

Intro to AI - Sammendrag

1. What is Artificial Intelligence (AI)?

- A) A system for collecting data.
- B) A program with explicit human instructions.
- C) A machine mimicking human intelligence with decision-making capabilities.
- D) A purely physical machine with no learning ability.

Answer: C

2. What does supervised learning require?

- A) Unstructured data.
- B) Labeled data.
- C) Randomized inputs.
- D) Trial-and-error learning.

Answer: B

3. Which of these is NOT a step in designing an AI system?

- A) Identify the problem.
- B) Label the data automatically.
- C) Choose algorithms.
- D) Train the algorithm with data.

Answer: B

4. What is the primary focus of Deep Learning (DL)?

- A) Data encryption.
- B) Generative algorithms.
- C) High-dimensional data processing.
- D) Low-complexity problems.

Answer: C

5. What does clustering in Machine Learning aim to do?

- A) Predict future trends.
- B) Classify unknown labels.
- C) Group similar data points.
- D) Anonymize sensitive data.

Answer: C

6. What is Explainable AI (XAI)?

- A) An AI system that self-learns without transparency.
- B) Processes and methods to make AI outputs understandable to humans.
- C) A term for autonomous learning models.
- D) AI systems optimized for speed over accuracy.

Answer: B

7. What is a key characteristic of logistic regression?

- A) Works with continuous data.
- B) Outputs probabilities for classification problems.
- C) Requires unsupervised learning data.
- D) Operates without a decision boundary.

Answer: B

8. Which use case is best suited for regression?

- A) Spam detection.
- B) House price prediction.
- C) Gender classification.
- D) Document clustering.

Answer: B

9. What is Naive Bayes primarily used for?

- A) Data clustering.
- B) Text and sentiment classification.
- C) Detecting image edges.
- D) Self-driving car training.

Answer: B

10. What is a hyperplane in Support Vector Machines?

- A) A clustering method.
- B) A feature selection tool.
- C) A decision boundary for classification.
- D) A data preparation technique.

Answer: C

11. Which algorithm is density-based for clustering?

- A) K-Means.
- B) DBSCAN.
- C) Naive Bayes.
- D) Random Forest.

Answer: B

12. What does feature engineering involve?

- A) Predicting data labels.
- B) Removing redundant data points.
- C) Constructing new data features from existing ones.
- D) Automatically clustering data into groups.

Answer: C

13. Which learning model uses reward and penalty mechanisms?

- A) Supervised learning.
- B) Unsupervised learning.
- C) Reinforcement learning.
- D) Deep learning.

Answer: C

14. What is the focus of recommender systems?

- A) Improving data clustering.
- B) Predicting future weather conditions.
- C) Suggesting relevant items to users.
- D) Enhancing supervised learning models.

Answer: C

15. What does "outlier" mean in data processing?

- A) A missing data point.
- B) A data point far from the rest of the dataset.
- C) A common error in labeled data.
- D) A type of feature used for classification.

Answer: B

16. What does synthetic data aim to achieve?

- A) Clean and organize real data.
- B) Replicate the structure of real data artificially.
- C) Replace raw data with encrypted formats.
- D) Extract meaningful features from data.

Answer: B

17. Which clustering method uses "sliding windows"?

- A) K-Means.
- B) Mean Shift Clustering.
- C) Gaussian Mixture Model.
- D) DBSCAN.

Answer: B

18. What is a major challenge in reinforcement learning?

- A) Lack of labeled data.
- B) Preparing the simulation environment.
- C) Processing small datasets.
- D) Training without external input.

Answer: B

19. What type of learning is used for stock market price prediction?

- A) Reinforcement learning.
- B) Regression.
- C) Clustering.
- D) Classification.

Answer: D

20. Which of the following is NOT a supervised learning algorithm?

- A) Decision Trees.
- B) K-Means.
- C) Random Forest.
- D) Logistic Regression.

