Module 1: Overview of Machine Learning

---

1. Introduction to Machine Learning

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that focuses on building systems that learn from data and improve over time without being explicitly programmed. It enables computers to recognize patterns, make decisions, and adapt to new data.

---

2. Machine Learning Paradigms

- **Supervised Learning:** Learns from labeled data (input-output pairs). Example: classification, regression.
- **Unsupervised Learning:** Finds patterns in unlabeled data. Example: clustering, dimensionality reduction.
- **Semi-Supervised Learning:** Combines a small amount of labeled data with a large amount of unlabeled data.
- **Reinforcement Learning:** An agent learns to make decisions by interacting with an environment to maximize a reward signal.

---

3. Supervised Learning Concepts

- **Input Representation:** The way data is presented to the model. It includes features (independent variables) used to make predictions.
- **Hypothesis Class:** A set of possible functions/models the algorithm considers for mapping inputs to outputs.
- **Version Space:** The subset of the hypothesis class that is consistent with all observed training examples.
- **VC (Vapnik-Chervonenkis) Dimension:** A measure of the capacity (complexity) of a hypothesis class — higher VC dimension means a more complex model that can fit more patterns.
- **PAC (Probably Approximately Correct) Learning:** A theoretical framework that defines when a learning algorithm can learn a function with high probability and low error given sufficient data.
- **Noise:** Refers to errors or randomness in data that can make learning more difficult (e.g., mislabeled examples).

- **Learning Multiple Classes:** Extending binary classification algorithms to handle multi-class problems using strategies like one-vs-all or one-vs-one.
- **Model Selection and Generalization:**
    - **Model Selection** involves choosing the best model or parameters based on performance.
    - **Generalization** is the model's ability to perform well on unseen data, not just the training set.

---

1. Introduction to Machine Learning

1. **What is the primary goal of machine learning?**
   a) Store data
   b) Manually code logic
   c) Learn from data to make predictions ✅
   d) Encrypt data
2. **Which of the following is *not* a machine learning paradigm?**
   a) Supervised learning
   b) Reinforcement learning
   c) Hard-coded programming ✅
   d) Unsupervised learning
3. **In ML, what is 'training data'?**
   a) Randomly generated data
   b) Data used to test a model
   c) Data used to learn model parameters ✅
   d) Data with no labels

2. Machine Learning Paradigms

4. **Which learning paradigm uses labeled data?**
   a) Supervised learning ✅
   b) Unsupervised learning
   c) Reinforcement learning
   d) Clustering
5. **Which learning paradigm learns through interaction with an environment?**
   a) Supervised
   b) Reinforcement ✅

c) Unsupervised
d) Semi-supervised

6. **Which of the following best describes semi-supervised learning?**
   a) All data is labeled
   b) No data is labeled
   c) Some data is labeled, some is not ✅
   d) Data is labeled with noise

7. **Clustering is an example of which learning type?**
   a) Supervised
   b) Reinforcement
   c) Semi-supervised
   d) Unsupervised ✅

8. **Which of the following is a typical application of reinforcement learning?**
   a) Spam detection
   b) Image classification
   c) Game playing ✅
   d) Clustering

3. Supervised Learning - Input Representation & Hypothesis Class

9. **In supervised learning, the input is usually represented as:**
   a) A list of rules
   b) A labeled dataset of feature vectors ✅
   c) Only a target variable
   d) A random noise signal

10. **A hypothesis class is defined as:**
    a) A set of all training examples
    b) A set of possible learning models ✅
    c) A set of labeled data
    d) A set of predictions

11. **Which of the following is *not* a part of a supervised learning setup?**
    a) Training data
    b) Hypothesis
    c) Clustering algorithm ✅
    d) Target variable

12. **What is the purpose of the hypothesis space?**
    a) To store the input data

b) To hold all possible models ✅
c) To map data to random noise
d) To eliminate target variables

4. Version Space

13. **What does a version space contain?**
    a) A single hypothesis
    b) A set of inconsistent hypotheses
    c) Hypotheses consistent with training data ✅
    d) All incorrect predictions

14. **In the Candidate Elimination algorithm, the version space is updated by:**
    a) Removing all hypotheses
    b) Adding noisy data
    c) Narrowing consistent hypotheses ✅
    d) Selecting random hypotheses

15. **What happens to the version space as more training data is added?**
    a) It expands
    b) It becomes more uncertain
    c) It gets refined ✅
    d) It disappears

5. VC Dimension

16. **The VC dimension measures:**
    a) Size of training data
    b) Complexity of a hypothesis class ✅
    c) Number of input features
    d) Training time

17. **Which of the following leads to a higher VC dimension?**
    a) Simpler models
    b) More complex models ✅
    c) Lower training error
    d) Fewer classes

18. **If a classifier can shatter 3 points, its VC dimension is at least:**
    a) 1
    b) 2

c) 3 ✅
d) 4

19. **VC dimension is defined as the largest number of points that can be:**
    a) Merged
    b) Labeled with no noise
    c) Correctly classified in all possible ways ✅
    d) Clustered using K-means

6. PAC Learning

20. **PAC stands for:**
    a) Probably Absolutely Certain
    b) Probably Approximately Correct ✅
    c) Partial Approximate Classification
    d) Predictive Accuracy Condition

21. **In PAC learning, we expect the learner to:**
    a) Always predict perfectly
    b) Sometimes guess
    c) Perform well with high probability ✅
    d) Work without any data

22. **In PAC learning, the concept of ε (epsilon) refers to:**
    a) Size of dataset
    b) Allowed error ✅
    c) Confidence level
    d) Training time

23. **In PAC learning, δ (delta) represents:**
    a) Margin of confidence ✅
    b) Accuracy of the classifier
    c) Learning rate
    d) Noise tolerance

7. Noise in Learning

24. **Noisy data can affect:**
    a) Model accuracy ✅
    b) Data storage

c) Data entry speed
d) Output size

25. **Which is a common method to handle noise?**
    a) Ignore all data
    b) Use data augmentation
    c) Regularization ✅
    d) Increase model size

26. **Label noise refers to:**
    a) Unlabeled data
    b) Incorrect output labels ✅
    c) Irrelevant features
    d) Clustering errors

8. Learning Multiple Classes

27. **Which technique converts a multi-class problem into multiple binary problems?**
    a) Softmax regression
    b) One-vs-All (OvA) ✅
    c) PCA
    d) Bagging

28. **What does One-vs-One classification do?**
    a) Trains a single classifier
    b) Trains one classifier per class
    c) Trains classifiers for every pair of classes ✅
    d) Ignores class labels

