Artificial Intelligence Systems I

Exam Revisions

Lecture 2: Core Concepts of Artificial Intelligence

1. **Which of these best describes Artificial Intelligence (AI)?**
   A. The field that builds systems to simulate or extend human intelligence in real-world tasks
   B. Only computer vision applications
   C. Just another name for programming
   D. Hardware design for smart devices
2. **What is Machine Learning (ML)?**
   A. Algorithms that learn from data to improve performance without explicit programming
   B. Traditional rule-based programming
   C. Only deep neural networks
   D. Manual data analysis techniques
3. **How is Deep Learning (DL) related to Machine Learning and AI?**
   A. Deep Learning ⊂ Machine Learning ⊂ AI
   B. AI ⊂ Machine Learning ⊂ Deep Learning
   C. Machine Learning ⊂ AI ⊂ Deep Learning
   D. They are completely separate fields
4. **Which statement about Deep Learning is correct?**
   A. A subset of ML that uses multi-layer neural networks to learn rich representations
   B. Another name for traditional programming
   C. Only used for computer vision tasks
   D. Doesn't require data to learn
5. **What is the correct hierarchical relationship among these fields?**
   A. AI (broadest) → ML (subset) → DL (subset of ML)
   B. ML (broadest) → DL → AI (most specific)
   C. DL (broadest) → AI → ML (most specific)
   D. All three are independent fields
6. **What characterizes Narrow or Weak AI?**
   A. Specialized in performing a single task
   B. Has human-level intelligence across all domains
   C. Possesses consciousness and self-awareness
   D. Can perform any intellectual task a human can
7. **Which of these is an example of Narrow AI?**
   A. Siri or Alexa voice assistants
   B. A hypothetical system with human-level general intelligence
   C. Superintelligent AI from science fiction
   D. Self-aware robots

8. **What does General or Strong AI refer to?**
   A. Hypothetical systems with human-level intelligence across all intellectual tasks
   B. Current specialized AI systems
   C. Basic rule-based systems
   D. AI with physical bodies
9. **Superintelligent AI is best described as:**
   A. A theoretical AI that surpasses human intelligence across all domains
   B. Current AI systems we use daily
   C. AI specialized in one task
   D. AI that can only react to current inputs
10. **Which capability-based AI type describes most current AI systems?**
    A. Narrow or Weak AI
    B. General or Strong AI
    C. Superintelligent AI
    D. Reactive AI
11. **What characterizes Reactive Machines?**
    A. Basic AI with no memory, only responding to current input
    B. AI that uses past data to make decisions
    C. AI that understands human emotions
    D. AI with self-awareness
12. **Which AI functionality type uses past data to make decisions?**
    A. Limited Memory AI
    B. Reactive Machines
    C. Theory of Mind AI
    D. Self-aware AI
13. **Where is Limited Memory AI commonly found?**
    A. Self-driving cars
    B. Basic chess programs
    C. Simple thermostats
    D. Calculators
14. **What is Theory of Mind AI (still in development)?**
    A. AI that understands emotions, beliefs, and intentions of others
    B. AI with no memory
    C. AI that only reacts to current inputs
    D. AI with physical bodies
15. **Which functionality-based AI type is purely hypothetical?**
    A. Self-aware AI
    B. Limited Memory AI

C. Reactive Machines
D. Theory of Mind AI

16. **Which school of AI is based on rules, logic, and symbols?**
    A. Symbolism (Symbolic AI)
    B. Connectionism
    C. Behaviorism
    D. Empiricism
17. **What is an example of Symbolism?**
    A. Expert systems and rule-based reasoning
    B. Deep learning and neural networks
    C. Reinforcement learning
    D. Genetic algorithms
18. **Which school is inspired by the human brain and learns from data using interconnected nodes?**
    A. Connectionism
    B. Symbolism
    C. Behaviorism
    D. Functionalism
19. **What are examples of Connectionism?**
    A. Deep learning, ANN, CNN, RNN
    B. Expert systems
    C. Rule-based programming
    D. Logical reasoning
20. **Which school focuses on agent behavior and responses to the environment (stimulus → action)?**
    A. Behaviorism
    B. Symbolism
    C. Connectionism
    D. Structuralism
21. **What is an example of Behaviorism?**
    A. Reinforcement learning, robotics
    B. Neural networks
    C. Expert systems
    D. Genetic algorithms
22. **What does NLP stand for and what does it do?**
    A. Natural Language Processing - enables computers to understand and generate human languages
    B. Neural Language Programming - creates neural networks
    C. Natural Learning Process - learns from nature
    D. Network Language Protocol - manages network communications

23. **Which of these is NOT an NLP application?**
    A. Sentiment analysis
    B. Machine translation
    C. Image recognition
    D. Chatbots
24. **What is Computer Vision (CV)?**
    A. Technology that gives machines the ability to 'see' and understand visual information
    B. Making computers understand speech
    C. Hardware for visual displays
    D. Programming visual interfaces
25. **Which CV task involves partitioning an image into multiple segments?**
    A. Image Segmentation
    B. Object Detection
    C. Object Tracking
    D. OCR
26. **What does Object Detection do?**
    A. Detects and locates objects in images
    B. Tracks objects across video frames
    C. Reads text from images
    D. Classifies entire images
27. **What is the process of following a specific object's movement across video frames?**
    A. Object Tracking
    B. Object Detection
    C. Image Segmentation
    D. Optical Character Recognition
28. **What does OCR stand for and what does it do?**
    A. Optical Character Recognition - recognizes text in images and converts to editable text
    B. Optical Computer Recognition - recognizes computers in images
    C. Optical Color Recognition - identifies colors
    D. Optical Character Reading - reads aloud text
29. **What are the three key elements needed for AI development according to the lecture?**
    A. Data, Algorithms, Computing Power
    B. Hardware, Software, Users
    C. Input, Processing, Output
    D. Sensors, Processors, Actuators

30. **Which industry applications were mentioned for AI technologies?**
    A. Healthcare, Finance, Transportation, Entertainment
    B. Only manufacturing
    C. Only scientific research
    D. Only military applications
31. **What development stages were mentioned for NLP?**
    A. Rules/linguistics → ML/DL → Large Language Models (LLMs)
    B. Only deep learning
    C. Hardware improvements only
    D. From simple to complex hardware
32. **Which of these is a Computer Vision task that helps in medical imaging to isolate tumors?**
    A. Image Segmentation
    B. Object Detection
    C. OCR
    D. Object Tracking
33. **In autonomous driving, which CV task helps understand exact road boundaries?**
    A. Image Segmentation
    B. Object Detection
    C. Face Recognition
    D. Text Recognition

Lecture 3: Machine Learning Concepts and Algorithms

1. **What is the main difference between traditional rule-based methods and machine learning?**
   A. ML learns patterns from data automatically, while rule-based methods require explicit programming
   B. ML is slower but more accurate
   C. Rule-based methods can handle more complex problems
   D. ML doesn't require data at all
2. **When should you consider using machine learning instead of traditional programming?**
   A. When rules are complex, change over time, or data distribution changes
   B. When problems are simple and well-defined
   C. When you have very little data
   D. When you need 100% deterministic results

3. **Which scenario is ideal for machine learning application?**
   A. Speech recognition where rules are complex to describe
   B. Adding two numbers
   C. Sorting a list alphabetically
   D. Calculating tax based on fixed percentages
4. **What happens when "task rules change over time" that makes ML suitable?**
   A. ML models can adapt to new data and changing patterns
   B. ML models become obsolete quickly
   C. ML requires less maintenance than rule-based systems
   D. ML doesn't need to be updated
5. **Why is ML better than rule-based systems when data distribution changes?**
   A. ML can adapt to new data patterns through retraining
   B. ML systems never need updating
   C. Rule-based systems are always better
   D. ML works perfectly without new data
6. **Which type of learning uses labeled data where each input has a known correct output?**
   A. Supervised Learning
   B. Unsupervised Learning
   C. Reinforcement Learning
   D. Semi-supervised Learning
7. **What characterizes Unsupervised Learning?**
   A. Works with unlabeled data, finding hidden patterns without correct answers provided
   B. Uses only labeled data with known outputs
   C. Learns from rewards and penalties
   D. Combines labeled and unlabeled data
8. **In Reinforcement Learning, how does the agent learn?**
   A. Through trial and error, receiving rewards for good actions and penalties for bad ones
   B. From labeled examples provided by a teacher
   C. By finding patterns in unlabeled data
   D. By following explicit programming rules
9. **What is Semi-supervised Learning?**
   A. Uses a small amount of labeled data combined with a large amount of unlabeled data
   B. Uses only unlabeled data
   C. Uses only labeled data
   D. Doesn't use any data