Answer: B

21. What is the function of the sigmoid activation in logistic regression?

- A) Filter outliers in data.
- B) Create continuous output.
- C) Map predictions to probabilities.
- D) Optimize feature engineering.

Answer: C

22. What does the Apriori algorithm analyze?

- A) Frequent itemsets for association rules.
- B) Text for sentiment classification.
- C) Image data for clustering.
- D) Data for hyperparameter tuning.

Answer: A

23. What is a "centroid" in K-Means clustering?

- A) A decision boundary.
- B) The center of a cluster.
- C) A classification algorithm.
- D) A feature extraction tool.

Answer: B

24. What is the primary use of Gaussian Mixture Models?

- A) Analyze density-based patterns.
- B) Generalize linear models.
- C) Handle irregular clustering formats.
- D) Replace supervised learning.

Answer: C

25. What is the Turing Test designed to evaluate?

- A) The accuracy of supervised learning.
- B) The ethical concerns of AI applications.
- C) If a computer can think like a human.
- D) The robustness of reinforcement learning.

Answer: C

26. What is Artificial Intelligence (AI)?

- a) A set of rules programmed by humans
- b) Machines mimicking human intelligence
- c) Software that follows a fixed path
- d) A hardware component in robots

Answer: b

27. Which of the following is NOT a type of supervised learning?

- a) Regression
- b) Clustering
- c) Classification
- d) Decision Trees

Answer: b

28. What is the primary goal of unsupervised learning?

- a) Predict labels based on input
- b) Create clusters from unlabeled data
- c) Generate large datasets
- d) Improve supervised models

Answer: b

29. The Turing Test is used to:

- a) Test the speed of a computer
- b) Measure AI's ability to mimic human behavior
- c) Compare algorithms
- d) Evaluate software development

Answer: b

30. What does "GOF AI" stand for?

- a) Generalized Object Features in AI
- b) Goals of Future Artificial Intelligence
- c) Good Old-Fashioned Artificial Intelligence
- d) Generated Output for AI

Answer: c

31. Which is a type of clustering algorithm?

- a) K-Means
- b) Linear Regression
- c) Random Forest
- d) Naive Bayes

Answer: a

32. What is a significant limitation of Deep Learning?

- a) It cannot process images.
- b) It struggles with edge cases.
- c) It only works with labeled data.
- d) It is not scalable.

Answer: b

33. What does "Explainable AI" aim to achieve?

- a) Higher efficiency in algorithms
- b) Better prediction accuracy
- c) Transparency in AI outputs
- d) Increased training speed

Answer: c

34. In supervised learning, what is "regression" used for?

- a) Grouping similar data
- b) Predicting continuous values
- c) Generating random outcomes
- d) Identifying categorical labels

Answer: b

35. Which algorithm is commonly used for spam detection?

- a) K-Means Clustering
- b) Naive Bayes
- c) DBSCAN
- d) Decision Trees

Answer: b

36. What does "Gradient Descent" optimize in machine learning?

- a) The data collection process
- b) The accuracy of predictions
- c) The cost function
- d) The feature engineering process

Answer: c

37. Which of the following is a type of logistic regression?

- a) Binary
- b) Decision Trees
- c) Polynomial
- d) Ridge

Answer: a

38. What is synthetic data?

- a) Data extracted from raw datasets
- b) Artificially generated data mimicking real data
- c) Data collected directly from users
- d) Encrypted data for security purposes

Answer: b

39. What does DBSCAN stand for?

- a) Density-Based Statistical Classification Algorithm Network
- b) Distance-Based Spatial Clustering of Applications with Noise
- c) Data-Based Systems Clustering and Analysis Network
- d) Density-Based Spatial Clustering of Applications with Noise

Answer: d

40. Which of the following is NOT an example of AI in daily life?

- a) Voice assistants like Alexa
- b) Self-driving cars
- c) Regular spreadsheet calculations
- d) Spam email filters