29. **Which model is naturally suited for multi-class classification?**
    a) Naive Bayes ✅
    b) Linear Regression
    c) SVM (without modification)
    d) Perceptron

9. Model Selection and Generalization

30. **Model generalization refers to performance on:**
    a) Training data
    b) Noise

c) Unseen data ✅
d) Overfit data

31. **Which is a common method to select the best model?**
    a) Use training accuracy
    b) Use test data only
    c) Cross-validation ✅
    d) Use no labels

32. **Overfitting occurs when the model:**
    a) Performs well on training and test data
    b) Learns general patterns
    c) Fits noise in training data ✅
    d) Has high bias

33. **Underfitting occurs when:**
    a) Model is too complex
    b) Model generalizes well
    c) Model cannot capture patterns in training data ✅
    d) Training accuracy is 100%

34. **Which metric is helpful to detect overfitting?**
    a) Training error only
    b) Test error only
    c) Difference between training and test errors ✅
    d) Model complexity alone

35. **Which approach helps improve generalization?**
    a) Memorizing training data
    b) Increasing model parameters
    c) Regularization ✅
    d) Using fewer features always

Module 2: Supervised Learning

---

1. Dimensionality Reduction

- **Subset Selection:** The process of selecting a subset of relevant features from the original set to reduce the feature space, improving model performance and reducing overfitting.
- **Principal Component Analysis (PCA):** A mathematical technique that transforms the original features into a smaller set of uncorrelated variables (principal components) that capture most of the variance in the data.

---

2. Regression

- **Linear Regression with One Variable:** A method to model the relationship between a single independent variable and a continuous dependent variable using a straight line.
- **Linear Regression with Multiple Variables:** Extension of simple linear regression where multiple input features are used to predict the output.
- **Solution using Gradient Descent Algorithm:** An iterative optimization technique to minimize the cost function by updating model parameters step-by-step.
- **Solution using Matrix Method:** A direct method using linear algebra (Normal Equation) to compute the best fit parameters without iteration.
- **Basic Idea of Overfitting in Regression:** When a model learns noise in the training data and performs poorly on new data due to excessive complexity.

---

3. Linear Methods for Classification

- **Logistic Regression:** A classification algorithm that models the probability of a binary outcome using the logistic function.
- **Naive Bayes:** A probabilistic classifier based on Bayes' theorem assuming independence between features.
- **Decision Tree Algorithm (ID3):** A tree-structured classifier that splits data based on attribute values to classify examples; ID3 uses information gain to decide splits.

1. Dimensionality Reduction – Subset Selection, PCA

1. **Dimensionality reduction aims to:**
   a) Increase the number of features
   b) Reduce computational time by removing irrelevant features ✅
   c) Increase model complexity
   d) Add noise to data

2. **Subset selection selects:**
   a) New feature combinations
   b) A subset of original features ✅
   c) Features based on their names
   d) All features

3. **Principal Component Analysis (PCA) works by:**
   a) Selecting features with lowest variance
   b) Projecting data to a lower-dimensional space ✅
   c) Clustering data
   d) Labeling unlabeled data

4. **The first principal component captures:**
   a) Maximum variance in data ✅
   b) Least variance
   c) Random noise
   d) Irrelevant features

5. **PCA assumes that features are:**
   a) Independent
   b) Linearly correlated ✅
   c) Non-numeric
   d) Discrete

6. **What is the result of PCA?**
   a) Decision tree
   b) Reduced dataset with fewer dimensions ✅
   c) Classifier
   d) Confusion matrix

7. **Which of the following is true for PCA?**
   a) It works only for classification
   b) It generates new transformed features ✅
   c) It keeps all original features
   d) It adds features to the dataset

8. **Which mathematical concept does PCA use?**
   a) Probability
   b) Eigenvalues and Eigenvectors ✅
   c) Gradient descent
   d) Entropy

2. Regression – Linear Regression

9. **Linear regression models the relationship between:**
   a) Two categorical variables
   b) One dependent and one/multiple independent variables ✅
   c) Labels only
   d) Features only

10. **In simple linear regression, the model is of the form:**
    a) y=ax2+by = ax^2 + b
    b) y=mx+cy = mx + c ✅
    c) y=exy = e^x
    d) y=log(x)y = \log(x)

11. **Multivariate linear regression includes:**
    a) One independent variable
    b) Multiple dependent variables
    c) Multiple independent variables ✅
    d) Only binary outputs

12. **The goal of linear regression is to:**
    a) Minimize entropy
    b) Maximize accuracy
    c) Minimize Mean Squared Error (MSE) ✅
    d) Cluster data

13. **Which of the following is used to evaluate linear regression performance?**
    a) Accuracy
    b) Entropy
    c) Mean Squared Error ✅
    d) ROC curve

14. **Which method is *not* used to solve linear regression?**
    a) Matrix inversion
    b) Gradient descent

c) Decision tree ✅
d) Normal equation

15. **Gradient descent is used to:**
    a) Build decision trees
    b) Find local minima of a loss function ✅
    c) Normalize data
    d) Increase dimensionality

16. **Which of the following is a limitation of gradient descent?**
    a) It always converges
    b) It is non-iterative
    c) It can converge to a local minimum ✅
    d) It requires no hyperparameters

17. **The learning rate in gradient descent controls:**
    a) Model complexity
    b) Output labels
    c) Step size during optimization ✅
    d) Number of iterations

3. Overfitting in Regression

18. **Overfitting occurs when the model:**
    a) Generalizes well
    b) Has low variance
    c) Fits training data too closely ✅
    d) Has high bias

19. **Which method helps avoid overfitting?**
    a) Using more complex models
    b) Increasing number of features
    c) Regularization ✅
    d) Reducing training data

20. **Which of the following is a symptom of overfitting?**
    a) Low training error, high test error ✅
    b) High training and test error
    c) Low variance
    d) Good generalization

4. Logistic Regression

21. **Logistic regression is used for:**
    a) Regression tasks
    b) Binary classification ✅
    c) Clustering
    d) Dimensionality reduction

22. **The output of logistic regression is:**
    a) A continuous value
    b) A probability score ✅
    c) A label only
    d) A clustering group

23. **Which function is used in logistic regression?**
    a) Linear
    b) Sigmoid ✅
    c) ReLU
    d) Tanh

24. **Sigmoid function squashes input to the range:**
    a) (-1, 1)
    b) (0, 1) ✅
    c) (-∞, ∞)
    d) (0, ∞)

25. **Which loss function is used in logistic regression?**
    a) Mean squared error
    b) Cross-entropy loss ✅
    c) Hinge loss
    d) Kullback-Leibler divergence

5. Naive Bayes

26. **Naive Bayes is based on:**
    a) Gradient descent
    b) Bayes' theorem ✅
    c) Entropy maximization
    d) Distance metrics