10. **Which learning type would you use to predict whether an email is spam?**
    A. Supervised Learning
    B. Unsupervised Learning
    C. Reinforcement Learning
    D. Active Learning
11. **What type of problem is "predicting house prices based on features"?**
    A. Regression (predicting continuous values)
    B. Classification (predicting categories)
    C. Clustering (grouping similar items)
    D. Reinforcement (maximizing rewards)
12. **Which problem type assigns inputs to one of k categories?**
    A. Classification
    B. Regression
    C. Clustering
    D. Dimensionality reduction
13. **What does Clustering do?**
    A. Groups unlabeled data by similarity
    B. Predicts continuous values
    C. Assigns categories to labeled data
    D. Maximizes reward signals
14. **What's the main difference between Classification and Regression outputs?**
    A. Classification outputs discrete classes, Regression outputs continuous values
    B. Classification is faster than Regression
    C. Regression only works with numbers, Classification with text
    D. Classification requires more data
15. **Which is an example of a Regression problem?**
    A. Predicting stock prices next week
    B. Detecting spam emails
    C. Grouping customers by buying behavior
    D. Learning to play chess
16. **In supervised learning, what is the "teacher"?**
    A. The labeled data that provides correct answers
    B. The programmer who writes the rules
    C. The test data used for evaluation
    D. The algorithm itself
17. **What does the model do during supervised learning?**
    A. Learns by comparing predictions with correct answers and improves over time
    B. Finds patterns without any guidance

C. Explores the environment through trial and error
D. Only works with perfect data

18. **Which is an example of supervised learning for classification?**
    A. Determining if an email is spam based on labeled examples
    B. Grouping similar news articles together
    C. Recommending movies based on user behavior
    D. Learning to drive through trial and error
19. **What is the output of a regression model?**
    A. A continuous numerical value
    B. A category or class label
    C. A group or cluster
    D. A reward signal
20. **In the patient clinic example from lecture, what type of supervised learning was shown?**
    A. Classification (healthy vs sick)
    B. Regression (predicting temperature)
    C. Clustering (grouping patients)
    D. Reinforcement (treating patients)
21. **What is the main challenge in unsupervised learning?**
    A. No correct answers are provided, so evaluation is harder
    B. Too much labeled data is required
    C. It's slower than supervised learning
    D. It can't find any patterns
22. **Which algorithm is commonly used for clustering?**
    A. K-means
    B. Linear Regression
    C. Decision Tree
    D. Support Vector Machine
23. **What does "intra-cluster similarity" mean in clustering?**
    A. Items within the same cluster are more similar to each other
    B. Items in different clusters are similar
    C. All clusters are equally similar
    D. No similarity measures are used
24. **Which is an example of unsupervised learning application?**
    A. Recommending movies for users based on viewing patterns
    B. Predicting house prices
    C. Classifying emails as spam
    D. Playing chess against a computer
25. **How does unsupervised learning handle new data?**
    A. Calculates similarity with existing samples and classifies by similarity

B. Compares with known labels
C. Asks for user feedback
D. Randomly assigns to clusters

26. **What is the key characteristic of Reinforcement Learning?**
    A. Learns through interaction with environment using rewards/penalties
    B. Requires large amounts of labeled data
    C. Finds patterns in unlabeled data
    D. Works only with static environments
27. **What does the "reward function" do in RL?**
    A. Provides feedback to guide learning towards desired behavior
    B. Labels data for training
    C. Groups similar states together
    D. Initializes the model parameters
28. **What does the agent try to maximize in RL?**
    A. Cumulative reward over time
    B. Training accuracy
    C. Number of correct predictions
    D. Speed of learning
31. **Why is semi-supervised learning useful?**
    A. Labeled data is often expensive/time-consuming to obtain
    B. It's always more accurate than supervised learning
    C. It doesn't require any data
    D. It's simpler to implement
32. **What proportion of data is typically labeled in semi-supervised learning?**
    A. Small amount of labeled, large amount of unlabeled
    B. All data is labeled
    C. All data is unlabeled
    D. Equal amounts of labeled and unlabeled
33. **Which learning types does semi-supervised combine?**
    A. Supervised and Unsupervised Learning
    B. Reinforcement and Supervised
    C. Reinforcement and Unsupervised
    D. Only unsupervised techniques
34. **Why is RL good for robotics and gaming?**
    A. These involve sequential decisions where actions affect future states
    B. They have lots of labeled data
    C. They are simple classification problems
    D. They don't require learning
35. **What type of learning would you use for "defining fish of the same species" without labels?**

A. Clustering (Unsupervised Learning)
B. Classification (Supervised Learning)
C. Regression
D. Reinforcement Learning

36. **For "recommending movies for users" based on viewing history without explicit ratings, which learning type?**
A. Unsupervised Learning (finding patterns in viewing behavior)
B. Supervised Learning
C. Reinforcement Learning
D. Semi-supervised Learning

Lecture 4: The Machine Learning Process: From Raw Data to Effective Models

1. **What is the complete machine learning process sequence?**
A. Data preparation → Data cleansing → Feature extraction/selection → Model training → Model evaluation → Deployment → Feedback & iteration
B. Model training → Data cleansing → Feature selection → Deployment → Feedback
C. Data collection → Model training → Immediate deployment
D. Only model training and evaluation
2. **What is the purpose of the feedback and iteration step in ML?**
A. Gather real-world feedback and use it to improve the model iteratively
B. Stop the process once deployed
C. Only collect more data without model changes
D. Wait for the model to improve automatically
3. **Which step comes immediately after data cleansing?**
A. Feature extraction and selection
B. Model deployment
C. Model evaluation
D. Data collection
4. **What is a "dataset" in machine learning?**
A. A collection of data used in ML tasks, where each piece is called a sample
B. A single algorithm
C. Only the training data
D. The final model
5. **What are "features" in a dataset?**
A. Items or attributes that reflect the nature of a sample in a particular

aspect
B. Only the output labels
C. The algorithms used
D. The evaluation metrics

6. **What is the difference between a training set and a test set?**
A. Training set: used for building the model; Test set: used for evaluating the learned model
B. Both are used for training
C. Training set: for evaluation; Test set: for building
D. They are the same thing
7. **What does "learning" or "training" refer to in ML?**
A. The process of building a model from data
B. The process of testing a model
C. The process of collecting data
D. The process of deploying the model
8. **In the typical dataset composition diagram, what are the components?**
A. Feature 1, Feature 2, Feature 3, Label
B. Only features without labels
C. Only labels without features
D. Model parameters
9. **Why is data processing crucial in machine learning?**
A. Data determines the scope of model capabilities; all good models require good data
B. It's optional and can be skipped
C. Only needed for complex models
D. It only affects training speed
10. **What is the "curse of dimensionality"?**
A. Having too many features (dimensions) compared with the amount of data
B. Having too few features
C. Having too much data
D. Having too few samples
12. **What are the common operations in data preprocessing?**
A. Data filtering, handling data loss, handling errors/abnormal values, merging data from multiple sources, data consolidation
B. Only training the model
C. Only collecting more data
D. Only feature selection
13. **What does "dirty data" typically contain?**
A. Incompleteness, noise, inconsistency

B. Too many features
C. Perfect structure
D. Already cleaned information

14. **What is "incompleteness" in data quality problems?**
    A. Incomplete data or lack of relevant attributes/values
    B. Too much data
    C. Correct records
    D. Consistent data
15. **What is "noise" in data quality problems?**
    A. Data contains incorrect records or abnormal points
    B. Missing values
    C. Consistent formatting
    D. Too many features
16. **What is "inconsistency" in data quality problems?**
    A. Data contains conflicting records
    B. Missing data
    C. Too much data
    D. Perfect formatting
17. **What operations are typically used to convert data for ML models?**
    A. Encoding categorical data, converting numeric to categorical, word embedding, image processing, feature engineering
    B. Only using raw data as is
    C. Only collecting more data
    D. Only removing data
18. **What is feature engineering?**
    A. Normalizing/standardizing features and creating new features from existing ones
    B. Only selecting features
    C. Only removing features
    D. Only collecting data
19. **Why might you convert numeric data into categorical data?**
    A. To reduce the number of variable values (e.g., segmenting age data)
    B. To increase complexity
    C. To make data more random
    D. To remove all patterns
20. **Why is feature selection necessary?**
    A. Some features may be unnecessary or irrelevant to the values to be predicted
    B. To always increase model complexity

C. To remove all features
D. To make training slower

21. **What are the three main types of feature selection methods?**
A. Filter methods, Wrapper methods, Embedded methods
B. Only statistical methods
C. Only manual selection
D. Only automatic selection
22. **How do Filter methods work?**
A. Select features based on statistical tests, evaluating each feature independently with target variable
B. Use model performance to score feature subsets
C. Build feature selection into model training
D. Only use domain knowledge
23. **How do Wrapper methods work?**
A. Use a prediction model to score feature subsets based on model performance
B. Use statistical tests only
C. Don't use any models
D. Only consider one feature at a time
24. **How do Embedded methods work?**
A. Treat feature selection as part of the modeling process itself
B. Work independently of models
C. Only use statistical tests
D. Require manual intervention
25. **What is a limitation of Filter methods?**
A. Tend to select redundant variables because they don't consider relationships between features
B. Too slow
C. Don't use any statistics
D. Always pick the best features
26. **What is a limitation of Wrapper methods?**
A. Computationally intensive because they train a new model for each feature subset
B. Too simple
C. Don't use models at all
D. Always give perfect results
27. **Which feature selection method combines benefits of both filter and wrapper methods?**
A. Embedded methods
B. Only filter methods