Answer: c

41. What is the main characteristic of unsupervised learning?

- a) Training on labeled data
- b) Creating relationships between inputs and outputs
- c) Discovering patterns in unlabeled data
- d) Predicting specific outcomes

Answer: c

42. Which AI technique is used to group similar kinds of data?

- a) Classification
- b) Clustering
- c) Regression
- d) Reinforcement learning

Answer: b

43. The term "AI bias" refers to:

- a) Algorithms running slower than expected
- b) Errors due to incorrect labels
- c) Systematic unfairness in AI decisions
- d) Training data being incomplete

Answer: c

44. What is the purpose of feature engineering?

- a) Build hardware for AI models
- b) Clean and process raw data for better analysis
- c) Test the efficiency of neural networks
- d) Reduce the dimensionality of datasets

Answer: b

45. What is the main function of reinforcement learning?

- a) Labeling datasets automatically
- b) Learning through rewards and penalties
- c) Predicting continuous outcomes
- d) Identifying data outliers

Answer: b

46. What is the significance of "clustering" in AI?

- a) Creating patterns for labeled data
- b) Automating model evaluation
- c) Grouping similar data points together
- d) Optimizing feature extraction

Answer: c

47. Which of the following is an example of Explainable AI?

- a) An algorithm with no observable outputs
- b) A black-box model
- c) A transparent decision-making process
- d) A model trained on incomplete data

Answer: c

48. Which method is often used for customer segmentation?

- a) DBSCAN
- b) Naive Bayes
- c) Logistic Regression
- d) Reinforcement Learning

Answer: a

49. What is "capacity prediction" in AI?

- a) Predicting storage requirements for AI systems
- b) Estimating future hardware costs
- c) Predicting when a system's workload will exceed capacity
- d) Determining the maximum data size a model can process

Answer: c

50. Which of the following is used for spam detection?

- a) Polynomial Regression
- b) K-Means Clustering
- c) Naive Bayes Classifier
- d) Gaussian Mixture Model

Answer: c

51. Which AI algorithm uses hyperplanes?

- a) Support Vector Machines
- b) Logistic Regression
- c) K-Nearest Neighbor
- d) Random Forest

Answer: a

52. What does the term "overfitting" mean in machine learning?

- a) The model fails to capture patterns in training data
- b) The model works perfectly on new data
- c) The model performs well on training data but poorly on unseen data
- d) The model is too simple for the given dataset

Answer: c

53. Which AI system passed the Turing Test in 2018?

- a) AlphaGo
- b) Google Duplex
- c) IBM Watson
- d) OpenAI GPT-3

Answer: b

54. What is meant by "data anonymization"?

- a) Encrypting sensitive data to ensure privacy
- b) Generating synthetic data
- c) Cleaning data by removing duplicate entries
- d) Preparing data for AI training

Answer: a

55. What is the main challenge of reinforcement learning?

- a) Training with labeled data
- b) Preparing the simulation environment
- c) Optimizing cost functions
- d) Processing large datasets quickly

Answer: b

56. What does "feature extraction" focus on?

- a) Cleaning noisy datasets
- b) Identifying relevant features from raw data
- c) Aggregating multiple datasets
- d) Reducing model size

Answer: b

57. What is a common use case for Random Forest?

- a) Image processing
- b) Time series analysis
- c) Fraud detection
- d) Speech recognition

Answer: c

58. What does "association" in machine learning aim to do?

- a) Group data points into clusters
- b) Predict numerical outcomes
- c) Find relationships between items in datasets
- d) Optimize training algorithms

Answer: c

59. What is an "outlier" in a dataset?

- a) A data point that differs significantly from others
- b) A missing value in a dataset
- c) A duplicate entry in a database
- d) A mislabeled data instance

Answer: a

60. What is "Mean Shift Clustering" primarily used for?

- a) Reducing noise in datasets
- b) Finding dense regions in data
- c) Detecting fraud in transactions
- d) Analyzing decision boundaries

Answer: b

Theory Questions

1. What is the primary difference between supervised and unsupervised learning?

Answer:

Supervised learning requires labeled data to train the model, where the algorithm learns from examples to predict outcomes.