27. **Naive Bayes assumes:**
    a) Features are dependent
    b) Features are independent ✅

c) Non-linear decision boundaries
d) Logistic probabilities

28. **Which type of data is suitable for Naive Bayes?**
    a) Categorical ✅
    b) Continuous only
    c) Time-series only
    d) Graph-based

29. **In Naive Bayes, the prior is:**
    a) Conditional probability
    b) Initial class probability ✅
    c) Model complexity
    d) Feature vector

30. **Which of the following is an advantage of Naive Bayes?**
    a) High computational cost
    b) Robust to irrelevant features ✅
    c) No need for probability
    d) Requires large datasets

6. Decision Tree Algorithm – ID3

31. **ID3 algorithm uses which measure for splitting?**
    a) Gini index
    b) Entropy ✅
    c) Distance
    d) Mean

32. **Which criterion does ID3 use to choose the best attribute?**
    a) Minimum accuracy
    b) Maximum depth
    c) Information gain ✅
    d) Least variance

33. **What is entropy in decision trees?**
    a) Measure of depth
    b) Measure of uncertainty ✅
    c) Measure of purity
    d) Measure of distance

34. **Which is *not* a feature of decision trees?**
    a) Easy to interpret

b) Handles categorical data
c) Sensitive to small data changes ✅
d) Needs normalization

35. **A leaf node in a decision tree represents:**
    a) A new feature
    b) A final decision/class ✅
    c) A data point
    d) A missing value

36. **Which of the following is a disadvantage of ID3?**
    a) Works only on small datasets
    b) Overfitting on training data ✅
    c) Cannot handle categorical variables
    d) Cannot split data

7. Case Study – Face Detection Classifier

37. **Face detection is generally a:**
    a) Regression problem
    b) Classification problem ✅
    c) Clustering problem
    d) Reinforcement task

38. **Which features are commonly used in face detection?**
    a) Random noise
    b) Pixel intensities or edge features ✅
    c) Labels only
    d) Sorted values

39. **Haar-like features are used in:**
    a) Naive Bayes
    b) PCA
    c) Viola-Jones face detection ✅
    d) Logistic regression

40. **Which ML method is often used in real-time face detection?**
    a) SVM
    b) Viola-Jones with AdaBoost ✅
    c) K-means
    d) Naive Bayes

41. **In face detection, false positives occur when:**
    a) A face is missed
    b) A non-face is detected as a face ✅
    c) A face is correctly detected
    d) The model crashes

42. **Which of the following can help improve face detection accuracy?**
    a) Ignoring lighting conditions
    b) Data augmentation ✅
    c) Reducing training data
    d) Random guessing

8. General MCQs on Supervised Learning

43. **Which of the following is *not* supervised learning?**
    a) Linear regression
    b) Naive Bayes
    c) K-means clustering ✅
    d) Decision tree

44. **Which step comes first in supervised learning?**
    a) Model testing
    b) Model tuning
    c) Data labeling ✅
    d) Cross-validation

45. **Which is *not* true about supervised learning models?**
    a) They require labeled data
    b) They predict outputs
    c) They discover data structure ✅
    d) They use input-output pairs

46. **Supervised learning works best when:**
    a) Data is unlabeled
    b) Data is labeled ✅
    c) Data is clustered
    d) Features are missing

47. **Which algorithm is *not* used in regression?**
    a) Linear regression
    b) Ridge regression

c) Logistic regression ✅
d) Lasso

48. **Which classification method uses probability theory?**
    a) SVM
    b) Naive Bayes ✅
    c) ID3
    d) PCA

49. **Which of the following is *not* a linear classifier?**
    a) Logistic regression
    b) Perceptron
    c) SVM (linear kernel)
    d) K-nearest neighbors ✅

50. **Which method can be used for both regression and classification?**
    a) PCA
    b) Decision trees ✅
    c) Naive Bayes
    d) K-means

Module 3: Classification Assessment and Neural Networks

---

1. Classification Performance Measures

- **Precision:** The ratio of correctly predicted positive observations to the total predicted positives. It indicates the accuracy of positive predictions.
- **Recall (Sensitivity):** The ratio of correctly predicted positive observations to all actual positives. It measures the model's ability to find all positive cases.
- **Accuracy:** The ratio of correctly predicted observations to the total observations. It gives an overall performance metric.
- **F-Measure (F1 Score):** The harmonic mean of precision and recall, providing a balance between the two.
- **Receiver Operating Characteristic (ROC) Curve:** A plot of true positive rate (recall) against false positive rate for different threshold settings.
- **Area Under Curve (AUC):** The area under the ROC curve, representing the overall ability of the model to discriminate between classes.

---

2. Bootstrapping and Cross Validation

- **Bootstrapping:** A resampling technique to estimate statistics on a dataset by sampling with replacement, often used to assess model stability.
- **Cross Validation:** A technique for evaluating the generalization ability of a model by partitioning data into training and testing sets multiple times (e.g., k-fold cross-validation).

---

3. Perceptron

- A simple binary linear classifier that updates weights based on misclassification errors; foundational model for neural networks.

---

4. Neural Networks

- **Multilayer Feed Forward Network:** A network consisting of multiple layers of neurons where information flows forward from input to output layers.
- **Activation Functions:** Functions applied to neuron outputs to introduce non-linearity.

    - **Sigmoid:** Outputs values between 0 and 1, useful for binary classification.
    - **ReLU (Rectified Linear Unit):** Outputs zero if input is negative, else outputs input directly; popular due to efficiency.
    - **Tanh:** Outputs values between -1 and 1, centered at zero.
- **Backpropagation Algorithm:** A method for training neural networks by propagating errors backward through the network and updating weights using gradient descent.