C. Only wrapper methods
D. Manual selection

28. **What are the 6 steps in building an AI model according to the lecture?**
A. Split data → Train model → Validate model → Tune model → Deploy model → Test model
B. Only train and test
C. Collect data → Train → Deploy
D. Clean data → Train → Evaluate
29. **What is the purpose of splitting data into training, validation, and test sets?**
A. Training: build model; Validation: tune model; Test: evaluate generalization capability
B. All for training
C. All for testing
D. Only for data collection
30. **What is the validation set used for?**
A. To evaluate model effectiveness and tune the model
B. Only for final testing
C. Only for training
D. To collect more data
31. **What is the test set used for?**
A. To evaluate the generalization capability of the model
B. For training the model
C. For tuning parameters
D. For data cleaning
34. **What is "generalization" in machine learning?**
A. The model's ability to perform well on new, unseen samples
B. Performing well only on training data
C. Being very complex
D. Having many features
35. **What is training error?**
A. Error of the model on the training set
B. Error on new samples
C. Error during deployment
D. Error in data collection
36. **What is generalization error?**
A. Error of the model on new samples
B. Error on training data
C. Error during model training
D. Error in algorithm selection

37. **What is underfitting?**
    A. When training error is large (model too simple)
    B. When training error is small but generalization error is large
    C. When model is perfectly balanced
    D. When model has too many parameters
38. **What is overfitting?**
    A. When training error is small but generalization error is large (model too complex)
    B. When training error is large
    C. When model is too simple
    D. When data is perfectly clean
39. **What is model capacity (complexity)?**
    A. The capability of the model to fit various functions
    B. The speed of training
    C. The number of features
    D. The amount of data
40. **What happens with insufficient model capacity?**
    A. Underfitting may occur (can't handle complex tasks)
    B. Overfitting occurs
    C. Perfect fit always
    D. Model trains faster
41. **What happens with excessive model capacity?**
    A. Overfitting may occur (learns noise)
    B. Underfitting occurs
    C. Perfect generalization
    D. No training needed

Lecture 5: Machine Learning Model Evaluation - Comprehensive Revision Questions

1. **What is the primary goal of machine learning?**
   A. Ensure models perform well on new, unseen samples (generalization)
   B. Achieve perfect performance on training data only
   C. Create the most complex model possible
   D. Use all available features without selection
2. **What does "error" in machine learning refer to?**
   A. The difference between the model's prediction and the actual result
   B. The number of features used
   C. The training time required
   D. The model's complexity

3. **What is training error?**
   A. The error of the model on the training set
   B. The error on new, unseen samples
   C. The error during model deployment
   D. The error in data collection
4. **What is generalization error?**
   A. The error of the model on new samples
   B. The error on training data
   C. The error during parameter tuning
   D. The error in algorithm implementation
5. **Which type of error should we aim to minimize in practical applications?**
   A. Generalization error
   B. Training error only
   C. Both equally
   D. Neither - we want maximum error
6. **What happens when model capacity is too large?**
   A. Overfitting occurs (model learns noise)
   B. Underfitting occurs
   C. Both errors become zero
   D. Model cannot train

11. **In binary classification, what are the possible model outputs?**
    A. Correct or incorrect predictions
    B. Continuous values
    C. Multiple categories
    D. Probability distributions only
12. **What is the typical number of classes in binary classification?**
    A. Two possible classes (e.g., Dog or Cat)
    B. Three or more classes
    C. Only one class
    D. Unlimited classes
13. **What happens in the supervised learning process for classification?**
    A. Train on labeled training data → Test on testing data → Compare predictions with ground truth
    B. Only train without testing
    C. Test first then train
    D. Use only unlabeled data
14. **After training, what does the model predict?**
    A. Labels for X_test (testing data)
    B. Labels for training data only

C. Random labels
D. Only the most common class

15. **What are "ground truth labels" (y values)?**
A. The actual correct labels for the test data
B. The predicted labels
C. The feature values
D. The model parameters

16. **What metrics can be derived from a confusion matrix?**
A. Recall, Precision, Specificity, Accuracy, AUC-ROC
B. Only training error
C. Only model size
D. Only data collection rate

17. **What is the formula for accuracy?**
A. (TP + TN) / (TP + TN + FP + FN)
B. TP / (TP + FN)
C. TP / (TP + FP)
D. TN / (TN + FP)

18. **What does accuracy measure?**
A. Overall correctness: total correct predictions divided by total predictions
B. How well we catch positive cases
C. How accurate positive predictions are
D. How well we identify negative cases

21. **What is the formula for Recall?**
A. TP / (TP + FN)
B. TP / (TP + FP)
C. (TP + TN) / total
D. TN / (TN + FP)

22. **What does Precision measure?**
A. "Out of everything predicted as Positive, how many were actually correct?"
B. "Out of actual positives, how many were caught?"
C. Overall accuracy
D. Error rate

23. **What is the formula for Precision?**
A. TP / (TP + FP)
B. TP / (TP + FN)
C. TN / (TN + FP)
D. (TP + TN) / total

24. **What is the F1-score (F-score)?**
A. Harmonic mean of Precision and Recall

B. Arithmetic mean of Precision and Recall
C. Difference between Precision and Recall
D. Sum of Precision and Recall

25. **When is F1-score particularly useful?**
A. When we need to balance catching positives (Recall) and avoiding false alarms (Precision)
B. When we only care about accuracy
C. When we have unlimited data
D. When model is very simple

26. **What is the formula for F1-score?**
A. 2 * (Precision * Recall) / (Precision + Recall)
B. (Precision + Recall) / 2
C. Precision * Recall
D. Precision - Recall

30. **What does AUC measure?**
A. Overall performance of binary classifier across all thresholds
B. Training time
C. Model size
D. Data collection rate

31. **What is a perfect classifier's AUC value?**
A. 1.0
B. 0.5
C. 0.0
D. 0.75

32. **What does an AUC of 0.5 indicate?**
A. Random classifier (no discriminative power)
B. Perfect classifier
C. Good classifier
D. Overfitted classifier

46. **What are model parameters?**
A. Automatically learned by models from data (e.g., weights, biases)
B. Manually set before training
C. Only related to data preprocessing
D. Only for model evaluation

47. **What are hyperparameters?**
A. Manually set before training to control learning process
B. Learned automatically from data
C. Only for data cleaning
D. Only for model deployment

48. **What do hyperparameters control?**
    A. Learning rate, number of layers, batch size, regularization strength
    B. Model predictions
    C. Data collection
    D. Feature values
49. **How do we typically find good hyperparameters?**
    A. Through hyperparameter search (grid search, random search)
    B. They are fixed for all models
    C. Guessed randomly
    D. Only from theoretical calculation
50. **What is cross-validation?**
    A. A statistical method to check classifier performance by splitting data
    B. A type of neural network
    C. A feature selection method
    D. A data collection technique
51. **How does basic cross-validation work?**
    A. Split data into training and validation sets, train on one, test on other
    B. Use all data for both training and testing
    C. Only use test data
    D. Ignore validation sets
52. **What is k-fold cross-validation?**
    A. Split dataset into K equal parts, train on K-1 folds, test on remaining fold, repeat K times
    B. Use only one fold for everything
    C. Train and test on same data
    D. Only use validation data
53. **What is a common choice for K in k-fold CV?**
    A. 5 or 10
    B. 1 or 2
    C. 100
    D. Same as number of samples
54. **What does k-fold CV help reduce?**
    A. Bias (overfitting to one particular train/test split)
    B. Training time
    C. Model size
    D. Data collection effort
55. **What do we do with the k-fold CV results?**
    A. Average them to get a more stable estimate of performance
    B. Use only the best fold

C. Ignore variations
D. Only report minimum

56. **Why is cross-validation particularly useful?**
A. More reliable evaluation using all data for both training and testing
B. Faster training
C. Simpler implementation
D. Requires less data
57. **In 5-fold CV with 100 samples, how many samples in each fold?**
A. 20 samples per fold
B. 100 samples
C. 5 samples
D. 50 samples
58. **What does cross-validation help us choose?**
A. The best model that generalizes well
B. Only the fastest model
C. Only the simplest model
D. Only models with most parameters

Lecture 6: Linear Regression and Model Evaluation in Machine Learning - Comprehensive Revision Questions

1. **What does linear regression predict?**
A. Continuous values
B. Categories/classes
C. Clusters/groups
D. Sequences
2. **What type of machine learning is linear regression?**
A. Supervised learning
B. Unsupervised learning
C. Reinforcement learning
D. Semi-supervised learning
3. **What does linear regression assume?**
A. A linear relationship between input (independent) and output (dependent) variables
B. No relationship between variables
C. Complex non-linear relationships
D. Random relationships
4. **What is the independent variable in linear regression?**
A. The predictor variable (X)
B. The target variable (Y)