Unsupervised learning works with unlabeled data and focuses on discovering patterns, such as clustering or relationships within the data.

2. How does Explainable AI (XAI) enhance trust in AI systems?

Answer:

Explainable AI provides processes and methods that allow human users to understand and trust the outputs of machine learning models. It clarifies how decisions are made, making AI more transparent and reliable.

3. Why is feature engineering considered an essential part of machine learning?

Answer:

Feature engineering involves extracting relevant information from raw data to create meaningful features that improve the performance of machine learning models. It combines domain knowledge with statistical methods to enhance predictive accuracy.

4. What challenges are associated with reinforcement learning?

Answer:

Challenges in reinforcement learning include creating realistic simulation environments for training, scaling and tweaking the neural networks, and transferring models from training environments to real-world scenarios without errors.

5. What is the significance of the sigmoid function in logistic regression?

Answer:

The sigmoid function maps predicted values to probabilities between 0 and 1, allowing the logistic regression model to classify data based on a threshold, such as 0.5, for binary outcomes.

6. How does the DBSCAN algorithm detect outliers in clustering?

Answer:

DBSCAN identifies outliers by separating regions of low-density data from high-density clusters. It uses parameters like minPts (minimum points in a dense region) and eps (distance metric) to define cluster boundaries.

7. What is the role of synthetic data in AI and machine learning?

Answer:

Synthetic data is artificially generated to mimic real data while preserving its characteristics. It is useful for training models when real data is scarce, sensitive, or requires anonymization.

8. Why is gradient descent important in machine learning?

Answer:

Gradient descent is an optimization algorithm used to minimize the cost function by iteratively adjusting parameters in the direction

that reduces prediction errors, enabling the model to improve its accuracy.

9. What makes K-Means clustering an efficient algorithm for unsupervised learning?

Answer:

K-Means is efficient because it is simple to run and relies on distance-based calculations to group data points into clusters. It iteratively updates centroids until clusters stabilize or the maximum number of iterations is reached.

10. How does Naive Bayes handle feature independence in its classification?

Answer:

Naive Bayes assumes that the presence of a particular feature is independent of other features in a dataset. This simplification allows it to calculate probabilities efficiently, making it a powerful and easy-to-implement classification algorithm.

11. Discuss the main differences between classification and regression in machine learning.

Answer:

- **Classification:** Predicts discrete labels or categories (e.g., spam or not spam).
- **Regression:** Predicts continuous values (e.g., house prices, salary).
- **Key Difference:** Classification works with discrete output, while regression deals with continuous numerical output.

12. What are the advantages and disadvantages of using Support Vector Machines?

Answer:

- **Advantages:**
 - Effective for high-dimensional spaces.
 - Works well with smaller datasets.
 - Finds complex relationships between data.
- **Disadvantages:**
 - Computationally intensive for large datasets.
 - Does not provide probability estimates without additional computation.
 - Sensitive to the choice of kernel and parameters.

13. Explain the steps involved in logistic regression.

Answer:

1. Visualize data to understand relationships.
2. Apply the sigmoid activation function to map values to probabilities.
3. Determine a decision boundary for classification.
4. Train the model using gradient descent to minimize the cost function.
5. Map probabilities to class labels (e.g., 0 or 1).
6. Evaluate model accuracy through iterations.

14. Describe the challenges faced by self-driving cars in real-world scenarios.

Answer:

- Handling edge cases, like unusual objects or unexpected events (e.g., animals on the road).
- Adapting to diverse environmental conditions (e.g., snow obscuring lane markings).
- Avoiding over-reliance on biased or incomplete training data.
- Ensuring safety and reliability in high-stakes situations.