---

1. Classification Performance Measures

1. **Accuracy is defined as:**
   a) TP / (TP + FP)
   b) (TP + TN) / (TP + TN + FP + FN) ✅
   c) FN / (TP + FP)
   d) TN / (TP + FP + TN + FN)

2. **Precision is:**
   a) TP / (TP + FP) ✅
   b) TP / (TP + FN)
   c) TN / (TN + FP)
   d) (TP + TN) / Total

3. **Recall is also known as:**
   a) Specificity
   b) Sensitivity ✅
   c) Accuracy
   d) Precision

4. **F1 Score is the:**
   a) Harmonic mean of accuracy and recall
   b) Harmonic mean of precision and recall ✅
   c) Arithmetic mean of precision and recall
   d) Geometric mean of accuracy and recall

5. **Which metric is most useful when false negatives are costly?**
   a) Accuracy
   b) Recall ✅
   c) Precision
   d) F1-score

6. **Which metric is best when false positives are costly?**
   a) Accuracy

b) Recall
c) Precision ✅
d) AUC

7. **True Negative (TN) means:**
   a) Incorrectly predicted positive
   b) Correctly predicted negative ✅
   c) Incorrectly predicted negative
   d) True value is positive

8. **AUC stands for:**
   a) Average Under Curve
   b) Area Under Curve ✅
   c) All Unique Classes
   d) Accuracy Using Classifiers

9. **ROC Curve plots:**
   a) Precision vs Recall
   b) TP vs TN
   c) TPR vs FPR ✅
   d) Accuracy vs Precision

10. **The best ROC curve will be closest to:**
    a) Diagonal
    b) Bottom right
    c) Top left ✅
    d) Bottom left

2. Bootstrapping and Cross-validation

11. **Bootstrapping involves:**
    a) Splitting data without replacement
    b) Drawing samples with replacement ✅
    c) Using test data for training
    d) Applying PCA

12. **Purpose of bootstrapping is:**
    a) Feature selection
    b) Estimate model performance ✅
    c) Dimensionality reduction
    d) Hyperparameter tuning

13. **Cross-validation is used to:**
    a) Overfit a model
    b) Reduce dataset size
    c) Assess generalization performance ✅
    d) Increase bias

14. **In k-fold cross-validation, the data is:**
    a) Used all for testing
    b) Split into k test sets
    c) Split into k parts and each used once for testing ✅
    d) Trained on k parts and never tested

15. **In 10-fold cross-validation, how many times does training occur?**
    a) 1
    b) 5
    c) 10 ✅
    d) 20

16. **Leave-One-Out Cross-Validation (LOOCV) is a special case of:**
    a) Random search
    b) Bootstrapping
    c) k-fold where k = number of samples ✅
    d) Bagging

17. **Cross-validation helps in:**
    a) Creating new features
    b) Overfitting
    c) Selecting models ✅
    d) Speeding up training

3. Perceptron

18. **Perceptron is a type of:**
    a) Decision tree
    b) Neural network ✅
    c) Clustering algorithm
    d) Statistical test

19. **Perceptron is used for:**
    a) Linear classification ✅
    b) Regression

c) Clustering
d) Dimensionality reduction

20. **Perceptron learning rule updates weights based on:**
    a) Mean squared error
    b) Entropy
    c) Error between predicted and actual label ✅
    d) Gradient of accuracy

21. **Perceptron converges only if the data is:**
    a) Noisy
    b) Non-linear
    c) Linearly separable ✅
    d) Unlabeled

22. **The output of a perceptron is:**
    a) Continuous
    b) Sigmoid
    c) Binary (0 or 1) ✅
    d) Multinomial

23. **Which of the following is not part of a perceptron?**
    a) Input vector
    b) Bias
    c) Activation function
    d) Hidden layers ✅

24. **What does the bias in a perceptron do?**
    a) Scales inputs
    b) Reduces overfitting
    c) Shifts activation function ✅
    d) Increases training data

4. Neural Networks – Multilayer Feedforward Networks

25. **A feedforward neural network has:**
    a) Cycles
    b) Only one layer
    c) Information flow from input to output without loops ✅
    d) Feedback connections

26. **In a neural network, a hidden layer is:**
    a) Between input and output ✅

b) A visible output
c) Final classification layer
d) Not used during training

27. **Which activation function is not commonly used in modern neural networks?**
    a) ReLU
    b) Sigmoid
    c) Tanh
    d) Step function ✅

28. **ReLU activation function outputs:**
    a) Values between 0 and 1
    b) 0 or 1
    c) 0 if input < 0, else input ✅
    d) Negative values only

29. **Tanh activation function outputs values between:**
    a) 0 and 1
    b) -1 and 1 ✅
    c) 0 and ∞
    d) -∞ and 0

30. **Which of the following is *not* an advantage of ReLU?**
    a) Reduces vanishing gradient problem
    b) Simpler to compute
    c) Outputs negative values ✅
    d) Works well in deep networks

31. **Sigmoid activation function has a problem of:**
    a) Exploding gradients
    b) Vanishing gradients ✅
    c) Output exceeding bounds
    d) No differentiability

32. **Which layer performs non-linear transformation in neural networks?**
    a) Output layer
    b) Input layer
    c) Hidden layer ✅
    d) Data layer

5. Backpropagation Algorithm

33. **Backpropagation is used to:**
    a) Increase bias
    b) Normalize data
    c) Adjust weights in neural networks ✅
    d) Select features

34. **Backpropagation uses which method for optimization?**
    a) Gradient descent ✅
    b) Entropy maximization
    c) Random selection
    d) Nearest neighbor

35. **Which values are propagated backward in backpropagation?**
    a) Inputs
    b) Outputs
    c) Errors ✅
    d) Labels

36. **Backpropagation calculates gradients using:**
    a) Chain rule ✅
    b) Bayes' theorem
    c) Matrix inversion
    d) Statistical averaging

37. **The goal of backpropagation is to:**
    a) Maximize error
    b) Increase weights
    c) Minimize loss function ✅
    d) Increase classification time

38. **In backpropagation, weight update is influenced by:**
    a) Learning rate ✅
    b) Accuracy
    c) ROC
    d) Confusion matrix

39. **Which function is commonly used as a loss function in classification?**
    a) Mean Absolute Error
    b) Cross-entropy ✅
    c) Sum of squares
    d) Euclidean distance

40. **The first step in backpropagation is:**
    a) Weight update

b) Forward pass ✅
c) Gradient update
d) Activation calculation

6. General MCQs on Module Concepts

41. **Which model can be trained using backpropagation?**
    a) Decision Tree
    b) Perceptron
    c) Multilayer Neural Network ✅
    d) K-means

42. **Which component is not necessary in a single-layer perceptron?**
    a) Input
    b) Output
    c) Hidden layer ✅
    d) Weights

43. **A confusion matrix does not show:**
    a) TP
    b) FP
    c) ROC ✅
    d) FN

44. **Which function is typically used in output layer of binary classifier?**
    a) Tanh
    b) ReLU
    c) Sigmoid ✅
    d) Linear

45. **Classification performance measure not affected by class imbalance is:**
    a) Accuracy
    b) Precision
    c) Recall
    d) ROC-AUC ✅

46. **High variance in a model indicates:**
    a) Underfitting
    b) Overfitting ✅
    c) Good generalization
    d) Poor training

47. **Which is a linear classifier?**
    a) Logistic regression ✅
    b) Decision tree
    c) k-NN
    d) RBF network

48. **A good classifier will have:**
    a) High FP rate
    b) Low accuracy
    c) High AUC ✅
    d) High entropy

49. **Which function is differentiable and used in neural networks?**
    a) Step
    b) Sigmoid ✅
    c) Random
    d) Max

50. **The vanishing gradient problem is reduced by:**
    a) Using sigmoid activation
    b) Increasing layers
    c) Using ReLU ✅
    d) Adding dropout

Module 4: Parameter Estimation & SVM Classifier

---

1. Basics of Parameter Estimation

- **Maximum Likelihood Estimation (MLE):** A statistical method to estimate model parameters by maximizing the likelihood that the observed data was generated by the model.
- **Maximum a Posteriori Estimation (MAP):** Similar to MLE but incorporates prior knowledge (prior probability) about parameters using Bayes' theorem, resulting in a regularized estimate.
- **Bias-Variance Decomposition:** A framework to analyze the error of a model by splitting it into bias (error due to incorrect assumptions) and variance (error due to sensitivity to training data fluctuations).