C. The error term
D. The intercept

5. **What is the dependent variable in linear regression?**
   A. The target variable we want to predict (Y)
   B. The predictor variable (X)
   C. The coefficient
   D. The learning rate
6. **Which term represents the predicted value in linear regression?**
   A. ŷ (y-hat)
   B. x
   C. ε (epsilon)
   D. β (beta)
7. **What is the simplest form of linear regression?**
   A. Simple linear regression: y = b0 + b1*x
   B. Multiple regression with many variables
   C. Polynomial regression
   D. Logistic regression

27. **What is the regression line?**
   A. The best-fitting line through the data points that minimizes error
   B. Any line through the data
   C. A curved line
   D. A horizontal line
28. **In the salary vs. experience example, what does the regression line show?**
   A. How salary changes with experience
   B. Random relationship
   C. No relationship
   D. Categorical differences
29. **What is a residual?**
   A. The difference between actual value and predicted value
   B. The slope of the line
   C. The intercept
   D. The independent variable
30. **What does ε (epsilon) represent in regression?**
   A. The error term (difference between actual and predicted)
   B. The slope coefficient
   C. The intercept
   D. The learning rate
31. **What is the "point estimate" of the error term called?**
   A. Residual
   B. Coefficient

C. Intercept
D. Variance

32. **What method does linear regression use to find the best-fitting line?**
    A. Least Squares Method
    B. Gradient Descent only
    C. Random selection
    D. Maximum likelihood
33. **What does the Least Squares method minimize?**
    A. Sum of squared differences between actual and predicted values
    B. Sum of absolute differences
    C. Number of errors
    D. Model complexity
34. **What are the "squared differences" being minimized?**
    A. Residuals squared
    B. Features squared
    C. Coefficients squared
    D. Inputs squared
35. **What is the hypothesis (model) in linear regression?**
    A. $\hat{y} = b0 + b1*x$ (for simple regression)
    B. $y = mx + c$
    C. $y = ax^2 + bx + c$
    D. y = e^x
36. **What is the error for each data point?**
    A. $(\hat{y} - y)$ or $(y - \hat{y})$
    B. $(x - \hat{y})$
    C. (b0 - b1)
    D. (x - y)
37. **What cost function is commonly used in linear regression?**
    A. Mean Squared Error (MSE)
    B. Cross-entropy
    C. Accuracy
    D. Precision
38. **What is the formula for Mean Squared Error?**
    A. $(1/n) * \Sigma(\hat{y} - y)^2$
    B. $(1/n) * \Sigma|\hat{y} - y|$
    C. $\Sigma(\hat{y} - y)$
    D. $(\hat{y} - y)^2/n^2$
39. **What does MSE measure?**
    A. Average of squared differences between predicted and actual values
    B. Number of correct predictions

C. Model complexity
D. Training time

42. **What is simple linear regression?**
A. One independent variable predicting one dependent variable: y = b0 + b1*x
B. Multiple variables predicting one output
C. One variable predicting multiple outputs
D. No variables
43. **What is multiple linear regression?**
A. Multiple independent variables predicting one dependent variable: y = b0 + b1*x1 + b2*x2 + ...
B. One variable predicting one output
C. Multiple variables predicting multiple outputs
D. No clear relationship
44. **In multiple regression equation y = b0 + b1*x1 + b2*x2, what is b0?**
A. Constant (intercept)
B. Coefficient for x1
C. Coefficient for x2
D. Error term
45. **In multiple regression equation y = b0 + b1*x1 + b2*x2, what is b1?**
A. Coefficient for variable x1
B. Constant term
C. Coefficient for x2
D. Dependent variable
46. **How does adding more independent variables affect regression?**
A. Can capture more complex relationships but risks overfitting
B. Always improves accuracy
C. Makes model simpler
D. Reduces training time
47. **What happens if we use linear regression for non-linear data?**
A. Poor predictions due to incorrect model assumption
B. Perfect predictions
C. No predictions possible
D. Random results

Lecture 8: Random Forest and Ensemble Learning Algorithms

1. **How is a decision tree oriented?**
   A. Upside down with root at the top
   B. Left to right
   C. Circular
   D. Diagonal
2. **What type of questions does a decision tree ask?**
   A. True/False questions at each node
   B. Multiple choice questions
   C. Essay questions
   D. No questions
3. **What is the goal of decision tree questions?**
   A. To un-mix the data or reduce uncertainty
   B. To make the tree look complex
   C. To use all features equally
   D. To confuse the user
4. **How does a decision tree split data?**
   A. Into two or more subsets based on answers to questions
   B. Into equal parts randomly
   C. Into fixed number of parts
   D. Doesn't split data
5. **What algorithm step calculates how well features split data?**
   A. Using Gini Impurity (GI)
   B. Using accuracy only
   C. Using random selection
   D. Using human intuition
6. **What is Random Forest?**
   A. A machine learning algorithm that uses many decision trees to make better predictions
   B. A single decision tree
   C. A type of linear regression
   D. A clustering algorithm
7. **How does Random Forest combine tree results for classification?**
   A. Majority voting
   B. Averaging
   C. Multiplication
   D. Using the first tree only

8. **How does Random Forest combine tree results for regression?**
   A. Averaging
   B. Majority voting
   C. Random selection
   D. Weighted sum
9. **What helps Random Forest avoid overfitting?**
   A. Using random data and features for each tree
   B. Making all trees identical
   C. Using only one tree
   D. Using all features all the time
10. **What makes Random Forest predictions more trustworthy?**
   A. Combining multiple trees reduces individual errors
   B. Using only the best tree
   C. Ignoring some trees
   D. Always choosing the simplest answer
11. **What is the first step in Random Forest algorithm?**
   A. Bootstrap Sampling: creating random subsets with replacement
   B. Training all trees on same data
   C. Selecting features manually
   D. Pruning trees first
12. **What does "sampling with replacement" mean in bootstrap?**
   A. Same sample can appear multiple times in a subset
   B. Each sample appears exactly once
   C. No samples are repeated
   D. Only unique samples are used
13. **How are trees constructed in Random Forest?**
   A. Each tree considers only random subset of features at each split
   B. All trees use all features
   C. Trees use same features as each other
   D. No feature selection occurs
14. **What happens after all trees are built?**
   A. Predictions are aggregated using voting (classification) or averaging (regression)
   B. Only best tree is kept
   C. All trees are discarded
   D. Trees are combined into one big tree
15. **What does Random Forest specifically help reduce?**
   A. Overfitting and improves model accuracy
   B. Training time only

C. Model interpretability
D. Feature importance

16. **Is Random Forest a supervised or unsupervised algorithm?**
    A. Supervised algorithm
    B. Unsupervised algorithm
    C. Semi-supervised
    D. Reinforcement learning
17. **What happens when you increase number of trees in Random Forest?**
    A. Improves model stability and accuracy (up to a point)
    B. Always decreases accuracy
    C. No effect on performance
    D. Makes model less stable
18. **How does Random Forest differ from a single decision tree?**
    A. Introduces randomness in data and feature selection
    B. Uses same dataset and rules
    C. Is simpler to interpret
    D. Has fewer parameters
19. **What assumptions does Random Forest make?**
    A. Each tree makes its own decisions, uses random data parts, needs enough data
    B. All trees should agree perfectly
    C. Only one tree is needed
    D. Features are perfectly independent
20. **What is bootstrapping?**
    A. Taking random samples from data with replacement many times
    B. Sampling without replacement
    C. Using entire dataset always
    D. Manual selection of samples
21. **What is created through bootstrapping?**
    A. New "fake" datasets that are slightly different
    B. Exact copies of original data
    C. Completely unrelated datasets
    D. Only one dataset
22. **What do we calculate for each bootstrap dataset?**
    A. Results like average to understand reliability
    B. Nothing - just store them
    C. Only maximum value
    D. Random numbers
23. **What does bootstrapping help us understand?**
    A. How stable or reliable our original result is

B. Only the mean value
C. Only the variance
D. Nothing useful

24. **When is bootstrapping particularly useful?**
    A. When we have small dataset
    B. Only with large datasets
    C. Never useful
    D. Only for classification
25. **Where is bootstrapping used in machine learning?**
    A. In Random Forests to make models more reliable
    B. Only in linear regression
    C. Only in neural networks
    D. Not used in ML
26. **What is ensemble learning?**
    A. Combining multiple models to improve performance
    B. Using single best model
    C. Training models sequentially
    D. Only using similar models
27. **What problem do individual models often have?**
    A. May overfit training data or underperform on unseen instances
    B. Always perfect performance
    C. Never overfit
    D. Too simple
28. **How do ensembles solve individual model problems?**
    A. By aggregating models and balancing their errors
    B. By using only one model
    C. By making all models identical
    D. By ignoring errors
29. **In the ensemble diagram, how are individual model predictions combined?**
    A. Through aggregation to produce final ensemble classifier
    B. By selecting only one
    C. By averaging only
    D. By discarding some
30. **Why do ensembles often work better than individual models?**
    A. Different models make different errors, ensemble averages them out
    B. All models always agree
    C. Ensembles are always simpler
    D. Requires less computation
31. **What are the two main ways to build ensemble models?**
    A. Independent construction and coordinated construction