15. What are the ethical implications of bias in AI training datasets?

Answer:

- **Bias in data** can perpetuate discrimination based on race, gender, or socioeconomic status.
- AI systems trained on biased data may make unfair decisions, impacting hiring, healthcare, or law enforcement.
- Transparency and fairness are essential to address and mitigate such biases.

16. How does the K-Means algorithm determine the optimal number of clusters?

Answer:

The optimal number of clusters can be determined using the **elbow method**, where the total within-cluster sum of squares (WSS) is plotted against the number of clusters, and the "elbow point" indicates the optimal value.

17. What are hyperplanes, and how are they utilized in SVM?

Answer:

A hyperplane is a decision boundary that separates data points in an SVM model. In higher dimensions, hyperplanes maximize the margin between data points from different classes, ensuring better generalization.

18. Define Explainable AI and its significance in building trust with AI models.

Answer:

Explainable AI (XAI) refers to methods that make AI outputs understandable to humans. It helps build trust by clarifying how models reach decisions, particularly in critical applications like healthcare or finance.

19. What is the role of feature engineering in improving machine learning performance?

Answer:

Feature engineering creates meaningful features from raw data, enhancing model accuracy by making patterns more recognizable. It improves prediction accuracy and model interpretability.

20. Explain the use and significance of the Apriori algorithm in association analysis.

Answer:

The Apriori algorithm identifies frequent itemsets in transactional data to discover association rules. It powers recommendation systems (e.g., "customers who bought X also bought Y").

21. Explain the difference between supervised and unsupervised learning.

Answer:

- **Supervised Learning:** Requires labeled data to learn and predict outcomes.
- **Unsupervised Learning:** Works with unlabeled data to find patterns and relationships (e.g., clustering).

22. Describe the concept of "data pitfalls" in AI. Provide examples.

Answer:

Data pitfalls refer to issues that compromise model performance, including:

- **Assuming clean data:** Ignoring typos or outliers.
- **Ignoring seasonality:** Failing to account for time-based trends (e.g., Black Friday sales).
- **Bias in data:** Training on imbalanced datasets.

23. What are the steps involved in designing an AI system?

Answer:

1. Identify the problem.
2. Prepare and clean the data.
3. Choose the appropriate algorithms.
4. Train the model using data.
5. Deploy and monitor the model on the selected platform.

24. Discuss the limitations of AI in current real-world applications.

Answer:

- AI struggles with edge cases and generalizing from limited data.
- High computational and data requirements.
- Ethical issues such as bias and lack of transparency.
- Adoption is limited to large organizations due to costs and expertise.

25. What are hyperplanes in Support Vector Machines (SVM), and how are they used?

Answer:

Hyperplanes are decision boundaries in SVMs that separate different classes. The goal is to maximize the margin between data points of different classes to improve classification accuracy.

26. Explain the role of feature engineering in machine learning. Why is it crucial?

Answer:

Feature engineering transforms raw data into meaningful inputs for models, enhancing accuracy by making patterns more identifiable. It is crucial for creating relevant features and improving model performance.

27. How does the K-Means clustering algorithm work? Outline its advantages.

Answer:

- **Steps:**
 1. Initialize centroids randomly.
 2. Assign data points to the nearest centroid.
 3. Update centroids as the mean of assigned points.
 4. Repeat until centroids stabilize.
- **Advantages:**
 - Easy to implement.
 - Efficient for large datasets.

28. Define "Generative Algorithms" and provide an example of their application.

Answer:

Generative algorithms learn to model the data distribution to create new data similar to the training set.

Example: Generative Adversarial Networks (GANs) for creating realistic images.

29. What is the significance of the sigmoid function in logistic regression?

Answer:

The sigmoid function maps outputs to a range between 0 and 1, representing probabilities, which are then used to classify data points based on a threshold.

30. Discuss the ethical challenges associated with data bias in AI.

Answer:

Data bias can lead to discriminatory practices and reinforce societal inequalities. AI trained on biased datasets may produce unfair or harmful outcomes, requiring stringent measures to ensure fairness and accountability.