---

2. Support Vector Machines (SVM)

- **Introduction:** SVM is a powerful supervised learning algorithm for classification and regression that finds the best separating hyperplane.
- **Maximum Margin Hyperplanes:** SVM finds the hyperplane that maximizes the margin (distance) between different classes, improving generalization.
- **Mathematics Behind Maximum Margin Classification:** Formulation as a convex optimization problem ensuring the widest possible margin while correctly classifying training points.
- **Soft Margin SVM Classifier:** Allows some misclassifications (slack variables) to handle non-linearly separable data while still maximizing the margin.
- **Non-linear SVM:** Uses kernel functions to transform data into higher-dimensional space to make it linearly separable.
- **Kernels for Learning Non-linear Functions:** Functions that implicitly map data into higher dimensions without explicit transformation.
    - **Polynomial Kernel:** Maps data into polynomial feature space.
    - **Radial Basis Function (RBF):** Measures similarity based on distance, widely used for non-linear classification.
- **Kernel Trick:** A technique to apply kernels so that the SVM optimization problem can be solved efficiently without computing high-dimensional mappings explicitly.

1. Parameter Estimation – MLE & MAP

1. **Maximum Likelihood Estimation (MLE) estimates parameters by:**
   a) Minimizing error
   b) Maximizing posterior probability
   c) Maximizing likelihood of observed data ✅
   d) Minimizing variance

2. **MAP stands for:**
   a) Maximum Average Prediction
   b) Maximum A Posteriori ✅
   c) Mean Absolute Prediction
   d) Marginal Average Posterior

3. **MLE does not consider:**
   a) Data likelihood
   b) Prior distribution ✅
   c) Observed data
   d) Probability density

4. **MAP estimation differs from MLE by incorporating:**
   a) Loss function
   b) Bias
   c) Prior probability ✅
   d) Training set size

5. **When the prior is uniform, MAP becomes equivalent to:**
   a) Entropy
   b) MLE ✅
   c) Bayes Theorem
   d) Bias estimation

6. **MAP estimation maximizes:**
   a) P(Data | Parameters)
   b) P(Parameters | Data) ✅
   c) P(Data)
   d) Posterior variance

7. **Which of the following is a disadvantage of MLE?**
   a) High bias
   b) Overfitting for small data ✅
   c) Depends on prior
   d) Requires Bayesian methods

8. **MLE finds parameters θ that:**
   a) Minimize P(D|θ)
   b) Maximize P(D|θ) ✅
   c) Maximize P(θ|D)
   d) Minimize variance

9. **MAP estimation is commonly used in:**
   a) Frequentist methods
   b) Bayesian methods ✅
   c) Unsupervised learning
   d) Clustering algorithms

10. **Likelihood in MLE refers to:**
    a) P(θ)
    b) P(D|θ) ✅
    c) P(θ|D)
    d) P(D)

2. Bias-Variance Decomposition

11. **Bias measures the error due to:**
    a) Variability in training data
    b) Incorrect assumptions in the learning algorithm ✅
    c) Random noise
    d) Overfitting

12. **Variance measures the error due to:**
    a) Model complexity
    b) Changes in training data ✅
    c) Underfitting
    d) Bias

13. **High bias leads to:**
    a) Overfitting
    b) Underfitting ✅
    c) Good generalization
    d) Increased variance

14. **High variance leads to:**
    a) Overfitting ✅
    b) Underfitting

c) Good training performance
d) Robust prediction

15. **Bias-variance trade-off is important for:**
    a) Model testing
    b) Optimization
    c) Model generalization ✅
    d) Feature selection

16. **Which of the following models is more likely to have high bias?**
    a) Linear regression ✅
    b) Random forest
    c) Neural network
    d) SVM with RBF

17. **Variance can be reduced by:**
    a) Using more complex models
    b) Increasing training data ✅
    c) Increasing bias
    d) Decreasing model size

18. **In general, increasing model complexity:**
    a) Increases bias
    b) Increases variance ✅
    c) Decreases variance
    d) Increases error

19. **Which model is prone to low bias and high variance?**
    a) Simple linear regression
    b) Decision tree ✅
    c) Logistic regression
    d) Naive Bayes

20. **In bias-variance tradeoff, the total error is the sum of:**
    a) Accuracy and precision
    b) Bias, variance, and noise ✅
    c) Bias and accuracy
    d) Precision and recall

3. Support Vector Machines (SVM)

21. **SVM stands for:**
    a) Support Validation Method

b) Supervised Vector Machine
c) Support Vector Machine ✅
d) Sample Vector Model

22. **SVMs are primarily used for:**
    a) Clustering
    b) Regression
    c) Classification ✅
    d) Dimensionality reduction

23. **SVM finds the hyperplane that:**
    a) Minimizes classification error
    b) Maximizes margin ✅
    c) Minimizes variance
    d) Increases entropy

24. **The support vectors are:**
    a) Points farthest from the decision boundary
    b) Data points not used in training
    c) Points closest to the decision boundary ✅
    d) Randomly selected points

25. **In linearly separable data, SVM aims to:**
    a) Reduce accuracy
    b) Maximize distance between classes ✅
    c) Minimize margin
    d) Use all data points equally

26. **SVM is a type of:**
    a) Probabilistic model
    b) Non-parametric model
    c) Margin-based classifier ✅
    d) Generative model

27. **The optimal hyperplane in SVM is:**
    a) Arbitrary
    b) Perpendicular to data
    c) Equidistant from support vectors ✅
    d) Closest to the majority class

28. **Which function is used to define the margin in SVM?**
    a) Loss function
    b) Hinge loss ✅

c) Logistic loss
d) Entropy loss

4. Soft Margin SVM

29. **Soft margin SVM allows:**
    a) Linear separation
    b) No misclassification
    c) Some misclassification ✅
    d) Higher bias