B. Only sequential building
C. Only parallel building
D. Random building

32. **How does independent ensemble construction work?**
A. Each model trained separately, randomness added, results combined
B. Models built one after another
C. Only one model trained
D. No randomness added

33. **What is an example of independent ensemble construction?**
A. Random Forest (each tree built independently)
B. AdaBoost
C. Gradient Boosting
D. Neural networks

34. **How does coordinated ensemble construction work?**
A. Models built one after another, each learns from previous mistakes
B. All models trained at same time
C. Models don't communicate
D. Random training order

36. **What are the three main types of ensemble classifiers?**
A. Bagging, Boosting, Stacking
B. Only voting
C. Only averaging
D. Only weighting

37. **What does "Bagging" stand for?**
A. Bootstrap Aggregating
B. Best Aggregating
C. Binary Aggregating
D. Basic Aggregating

38. **How does Bagging work?**
A. Trains multiple models independently in parallel using bootstrap samples
B. Trains models sequentially
C. Uses only original dataset
D. No bootstrap sampling

39. **What does Bagging specifically reduce?**
A. Variance and helps prevent overfitting
B. Only bias
C. Only training time
D. Model accuracy

40. **What is an advantage of Bagging?**
A. Models trained in parallel, making it efficient

B. Always highest accuracy
C. Simplest to implement
D. No need for multiple models

41. **How are bootstrap datasets created in Bagging?**
    A. By randomly sampling with replacement
    B. By sampling without replacement
    C. By manual selection
    D. By using all data unchanged
42. **What type of base learner is often used in Bagging?**
    A. Decision tree
    B. Linear regression
    C. Neural network
    D. Support vector machine
43. **How are predictions combined in Bagging?**
    A. Averaging (regression) or majority voting (classification)
    B. Only using best model
    C. Weighted average
    D. Random selection
44. **What problem does Bagging specifically address?**
    A. High variance in unstable models like decision trees
    B. High bias
    C. Small datasets only
    D. Feature selection
45. **What is Boosting?**
    A. Builds models sequentially so each learns from errors of previous ones
    B. Builds models in parallel
    C. Uses only one model
    D. Random model building
46. **How does Boosting improve models?**
    A. Improving bias and accuracy by focusing on difficult instances
    B. Only reducing variance
    C. Only increasing speed
    D. Simplifying models
47. **What happens to misclassified samples in Boosting?**
    A. Receive higher weights in subsequent iterations
    B. Are ignored
    C. Are removed
    D. Get lower weights
48. **How long does the Boosting process continue?**
    A. For multiple iterations/specified number of models

B. Until perfect accuracy
C. Only one iteration
D. Random number of iterations

49. **What is an advantage of Boosting?**
    A. Can turn weak learners into strong ones
    B. Always fastest
    C. Simplest to implement
    D. Never overfits
50. **Where does Boosting work particularly well?**
    A. With structured data, provides high accuracy
    B. Only with images
    C. Only with text
    D. Only with small datasets
51. **What type of model does Boosting typically start with?**
    A. Weak base model (e.g., shallow decision tree)
    B. Complex neural network
    C. Perfect classifier
    D. Random guesser
52. **What happens to sample weights after each iteration in Boosting?**
    A. Increased for misclassified samples
    B. Decreased for all samples
    C. Stay same
    D. Only correct samples get weights
53. **How are predictions combined in Boosting?**
    A. Combine predictions of all models
    B. Use only last model
    C. Random selection
    D. Average of first and last
54. **What is Stacking?**
    A. Combines multiple models of different types using a meta-model
    B. Only uses same type models
    C. No meta-model used
    D. Simple averaging
55. **How are base models trained in Stacking?**
    A. Independently, then outputs used as inputs to meta-learner
    B. Sequentially
    C. Only one model trained
    D. With same data and features
56. **What does the meta-learner do?**
    A. Learns best way to merge base model predictions

B. Makes initial predictions
C. Selects features
D. Cleans data

57. **What is commonly used as the meta-learner?**
    A. Logistic regression
    B. Deep neural network
    C. Decision tree
    D. Random forest
58. **What type of models can Stacking combine?**
    A. Different types (e.g., decision trees, SVMs, neural networks)
    B. Only identical models
    C. Only two models
    D. Only regression models
59. **What advantage does Stacking provide?**
    A. Leverages strengths of various models, often improving accuracy
    B. Always simplest solution
    C. Fastest training
    D. Most interpretable
60. **What is the key idea behind all ensemble methods?**
    A. Wisdom of crowds - multiple models better than one
    B. Simplicity is best
    C. One perfect model enough
    D. More data always better
61. **How do Bagging and Boosting differ in model training?**
    A. Bagging: parallel, independent; Boosting: sequential, dependent
    B. Both are sequential
    C. Both are parallel
    D. No difference
62. **Which ensemble method reduces variance primarily?**
    A. Bagging
    B. Boosting
    C. Stacking
    D. All equally
63. **Which ensemble method reduces bias primarily?**
    A. Boosting
    B. Bagging
    C. Stacking
    D. All equally
64. **Which ensemble method can combine different model types?**
    A. Stacking

B. Bagging
C. Boosting
D. None can

65. **What is a common characteristic of all ensemble methods?**
    A. They combine multiple models to improve performance
    B. They use only one model
    C. They are always complex
    D. They don't need training data
66. **When would you choose Random Forest over single Decision Tree?**
    A. When you need better accuracy and reduced overfitting
    B. When interpretability is most important
    C. When you have very little data
    D. When training time must be minimized
67. **What problem does ensemble learning solve that single models struggle with?**
    A. Balancing bias-variance tradeoff
    B. Data collection
    C. Feature engineering
    D. Model interpretation
68. **What does "weak learner" mean in Boosting context?**
    A. A model that performs slightly better than random guessing
    B. A perfect classifier
    C. A very complex model
    D. A model that always fails
69. **How does Random Forest's feature randomness help?**
    A. Makes trees more diverse, reducing correlation between them
    B. Makes all trees identical
    C. Reduces training time only
    D. Increases overfitting
70. **What is the trade-off with ensemble methods?**
    A. Improved performance but increased complexity and computation
    B. Always simpler and faster
    C. Never any trade-off
    D. Worse performance but more interpretable

Lecture 9: K-Nearest Neighbors, Naïve Bayes & Neural Networks Introduction

1. **What is K-Nearest Neighbors (KNN)?**
   A. An algorithm that classifies based on majority class of k nearest neighbors
   B. A deep neural network architecture
   C. A clustering algorithm only
   D. A linear regression model
2. **For regression tasks, how does KNN make predictions?**
   A. Predicts the average of the k nearest neighbors' values
   B. Uses majority voting
   C. Always predicts the same value
   D. Uses a linear combination
3. **What does the parameter "k" represent in KNN?**
   A. Number of neighbors to consider
   B. Number of features
   C. Number of classes
   D. Number of training samples
4. **How does KNN determine which neighbors are "nearest"?**
   A. Using distance metrics like Euclidean distance
   B. Random selection
   C. Feature importance
   D. Time-based ordering
5. **What is a major disadvantage of KNN?**
   A. Computationally expensive for large datasets (needs to store all data)
   B. Cannot handle non-linear data
   C. Always overfits
   D. Only works with categorical data
6. **What is the fundamental assumption of Naïve Bayes?**
   A. Features are conditionally independent given the class
   B. All features are perfectly correlated
   C. Features have no relationship with class
   D. Only one feature matters
7. **Which theorem is Naïve Bayes based on?**
   A. Bayes' Theorem
   B. Central Limit Theorem
   C. Pythagorean Theorem
   D. Law of Large Numbers

8. **What does "naïve" refer to in Naïve Bayes?**
   A. The assumption that features are independent (which is often naive)
   B. That it's a simple algorithm
   C. That it always makes naive predictions
   D. That it only works with simple data
9. **Which type of data distribution is commonly assumed in Gaussian Naïve Bayes?**
   A. Normal (Gaussian) distribution
   B. Uniform distribution
   C. Exponential distribution
   D. No distribution assumed
10. **What is a key advantage of Naïve Bayes?**
    A. Fast training and prediction, works well with high-dimensional data
    B. Always achieves highest accuracy
    C. No need for any data
    D. Perfect for all problem types
11. **What is a neural network?**
    A. A composition of simple neurons into several layers
    B. A single mathematical formula
    C. Only input and output without hidden layers
    D. A database system
12. **What does each neuron compute?**
    A. Linear combination of inputs plus bias, passed through activation function
    B. Only the average of inputs
    C. Random function of inputs
    D. Direct copy of input
13. **What is the purpose of hidden layers in neural networks?**
    A. To learn new representations/features of the data
    B. To store training data
    C. To make the network slower
    D. To increase input size
14. **What is the final output of a neural network?**
    A. The target variable y = f(x)
    B. The input data
    C. Random noise
    D. Intermediate features only
15. **What are neurons in neural networks?**
    A. Basic units that receive inputs, governed by threshold and activation function

