolution to derive C2:

1. Resolution Technique: Use resolution techniques to find a clause C2 that, when combined with another clause, results in C or C1.
Possible C2: Combine C with C2 to form C1.
Example: If C2 = G, then C2 combined with C ($A \wedge B$) results in C1 ($A \wedge B \wedge G$).
2. Inverse Operation: Perform an inverse operation on the clauses.
Possible C2: Deduce what C2 must be to reach C1 from C.
Example: Since $C1 = C \wedge G$, C2 could be G.

8. What are the differences between inductive learning and analytical learning problems and explain the same with an example.

1. Definition:
   Inductive Learning: Learns general rules from specific instances.
   Analytical Learning: Uses background knowledge and logical reasoning.
2. Approach:
   Inductive: Empirical, based on observation.
   Analytical: Deductive, based on theory and reasoning.
3. Data Dependency:
   Inductive: Highly dependent on data.
   Analytical: Relies more on pre-existing knowledge.
4. Learning Process:
   Inductive: Generalizes from examples.
   Analytical: Uses problem-solving techniques.
5. Examples:
   Inductive: Learning to classify animals based on features.
   Analytical: Solving a geometry problem using theorems.
6. Knowledge Representation:
   Inductive: Often uses statistical models.
   Analytical: Uses logical representations.
7. Flexibility:
   Inductive: Adapts to new data easily.
   Analytical: More rigid, based on fixed rules.
8. Use Cases:
   Inductive: Suitable for pattern recognition.
   Analytical: Ideal for systems requiring logical explanation.
9. Uncertainty Handling:
   Inductive: Handles uncertain and noisy data well.
   Analytical: Struggles with uncertainty.
10. Outcome:
    Inductive: Produces probabilistic models.
    Analytical: Yields deterministic solutions.