30. **Soft margin is controlled by:**
    a) Kernel function
    b) Slack variable ✅
    c) Prior
    d) Number of classes

31. **Slack variable allows SVM to:**
    a) Increase margin
    b) Handle non-linear data
    c) Handle outliers ✅
    d) Remove noise

32. **Trade-off between margin and misclassification is managed by:**
    a) γ parameter
    b) C parameter ✅
    c) α value
    d) Threshold

33. **High value of C in SVM leads to:**
    a) Wider margin
    b) More misclassification
    c) Less misclassification ✅
    d) Lower variance

5. Non-linear SVM and Kernel Trick

34. **Kernel trick allows SVM to:**
    a) Work with labeled data
    b) Reduce dimensionality

c) Learn non-linear boundaries ✅
d) Avoid bias

35. **Kernel function helps in:**
    a) Visualizing data
    b) Projecting data into higher dimension ✅
    c) Reducing training time
    d) Enhancing loss function

36. **A common kernel used in non-linear SVM is:**
    a) Euclidean
    b) Tanh
    c) Radial Basis Function (RBF) ✅
    d) Step function

37. **Kernel trick avoids:**
    a) Overfitting
    b) Explicit computation of high-dimensional mappings ✅
    c) Data transformation
    d) Linear transformation

38. **Polynomial kernel is used for:**
    a) Handling binary classification
    b) Modeling polynomial decision boundaries ✅
    c) Learning linear functions
    d) Normalizing features

39. **The degree of the polynomial kernel determines:**
    a) Output class
    b) Margin size
    c) Model complexity ✅
    d) Distance metric

40. **Which is not a valid kernel?**
    a) Linear
    b) RBF
    c) Cosine
    d) Manhattan ✅

6. Advanced & General Questions

41. **Which of the following is not an advantage of kernel SVMs?**
    a) Can handle non-linear data

b) Avoids explicit transformation
c) High interpretability ✅
d) Flexibility in decision boundary

42. **RBF kernel is a function of:**
a) Distance between data points ✅
b) Sum of features
c) Number of samples
d) Data dimensionality

43. **Which parameter in RBF controls the spread of the kernel?**
a) α
b) γ (gamma) ✅
c) θ
d) δ

44. **SVM can be used for regression via:**
a) SVR ✅
b) RBF
c) MLP
d) PCA

45. **SVM is considered a:**
a) Generative model
b) Discriminative model ✅
c) Bayesian model
d) Clustering model

46. **Support vectors are those that:**
a) Lie outside the margin
b) Maximize error
c) Influence the decision boundary ✅
d) Minimize bias

47. **Increasing the number of support vectors generally:**
a) Improves model speed
b) Decreases generalization ✅
c) Increases interpretability
d) Removes outliers

48. **Which kernel maps data into infinite-dimensional space?**
a) Linear
b) Polynomial

c) RBF ✅
d) Sigmoid

49. **The dual formulation of SVM optimization problem helps in:**
    a) Faster matrix computation
    b) Avoiding local minima
    c) Using kernel trick ✅
    d) Increasing C

50. **What is the main goal of SVM optimization?**
    a) Maximize classification error
    b) Maximize margin ✅
    c) Minimize number of features
    d) Minimize AUC

Module 5: Unsupervised Learning

---

1. Ensemble Methods

- **Voting:** Combines predictions from multiple models by majority vote (for classification) or averaging (for regression) to improve accuracy.
- **Bagging (Bootstrap Aggregating):** Builds multiple models by training on different bootstrap samples of the data and aggregates their outputs to reduce variance and avoid overfitting.
- **Boosting:** Sequentially builds models where each new model focuses on correcting errors made by previous models, improving overall prediction accuracy.

---

2. Unsupervised Learning Clustering Methods

- **Similarity Measures:** Metrics used to quantify how alike two data points are, such as Euclidean distance, cosine similarity, etc., which guide clustering.
- **K-means Clustering:** A popular clustering algorithm that partitions data into k clusters by minimizing the variance within each cluster.
- **Expectation-Maximization (EM) for Soft Clustering:** An iterative algorithm that assigns probabilities to data points belonging to clusters, allowing overlapping clusters.
- **Hierarchical Clustering Methods:** Build a tree-like structure (dendrogram) representing nested clusters, either by merging (agglomerative) or splitting (divisive) clusters.
- **Density-Based Clustering:** Groups data points based on regions of high density separated by regions of low density, effective for arbitrary shaped clusters and noise detection (e.g., DBSCAN).

---

1. Ensemble Methods – Voting, Bagging, Boosting

1. **What is the main purpose of ensemble methods?**
   a) Reduce training time
   b) Improve model accuracy ✅
   c) Increase bias
   d) Use single models effectively

2. **Which of the following is a type of ensemble method?**
   a) Decision Tree
   b) Bagging ✅
   c) K-means
   d) PCA

3. **Voting in ensemble learning combines:**
   a) Feature vectors
   b) Predictions from multiple models ✅
   c) Similarity measures
   d) Distance metrics

4. **In hard voting, the final output is determined by:**
   a) Weighted sum of probabilities
   b) Majority class label ✅
   c) Averaging errors
   d) Ensemble loss

5. **In soft voting, the prediction is based on:**
   a) Random decision
   b) Mode of predictions
   c) Weighted probabilities ✅
   d) Least number of models

6. **Bagging stands for:**
   a) Bias Aggregation
   b) Bootstrap Aggregation ✅
   c) Best Average Group
   d) Binned Algorithm Grouping

7. **Bagging reduces:**
   a) Bias
   b) Variance ✅
   c) Overfitting
   d) Dimensionality

8. **A popular algorithm using bagging is:**
   a) Decision Tree
   b) SVM
   c) Random Forest ✅
   d) Naive Bayes

9. **Bootstrapping involves:**
   a) Data normalization

b) Sampling with replacement ✅
c) Principal component analysis
d) Standardization

10. **Boosting focuses on:**
    a) Averaging predictions
    b) Giving equal weight to all models
    c) Combining weak learners sequentially ✅
    d) Reducing dimensionality

11. **In boosting, each new model tries to:**
    a) Improve variance
    b) Correct previous errors ✅
    c) Reduce clustering
    d) Maximize prior

12. **Which is a boosting algorithm?**
    a) AdaBoost ✅
    b) KNN
    c) PCA
    d) K-means

13. **Boosting reduces:**
    a) Variance
    b) Noise
    c) Bias ✅
    d) Number of features

14. **Which ensemble method works best with high variance models?**
    a) Voting
    b) Bagging ✅
    c) Boosting
    d) Clustering