B. Only output devices
C. Data storage units
D. Connection points only

16. **What do connections between neurons carry?**
    A. Information, regulated by weights and biases
    B. Only electricity
    C. Random signals
    D. No information
17. **What is the role of weights and biases?**
    A. Determine strength and influence of connections
    B. Store training data
    C. Measure accuracy
    D. Count layers
18. **What are propagation functions?**
    A. Mechanisms that process and transfer data across layers
    B. Learning algorithms
    C. Activation functions only
    D. Error calculations
19. **What is the learning rule?**
    A. Method that adjusts weights and biases to improve accuracy
    B. Fixed mathematical formula
    C. Data preprocessing step
    D. Output calculation
20. **What activation function is used for multi-class classification output layer?**
    A. Softmax
    B. Sigmoid
    C. ReLU
    D. Linear
21. **What activation function is used for binary classification output layer?**
    A. Sigmoid
    B. Softmax
    C. ReLU
    D. Tanh
22. **What activation function is used for regression output layer?**
    A. Linear
    B. Sigmoid
    C. Softmax
    D. ReLU
23. **What is the purpose of the loss function?**
    A. Evaluates error by comparing predicted output with actual values

B. Activates neurons
C. Initializes weights
D. Counts layers

24. **What happens during forward propagation?**
    A. Network produces output based on current weights and biases
    B. Weights are updated
    C. Error is calculated backward
    D. Data is collected
25. **What is the general neuron equation?**
    A. Output = activation_function(Σ(weight × input) + bias)
    B. Output = Σ(input)
    C. Output = weight × input
    D. Output = activation_function(input)
26. **What do activation functions decide?**
    A. Whether a neuron should fire/activate
    B. Only the input values
    C. Only the learning rate
    D. The data source
27. **Why do we need activation functions?**
    A. To introduce non-linearity, allowing network to learn complex patterns
    B. To make network linear
    C. To reduce number of neurons
    D. To increase training speed
28. **What happens if all layers use linear activation functions?**
    A. Whole network becomes just a linear model
    B. Network becomes more powerful
    C. Network cannot train at all
    D. Network becomes non-linear
29. **Why can't linear functions solve problems like image recognition?**
    A. Real-world data is not linear
    B. Linear functions are too complex
    C. Linear functions require too much data
    D. Linear functions are always wrong
30. **What is the range of the Sigmoid function?**
    A. (0, 1)
    B. (-1, 1)
    C. (-∞, ∞)
    D. (0, ∞)
31. **What is a problem with Sigmoid activation?**
    A. Vanishing gradient problem

B. Exploding gradients
C. Linear outputs
D. Too simple

32. **What is the range of the Tanh function?**
    A. (-1, 1)
    B. (0, 1)
    C. (-∞, ∞)
    D. (0, ∞)
33. **Why is Tanh sometimes preferred over Sigmoid in hidden layers?**
    A. Zero-centered output facilitates easier learning
    B. Simpler to compute
    C. Always gives better accuracy
    D. No vanishing gradient
34. **What is the ReLU function formula?**
    A. f(x) = max(0, x)
    B. f(x) = 1/(1 + e^(-x))
    C. f(x) = (e^x - e^(-x))/(e^x + e^(-x))
    D. f(x) = x
35. **What problem does Leaky ReLU address compared to standard ReLU?**
    A. "Dying ReLU" problem (neurons that never activate)
    B. Vanishing gradient
    C. Exploding gradient
    D. Too much computation
36. **What happens after the forward pass in neural network training?**
    A. Network computes loss (error between predicted and true value)
    B. Network stops training
    C. Weights are randomly changed
    D. Data is collected again
37. **How does backpropagation find how much each weight contributed to error?**
    A. Using gradient calculation (chain rule)
    B. Random guessing
    C. Asking the user
    D. Using fixed rules
38. **How are weights and biases updated in backpropagation?**
    A. In opposite direction of gradient using optimizer
    B. Randomly
    C. Always increased
    D. Only at output layer

39. **What does the learning rate control?**
    A. How big each weight update step is
    B. Number of layers
    C. Activation function type
    D. Input data size
40. **What is the ultimate goal of backpropagation?**
    A. Continuously reduce loss and improve predictions
    B. Increase training time
    C. Make network more complex
    D. Store more data
41. **What is a perceptron?**
    A. A single neuron receiving multiple inputs, producing one output
    B. Deep neural network
    C. Clustering algorithm
    D. Regression model
42. **What problem demonstrates the limitation of single-layer perceptrons?**
    A. XOR problem
    B. AND problem
    C. OR problem
    D. Linear regression
43. **Why can't a single-layer perceptron solve XOR?**
    A. XOR is not linearly separable
    B. XOR requires too much data
    C. XOR is too simple
    D. XOR requires non-binary inputs
44. **How is the XOR problem solved?**
    A. Using multi-layer perceptron with non-linear activation
    B. Using single-layer perceptron
    C. Using linear regression
    D. Using KNN only
46. **What are the three main parts of an MLP?**
    A. Input layer, hidden layers, output layer
    B. Only input and output
    C. Multiple output layers only
    D. No hidden layers
47. **How does an MLP learn non-linear patterns?**
    A. Thanks to non-linear activation functions
    B. Through linear combinations only
    C. By ignoring non-linearity
    D. Through data preprocessing only

48. **What is the source of non-linearity in MLP?**
    A. Non-linear activation functions
    B. Linear weights
    C. Number of layers only
    D. Input data size
49. **Why are MLPs widely used?**
    A. For classification, regression, and pattern recognition tasks
    B. Only for simple linear problems
    C. Only for data storage
    D. Only for visualization
61. **What does the Softmax function do?**
    A. Converts vector of real numbers into probability distribution
    B. Makes all values negative
    C. Returns binary outputs
    D. Increases vector size
62. **What is the range of Softmax outputs?**
    A. (0, 1) and sum of all elements = 1
    B. (-1, 1)
    C. (-∞, ∞)
    D. (0, ∞)
63. **Why is Softmax used for multi-class classification?**
    A. Output can be interpreted as probabilities of different classes
    B. It's the simplest function
    C. It always gives highest accuracy
    D. It works only with two classes
64. **When would you choose KNN over a neural network?**
    A. When you need simple implementation and interpretability
    B. When you have huge dataset and need fast training
    C. When you need to learn complex non-linear patterns
    D. When you have limited memory
65. **What is a key difference between Naïve Bayes and neural networks?**
    A. Naïve Bayes is probabilistic with independence assumption; neural networks learn hierarchical features
    B. Both are exactly the same
    C. Neural networks are always better
    D. Naïve Bayes is more complex

Functions, and Backpropagation

1. **What is the key question that distinguishes traditional ML from automated procedures?**
   A. Can you use an algorithm to automatically execute the procedure?
   B. Do you have enough data?
   C. Is the problem linear?
   D. Can humans do it manually?
2. **What are neural networks inspired by?**
   A. How the human brain learns
   B. Traditional programming logic
   C. Mathematical formulas only
   D. Database systems
3. **What do neural networks consist of?**
   A. Interconnected layers of nodes (neurons) that process information
   B. Single mathematical equations
   C. Only input and output without processing
   D. Random connections
4. **What does each neuron in a neural network do?**
   A. Receives inputs, applies weights, and passes result through activation function
   B. Only stores data
   C. Only outputs random values
   D. Doesn't process information
5. **How do neural networks learn patterns?**
   A. By adjusting weights during training
   B. By memorizing all data
   C. By asking users questions
   D. By following fixed rules
6. **Where are neural networks widely used?**
   A. Image classification, speech recognition, recommendation systems
   B. Only simple calculations
   C. Only data storage
   D. Only text editing
7. **What is essential for mastering deep learning?**
   A. Understanding forward propagation and backpropagation
   B. Only knowing activation functions
   C. Only collecting data
   D. Only using pre-trained models