15. **Which method uses weighted samples in training?**
    a) Bagging
    b) Boosting ✅
    c) Voting
    d) Random sampling

2. Clustering: Similarity Measures

16. **Which of the following is a similarity measure?**
    a) Euclidean distance ✅
    b) Entropy
    c) Cross entropy
    d) ROC

17. **Manhattan distance is also known as:**
    a) L1 norm ✅
    b) L2 norm
    c) Cosine distance
    d) Inner product

18. **Cosine similarity measures the:**
    a) Angle between vectors ✅
    b) Magnitude of vectors
    c) Euclidean distance
    d) Variance

19. **Jaccard similarity is best used for:**
    a) Categorical data ✅
    b) Continuous values
    c) Image data
    d) Noisy signals

20. **Euclidean distance is not suitable for:**
    a) Sparse data ✅
    b) Numerical data
    c) High-dimensional data
    d) Feature selection

3. K-Means Clustering

21. **K-means is a type of:**
    a) Supervised algorithm
    b) Unsupervised algorithm ✅
    c) Reinforcement method
    d) Semi-supervised model

22. **K in K-means refers to:**
    a) Number of dimensions
    b) Number of clusters ✅

c) Distance function
d) Number of iterations

23. **K-means clustering minimizes:**
    a) Between-cluster distance
    b) Within-cluster sum of squares ✅
    c) Euclidean norm
    d) Similarity

24. **The K-means algorithm is sensitive to:**
    a) Feature scaling ✅
    b) Model complexity
    c) Variance
    d) Number of features

25. **The first step in K-means is:**
    a) Update centroids
    b) Assign points to clusters
    c) Randomly initialize centroids ✅
    d) Calculate intra-cluster distance

26. **Which is a drawback of K-means?**
    a) High bias
    b) Requires number of clusters a priori ✅
    c) Works with categorical data
    d) Non-deterministic

27. **K-means works best with:**
    a) Circular clusters ✅
    b) Irregular shapes
    c) Categorical features
    d) High-dimensional sparse data

28. **In K-means, cluster assignment is based on:**
    a) Kernel function
    b) Probability
    c) Distance to nearest centroid ✅
    d) Majority vote

29. **Which method can help in selecting K value?**
    a) Gradient descent
    b) Elbow method ✅
    c) Bagging
    d) ROC curve

4. EM for Soft Clustering

30. **EM stands for:**
    a) Expectation Mapping
    b) Expectation Maximization ✅
    c) Empirical Method
    d) Error Minimization

31. **In soft clustering, each point belongs to:**
    a) One cluster
    b) All clusters equally
    c) Multiple clusters with probabilities ✅
    d) None of the clusters

32. **The E-step in EM algorithm:**
    a) Maximizes parameters
    b) Updates weights
    c) Calculates membership probabilities ✅
    d) Chooses clusters

33. **The M-step in EM algorithm:**
    a) Updates model parameters ✅
    b) Estimates centroids
    c) Computes distances
    d) Calculates mean

34. **EM algorithm is commonly used with:**
    a) K-means
    b) Gaussian Mixture Models ✅
    c) PCA
    d) SVM

35. **EM is more suitable than K-means when:**
    a) Data is categorical
    b) Clusters overlap ✅
    c) Dimensionality is low
    d) Clusters are sparse

36. **Soft clustering differs from hard clustering by:**
    a) Allowing overlapping memberships ✅
    b) Using labels

c) Having one cluster
d) Supervised learning

5. Hierarchical Clustering

37. **Hierarchical clustering builds:**
    a) Disjoint clusters
    b) Flat cluster assignment
    c) Nested clusters ✅
    d) Random partitions

38. **Which type of hierarchical clustering starts from individual points?**
    a) Divisive
    b) Agglomerative ✅
    c) Soft
    d) Non-parametric

39. **Divisive clustering starts with:**
    a) One large cluster ✅
    b) Many small clusters
    c) Distance matrix
    d) Zero clusters

40. **Dendrogram is a tool used in:**
    a) K-means
    b) Boosting
    c) Hierarchical clustering ✅
    d) PCA

41. **Linkage criteria in hierarchical clustering includes:**
    a) Average, Single, Complete ✅
    b) Hard, Soft, Medium
    c) Random, Max, Min
    d) Cosine, Jaccard, Manhattan

42. **Single linkage considers:**
    a) Centroid distance
    b) Maximum distance
    c) Minimum distance ✅
    d) All pairwise distances

43. **Complete linkage uses:**
    a) Min distance

b) Max distance ✅
c) Average
d) Variance

6. Density-Based Clustering

44. **DBSCAN is a:**
    a) Partitioning method
    b) Density-based clustering method ✅
    c) Boosting algorithm
    d) Dimensionality reduction

45. **DBSCAN stands for:**
    a) Distance-Based Clustering
    b) Density-Based Spatial Clustering of Applications with Noise ✅
    c) Decision-Based Spatial Clustering
    d) Dense Binary Scan

46. **DBSCAN is good at detecting:**
    a) Linear boundaries
    b) Uniform distributions
    c) Arbitrary shaped clusters ✅
    d) Labeled data

47. **DBSCAN requires two parameters:**
    a) Learning rate and epochs
    b) K and centroids
    c) Epsilon and MinPts ✅
    d) Distance and similarity

48. **Core points in DBSCAN are:**
    a) Distant from clusters
    b) Inside noise
    c) Surrounded by minimum points ✅
    d) Starting points

49. **Noise points in DBSCAN are:**
    a) Core points
    b) Points in dense regions
    c) Not assigned to any cluster ✅
    d) Always classified

50. **Which method is most robust to outliers?**
    a) K-means
    b) DBSCAN ✅
    c) Hierarchical
    d) Voting

General Machine Learning MCQs

1. **Which of the following is an example of reinforcement learning?**
   a) Linear regression
   b) K-means clustering
   c) Q-learning ✅
   d) PCA

2. **Which machine learning approach is best for anomaly detection?**
   a) Supervised learning
   b) Semi-supervised learning ✅
   c) Reinforcement learning
   d) Classification

3. **Which algorithm is best suited for spam detection in emails?**
   a) K-means
   b) Naive Bayes ✅
   c) PCA
   d) Random Forest

4. **What does the ROC curve plot?**
   a) True Positive Rate vs False Positive Rate ✅
   b) Precision vs Recall
   c) Accuracy vs Error
   d) Specificity vs Sensitivity

5. **Which of the following is a lazy learning algorithm?**
   a) KNN ✅
   b) Decision Tree
   c) SVM
   d) Logistic Regression

6. **Which of the following is used for feature selection?**
   a) PCA
   b) Information Gain ✅
   c) K-means
   d) Gradient descent