8. **What are neurons in neural networks?**
   A. Basic processing units that receive inputs and produce outputs
   B. Only data storage units
   C. Connection points without processing
   D. Output devices only
9. **What do connections between neurons do?**
   A. Pass information, controlled by weights
   B. Store data permanently
   C. Create noise
   D. Don't affect information flow
10. **What do weights and biases determine?**
    A. How strongly inputs affect a neuron
    B. Only the learning speed
    C. Only the data source
    D. Only the output format
11. **What are propagation functions?**
    A. The process of moving information forward through the network
    B. Only weight updates
    C. Only error calculations
    D. Data collection methods
12. **What is the learning rule?**
    A. How the network updates weights using error feedback
    B. Fixed mathematical formula
    C. Data preprocessing step
    D. User input method
13. **What does the output layer generate?**
    A. The final prediction
    B. Intermediate features
    C. Input data
    D. Random noise
14. **What activation function is used for multi-class classification output?**
    A. Softmax
    B. Sigmoid
    C. Linear
    D. ReLU
15. **What activation function is used for binary classification output?**
    A. Sigmoid
    B. Softmax
    C. Linear
    D. Tanh

16. **What activation function is used for regression output?**
    A. Linear
    B. Sigmoid
    C. Softmax
    D. ReLU
17. **What does the network produce during prediction?**
    A. Output based on current weights and biases
    B. Random guesses
    C. Always same output
    D. User-defined values
18. **What evaluates the error in predictions?**
    A. The loss function
    B. The activation function
    C. The input data
    D. The learning rate
19. **What characterizes supervised learning?**
    A. Learns from labeled data, compares prediction with true label
    B. Learns without any guidance
    C. Learns from rewards only
    D. Doesn't use data
20. **What does supervised learning adjust to reduce error?**
    A. Weights
    B. Input data
    C. Output labels
    D. Learning type
21. **What characterizes unsupervised learning?**
    A. Learns from unlabeled data, finds patterns or clusters
    B. Needs correct answers
    C. Uses reward signals
    D. Only works with images
22. **What is missing in unsupervised learning?**
    A. Teacher or correct output
    B. Data
    C. Algorithms
    D. Computation power
23. **What characterizes reinforcement learning?**
    A. Learns by interacting with environment, gets rewards or penalties
    B. Needs labeled examples
    C. Finds clusters in data
    D. Only works with text

24. **What does reinforcement learning try to maximize?**
    A. Reward
    B. Training speed
    C. Data amount
    D. Model complexity
25. **What are the main training steps in neural networks?**
    A. Forward pass → loss → backward pass → weight update
    B. Only forward pass
    C. Random updates
    D. Data collection only
26. **What happens in the forward pass?**
    A. Information flows from input to output, making predictions
    B. Weights are updated
    C. Errors are calculated backward
    D. Data is collected
27. **What happens after forward pass?**
    A. Loss is calculated
    B. Training stops
    C. New data is loaded
    D. Network structure changes
28. **What happens during backward pass?**
    A. Error is propagated back to update weights
    B. New predictions are made
    C. Data is preprocessed
    D. Network expands
29. **What is the final step in the training loop?**
    A. Weight update
    B. Data collection
    C. Network design
    D. User evaluation
30. **What is the general neuron equation?**
    A. Output = activation_function(Σ(weight × input) + bias)
    B. Output = Σ(input)
    C. Output = weight × input
    D. Output = activation_function(input)
31. **What do activation functions decide?**
    A. Whether a neuron should fire
    B. Input data values
    C. Learning rate
    D. Number of layers

32. **What do activation functions introduce?**
    A. Non-linearity
    B. Linearity only
    C. Randomness
    D. Data storage
33. **What happens without activation functions?**
    A. Network becomes just a linear model
    B. Network becomes more powerful
    C. Network stops working
    D. Network becomes faster
34. **What do activation functions help the network understand?**
    A. Curves, edges, images, language, non-linear boundaries
    B. Only linear relationships
    C. Only simple patterns
    D. Only mathematical formulas
35. **Why are activation functions needed?**
    A. Real-world data is not linear
    B. To make networks slower
    C. To reduce data requirements
    D. To simplify programming
36. **What happens if all layers use linear activation?**
    A. Whole network becomes just a linear model
    B. Network becomes more powerful
    C. Network cannot learn at all
    D. Network becomes non-linear
37. **What can't a network with only linear activation learn?**
    A. Complex or non-linear patterns
    B. Simple patterns
    C. Linear relationships
    D. Any patterns
38. **What problem is associated with Sigmoid activation?**
    A. Vanishing gradient problem
    B. Exploding gradient
    C. Linear outputs
    D. Too fast learning
41. **What is Tanh function?**
    A. Hyperbolic tangent function, shifted version of sigmoid
    B. Linear function
    C. Step function
    D. Constant function

42. **What is the range of Tanh outputs?**
    A. -1 to +1
    B. 0 to 1
    C. -∞ to +∞
    D. 0 to ∞
43. **Why is Tanh non-linear?**
    A. Enables modeling of complex data patterns
    B. Makes everything linear
    C. Simplifies calculations
    D. Reduces accuracy
44. **Where is Tanh commonly used?**
    A. Hidden layers
    B. Output layer only
    C. Input layer only
    D. Never used
45. **Why is Tanh preferred in hidden layers?**
    A. Zero-centered output facilitates easier learning
    B. Always gives best accuracy
    C. Simplest to compute
    D. No gradient problems
46. **What happens around 0 in Tanh function?**
    A. Curve is steep, function very sensitive to input changes
    B. Function is flat
    C. Function is discontinuous
    D. Function is random
47. **What is ReLU function?**
    A. Rectified Linear Unit: max(0, x)
    B. Sigmoid function
    C. Linear function
    D. Tanh function
48. **What is Leaky ReLU?**
    A. Enhanced ReLU with non-zero slope for negative values
    B. Same as standard ReLU
    C. Linear function
    D. Sigmoid variant
49. **What is Exponential Linear Unit (ELU)?**
    A. Another non-linear activation with smooth curve
    B. Linear activation
    C. Step function
    D. Constant function

50. **What does Softmax function transform input values into?**
    A. Probabilities
    B. Binary outputs
    C. Random numbers
    D. Input copies
51. **What happens to very small (negative) input values in Softmax?**
    A. Output close to zero
    B. Output becomes 1
    C. Output becomes random
    D. No change
52. **What happens as input value becomes larger in Softmax?**
    A. Output increases sharply
    B. Output decreases
    C. Output becomes zero
    D. Output becomes negative
53. **What does Softmax demonstrate about bigger scores?**
    A. Become bigger probabilities
    B. Become smaller probabilities
    C. Become zero
    D. Become negative
54. **What does Softmax demonstrate about smaller scores?**
    A. Become very small probabilities
    B. Become large probabilities
    C. Become 1
    D. Become random
55. **What helps the model choose the most likely class in multi-class classification?**
    A. Softmax giving higher probability to largest input value
    B. Random selection
    C. Always choosing first class
    D. User input
56. **What do different activation functions affect?**
    A. Model performance and learning characteristics
    B. Only training speed
    C. Only model size
    D. Only input data
57. **What does the network compute after forward pass?**
    A. Loss (error between predicted and true value)
    B. New input data

C. Network structure
D. Learning rate

58. **How does backpropagation find weight contributions to error?**
    A. Using gradient calculation
    B. Random guessing
    C. Asking user
    D. Fixed formulas
59. **How are weights and biases updated?**
    A. In opposite direction of gradient using optimizer
    B. Randomly
    C. Always increased
    D. Only at output layer
60. **What does learning rate control?**
    A. How big each update step is
    B. Number of layers
    C. Activation function type
    D. Data size
61. **What does backpropagation continuously do?**
    A. Reduces loss and improves predictions
    B. Increases error
    C. Changes network structure
    D. Collects more data
62. **How is error calculated?**
    A. Error = predicted output - actual output
    B. Random number
    C. Always zero
    D. Input value
63. **What is a common way to measure error?**
    A. Mean Squared Error (MSE)
    B. Random selection
    C. User judgment
    D. Fixed value
64. **How is error sent through the network?**
    A. Backward through network layer by layer
    B. Forward only
    C. Random direction
    D. Not sent at all
65. **What are computed to see how each weight affected error?**
    A. Gradients
    B. Inputs

C. Outputs
D. Layers

66. **In what direction are weights updated?**
    A. Direction that reduces error
    B. Random direction
    C. Always increase
    D. Always decrease
67. **What is used to calculate gradients?**
    A. Derivative of activation function
    B. Input data only
    C. Output data only
    D. Random numbers
68. **What does forward propagation compute?**
    A. Predictions using current weights
    B. New weights
    C. Error values
    D. Learning rates
69. **What function is applied at each neuron during forward propagation?**
    A. Activation function
    B. Loss function
    C. Gradient function
    D. Data function
70. **How are individual neuron outputs computed?**
    A. Weighted sum of inputs plus bias, passed through activation
    B. Random values
    C. Copy of input
    D. Fixed values
71. **How is final output computed?**
    A. From last layer's activations
    B. From first layer only
    C. Random selection
    D. User input
72. **What is compared to calculate error?**
    A. Predicted output vs true output
    B. Input vs output
    C. Weights vs biases
    D. Layers vs neurons
73. **What are gradients used for in weight updates?**
    A. Determine how much to adjust each weight
    B. Change input data