7. **Which algorithm works well with non-linearly separable data?**
   a) Linear regression
   b) Logistic regression
   c) SVM with RBF kernel ✅
   d) Perceptron

8. **Which method is used to reduce overfitting?**
   a) Dropout ✅
   b) Gradient boosting
   c) K-means
   d) Sigmoid function

9. **What is the main goal of cross-validation?**
   a) Increase training speed
   b) Improve testing error
   c) Estimate model performance ✅
   d) Normalize data

10. **Which of these is an ensemble method?**
    a) AdaBoost ✅
    b) PCA
    c) KNN
    d) K-means

Model Evaluation, Tuning, and Optimization

11. **Hyperparameter tuning is performed using:**
    a) Gradient descent
    b) Backpropagation
    c) Grid Search ✅
    d) Clustering

12. **Which method helps prevent overfitting?**
    a) Increasing depth of decision trees
    b) Using more training data ✅
    c) Using training error only
    d) Ignoring regularization

13. **The penalty term in L2 regularization is based on:**
    a) Absolute value
    b) Square of coefficients ✅
    c) Number of features
    d) Feature importance

14. **Which of the following is NOT a metric for regression?**
    a) MAE
    b) RMSE

c) Accuracy ✅
d) R^2

15. **Early stopping is a form of:**
    a) Feature selection
    b) Model optimization
    c) Regularization ✅
    d) Activation function

Feature Engineering and Data Preparation

16. **Which of the following is a feature scaling technique?**
    a) Label encoding
    b) One-hot encoding
    c) Min-Max normalization ✅
    d) Cross-validation

17. **Which technique converts categorical variables into binary vectors?**
    a) Label encoding
    b) PCA
    c) One-hot encoding ✅
    d) Z-score normalization

18. **Outliers can be detected using:**
    a) Logistic regression
    b) Boxplot or Z-score ✅
    c) PCA
    d) Bagging

19. **Which technique is useful for reducing multicollinearity?**
    a) PCA ✅
    b) SVM
    c) K-means
    d) Random Forest

20. **The curse of dimensionality affects:**
    a) Classification accuracy
    b) Distance-based algorithms ✅
    c) Regularization
    d) Regression only

Advanced & Practical ML Concepts

21. **Which concept helps in improving fairness in ML models?**
    a) Hyperparameter tuning
    b) Bias mitigation ✅
    c) Ensemble learning
    d) Dropout

22. **Transfer learning is most often used in:**
    a) Text data
    b) Tabular data
    c) Image and NLP tasks ✅
    d) Time series

23. **Which of the following is not a neural network architecture?**
    a) CNN
    b) RNN
    c) SVM ✅
    d) GAN

24. **Which method is used for dimensionality reduction in NLP?**
    a) Word2Vec
    b) PCA ✅
    c) Bag of Words
    d) Label encoding

25. **In time series forecasting, which model is commonly used?**
    a) ARIMA ✅
    b) KNN
    c) Logistic regression
    d) PCA

Application-Oriented Questions

26. **Which ML application is used in recommendation systems?**
    a) Reinforcement learning
    b) Collaborative filtering ✅
    c) Clustering
    d) Classification

27. **Sentiment analysis is a problem of:**
    a) Clustering
    b) Classification ✅

c) Regression
d) Feature extraction

28. **Which model is best for predicting house prices?**
    a) Logistic regression
    b) KNN
    c) Linear regression ✅
    d) DBSCAN

29. **Face recognition is an application of:**
    a) Regression
    b) Classification ✅
    c) Dimensionality reduction
    d) Anomaly detection

30. **Self-driving cars use which type of learning for decision-making?**
    a) Reinforcement learning ✅
    b) Clustering
    c) Regression
    d) Unsupervised learning

Data & Bias

31. **Data leakage refers to:**
    a) Missing data
    b) Test data used in training ✅
    c) Outliers in dataset
    d) Poor model performance

32. **Which type of bias occurs when a model favors one group over another?**
    a) Selection bias
    b) Label bias
    c) Algorithmic bias ✅
    d) Statistical bias

33. **What is data imbalance?**
    a) Missing values
    b) Unequal number of samples per class ✅
    c) Unstructured data
    d) Overlapping features

34. **One way to handle class imbalance is:**
    a) PCA

b) Oversampling ✅
c) Feature scaling
d) Bagging

35. **Which cross-validation is good for imbalanced data?**
    a) Standard CV
    b) Leave-One-Out CV
    c) Stratified CV ✅
    d) Random split

Terminology and Theory

36. **A model's ability to perform well on new data is known as:**
    a) Bias
    b) Training error
    c) Generalization ✅
    d) Overfitting

37. **A convex function has:**
    a) Multiple local minima
    b) A single global minimum ✅
    c) Complex roots
    d) No gradient

38. **Gradient descent optimizes:**
    a) Performance
    b) Loss function ✅
    c) Activation
    d) Features

39. **Which of the following is NOT a supervised learning task?**
    a) Spam classification
    b) Movie recommendation ✅
    c) House price prediction
    d) Sentiment analysis

40. **Which technique is best for reducing bias in models?**
    a) Bagging
    b) Boosting ✅
    c) K-means
    d) PCA

Miscellaneous

41. **What is a hyperparameter in ML?**
    a) Learned from data
    b) Set before training ✅
    c) Hidden layer output
    d) Final model accuracy

42. **Which is true about Naive Bayes?**
    a) Features are dependent
    b) It's slow
    c) Assumes feature independence ✅
    d) Cannot handle text data

43. **Overfitting means:**
    a) Model performs well on new data
    b) Model learns noise ✅
    c) Model underperforms
    d) Model fails to converge

44. **Which is true about underfitting?**
    a) Model is too complex
    b) High training accuracy
    c) Model can't capture data trends ✅
    d) High test accuracy

45. **Which method helps in visualizing high-dimensional data?**
    a) t-SNE ✅
    b) KNN
    c) Gradient descent
    d) Logistic regression

46. **Which model is best for sequential data?**
    a) CNN
    b) RNN ✅
    c) PCA
    d) Naive Bayes

47. **Which algorithm updates weights using errors?**
    a) Naive Bayes
    b) KNN
    c) Backpropagation ✅
    d) PCA

48. **Which of these is a generative model?**
    a) GAN ✅
    b) Logistic Regression
    c) SVM
    d) Random Forest

49. **Which ML model is most interpretable?**
    a) Random Forest
    b) Neural Network
    c) Decision Tree ✅
    d) SVM

50. **Which of the following is not a type of ML?**
    a) Supervised
    b) Unsupervised
    c) Enhanced ✅
    d) Reinforcement