C. Modify network structure
D. Select activation functions

74. **What is hidden unit error?**
A. Error propagated back to hidden layers
B. Error at input layer
C. Error at output only
D. Random error
75. **How does backpropagation update each weight?**
A. Multiply neuron's error with input, shift weights slightly
B. Random changes
C. Large fixed changes
D. Only at output layer

Lecture 11: Multilayer Perceptron (MLP) and Backpropagation

1. **What does MLP stand for?**
A. Multilayer Perceptron
B. Multiple Linear Programming
C. Multi-Level Processing
D. Machine Learning Protocol
2. **What is an MLP?**
A. One of the simplest and most fundamental neural network architectures
B. A type of decision tree
C. A clustering algorithm
D. A linear regression model
3. **How many layers does an MLP have at minimum?**
A. At least three layers
B. Only one layer
C. Exactly two layers
D. Four or more layers
4. **What are the three required layers in an MLP?**
A. Input layer, Hidden layer(s), Output layer
B. Only input and output layers
C. Multiple output layers
D. Only hidden layers
5. **What does the input layer receive?**
A. The feature vector (e.g., x_1, x_2, x_3...)
B. The output predictions

C. The error values
D. The weight updates

6. **What do hidden layer neurons apply?**
   A. Weighted sum + activation function
   B. Only linear transformations
   C. Random operations
   D. Only data storage
7. **What does the output layer produce?**
   A. Final prediction (class, number, probability)
   B. Intermediate features
   C. Input data
   D. Error calculations
8. **How does information flow in an MLP?**
   A. One way: Input → Hidden Layers → Output
   B. Circular flow
   C. Backward only
   D. Random flow
9. **What is the "feedforward" characteristic?**
   A. Information flows forward without cycles
   B. Information flows backward
   C. Information flows in both directions
   D. No information flow
10. **When is a single layer perceptron insufficient?**
    A. When data requires non-linear separation
    B. Always insufficient
    C. For simple linear problems
    D. For any classification task
11. **What do multilayer perceptrons introduce?**
    A. Hidden layers and non-linearity
    B. Only more input features
    C. Only linear functions
    D. Simpler models
12. **What does the multilayer version enable?**
    A. Approximation of complex functions
    B. Only linear approximations
    C. Only data storage
    D. Only simple calculations
13. **What problem demonstrates the need for MLP?**
    A. XOR problem
    B. AND problem

C. OR problem
D. Simple addition

14. **What is XOR?**
    A. Exclusive OR - outputs 1 only when inputs are different
    B. Always outputs 1
    C. Outputs 1 when inputs are same
    D. Simple addition function
15. **What does XOR output when inputs are same?**
    A. 0
    B. 1
    C. Depends on values
    D. Random output
16. **What can't separate XOR classes?**
    A. Single straight line
    B. Multiple lines
    C. Curved lines
    D. Any separator
17. **What is needed to classify XOR points correctly?**
    A. MLP with non-linearity
    B. Single layer perceptron
    C. Linear regression
    D. Simple threshold
18. **What happens during forward propagation?**
    A. Information flows forward to make prediction
    B. Weights are updated
    C. Error is calculated backward
    D. Network structure changes
19. **What is the result of forward propagation?**
    A. Network output/prediction
    B. Weight updates
    C. Error values
    D. New input data
20. **What does a loss function measure?**
    A. How far the model's prediction is from true label
    B. Network complexity
    C. Training speed
    D. Number of layers
21. **What is backpropagation's relationship to loss function?**
    A. Uses loss to update weights
    B. Ignores loss

C. Increases loss
D. Creates loss function

22. **Which loss function is most common for classification?**
    A. Cross-Entropy Loss
    B. Mean Squared Error
    C. Absolute Error
    D. Hinge Loss
23. **What does Cross-Entropy Loss do?**
    A. Penalizes wrong predictions strongly
    B. Only measures small errors
    C. Always gives zero loss
    D. Only for regression
24. **What works well with Cross-Entropy?**
    A. Softmax (multi-class) or Sigmoid (binary)
    B. Linear activation
    C. ReLU only
    D. No activation
25. **Which loss function is good for regression?**
    A. Mean Squared Error (MSE)
    B. Cross-Entropy
    C. Hinge Loss
    D. Log Loss
26. **What characterizes MSE?**
    A. Simple and intuitive
    B. Complex and obscure
    C. Only for classification
    D. Always perfect
27. **What happens after forward pass?**
    A. Output compared to true value to calculate error
    B. Training stops
    C. Network expands
    D. Data is collected
28. **How is error propagated?**
    A. Backward through network (backpropagation)
    B. Forward only
    C. Not propagated
    D. Random direction
29. **What is adjusted during backpropagation?**
    A. Weights and biases to minimize error
    B. Input data

C. Output labels
D. Network structure

30. **What algorithms are typically used for optimization?**
    A. Gradient descent (SGD) and variants
    B. Random guessing
    C. Fixed rules
    D. User input
31. **What is the relationship between forward and backpropagation?**
    A. Forward makes prediction, backpropagation learns from mistake
    B. Both make predictions
    C. Both update weights
    D. No relationship
32. **What does the network compute after forward pass?**
    A. Loss (error between predicted and true)
    B. New input
    C. Network size
    D. Learning rate
33. **How does it find weight contributions to error?**
    A. Using gradient calculation
    B. Random assignment
    C. Asking user
    D. Fixed formulas
34. **How are weights and biases updated?**
    A. Opposite direction of gradient using optimizer
    B. Random direction
    C. Always increase
    D. Only at output
35. **What controls update step size?**
    A. Learning rate
    B. Number of layers
    C. Input size
    D. Output value
36. **What does backpropagation continuously do?**
    A. Reduces loss and improves predictions
    B. Increases error
    C. Changes network structure
    D. Collects data
37. **What is a common way to measure error?**
    A. Mean Squared Error (MSE)
    B. Random number

C. Fixed value
D. User judgment

38. **How is error sent through network?**
A. Backward layer by layer
B. Forward only
C. Randomly
D. Not sent

39. **What are computed to see weight effects on error?**
A. Gradients
B. Inputs
C. Outputs
D. Layers

40. **In what direction are weights updated?**
A. Direction that reduces error
B. Random direction
C. Always increase
D. Always decrease

41. **What is used to calculate gradients?**
A. Derivative of activation function
B. Input data only
C. Output data only
D. Random numbers

42. **What does forward propagation compute?**
A. Predictions using current weights
B. New weights
C. Error values
D. Learning rates

43. **What function is applied at each neuron?**
A. Activation function
B. Loss function
C. Gradient function
D. Data function

44. **Which activation function might be used in hidden layers?**
A. ReLU, Sigmoid, or Tanh
B. Only linear
C. Only Softmax
D. No activation

45. **What does activation function determine?**
A. Whether neuron fires and non-linearity
B. Only input values

C. Only learning rate
D. Only data source

46. **How are individual neuron outputs computed?**
    A. Weighted sum of inputs + bias, through activation
    B. Random values
    C. Copy of input
    D. Fixed values
47. **How is final output computed?**
    A. From last layer's activations
    B. From first layer only
    C. Random selection
    D. User input
48. **What is compared to calculate error?**
    A. Predicted output vs true output
    B. Input vs output
    C. Weights vs biases
    D. Layers vs neurons
49. **What is typical error calculation for regression?**
    A. MSE = (predicted - true)^2
    B. Random formula
    C. Always zero
    D. From input data
50. **What are gradients used for?**
    A. Determine how much to adjust each weight
    B. Change input data
    C. Modify network structure
    D. Select activation functions
51. **What is chain rule used for?**
    A. Computing derivatives through multiple layers
    B. Adding numbers
    C. Multiplying matrices
    D. Initializing weights
52. **What is hidden unit error?**
    A. Error propagated back to hidden layers
    B. Error at input layer
    C. Error at output only
    D. Random error
53. **How is hidden layer error calculated?**
    A. Based on errors from subsequent layers
    B. Random values

C. Fixed percentages
D. From input data

54. **How does backpropagation update each weight?**
A. Multiply neuron's error with input, shift weights slightly
B. Random changes
C. Large fixed changes
D. Only at output layer

55. **What is the weight update formula?**
A. Δw = learning_rate × error × input
B. Δw = random value
C. Δw = fixed value
D. Δw = input ÷ output

56. **What is the purpose of weight updates?**
A. Reduce future errors
B. Increase errors
C. Change network structure
D. Store more data

57. **How are weights "slightly shifted"?**
A. Small changes in direction that reduces error
B. Large random changes
C. Always set to zero
D. Only positive changes

58. **What does forward pass compute?**
A. Predictions using current weights
B. New network structure
C. Data collection
D. User preferences

59. **How is loss calculated?**
A. e.g., MSE = $(y\_pred - y\_true)^2$
B. Random formula
C. Always zero
D. From input data

60. **How is error sent through network?**
A. Backward through network
B. Forward only
C. Random direction
D. Not sent

61. **How are gradients computed?**
A. Using chain rule
B. Random numbers

C. Fixed values
D. User input