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

Answer:

- **Supervised Learning:** Uses labeled data to train predictive models.
    - Regression → Predicts continuous values (e.g., price, temperature).
    - Classification → Predicts categories (e.g., spam/not spam).
- **Unsupervised Learning:** Uses unlabeled data to discover patterns.
    - Clustering → Groups similar data (e.g., customer segmentation).
    - Dimensionality Reduction → Reduces features (e.g., PCA).
      Summary: Supervised = predict outcomes; Unsupervised = find hidden structure.

---

2. What is the difference between classification and regression?

Answer:

- **Classification:** Predicts discrete labels (e.g., spam or not).
    - Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC.
- **Regression:** Predicts continuous values (e.g., house price).
    - Metrics: MAE, MSE, R^2.
      Summary: Classification → categories; Regression → continuous values.

---

3. What is the bias–variance tradeoff?

Answer:

- **Bias:** Error from oversimplification → underfitting.

- **Variance:** Error from sensitivity to data → overfitting.
  Goal: Minimize both for optimal performance.

---

4. How to deal with overfitting and underfitting?

Answer:

- **Overfitting Fixes:** Regularization, cross-validation, early stopping, reduce complexity, more data.
- **Underfitting Fixes:** Increase complexity, add features, reduce regularization.

---

5. What is cross-validation and why is it important?

Answer:

- **Definition:** Splits data into k folds to train and validate multiple times.
- **Purpose:** Ensures model generalization, reduces overfitting, helps tuning.

---

6. What are precision, recall, and F1-score?

Answer:

- **Precision:** TP / (TP + FP) → Correct positives among predicted.
- **Recall:** TP / (TP + FN) → Correct positives among actual.
- **F1-score:** Harmonic mean of precision and recall.
  Trade-off: High precision lowers recall and vice versa.

---

7. How to choose the right evaluation metric?

Answer:

- **Classification:** Accuracy, Precision, Recall, F1, AUC-ROC.
- **Regression:** MAE, MSE, RMSE, R^2.
- **Imbalanced Data:** Prefer Precision, Recall, or F1.
  Example: For fraud detection → use Recall or AUC-ROC.

---

8. What is the difference between accuracy, precision, and recall?

Answer:

- **Accuracy:** (TP + TN) / Total predictions.
- **Precision:** TP / (TP + FP).
- **Recall:** TP / (TP + FN).
  When to Use:
- Accuracy → balanced data.
- Precision → false positives costly.
- Recall → false negatives costly.

---

9. What is a confusion matrix?

Answer:
A table comparing actual vs predicted outcomes:

- **TP:** Correct positives
- **FP:** Wrong positives
- **TN:** Correct negatives
- **FN:** Missed positives
  Metrics Derived: Accuracy, Precision, Recall, F1-score.

---

10. How to handle missing or corrupted data?

Answer:

- **Delete:** If minimal impact.
- **Impute:** Mean, median, mode, k-NN, or predictive models.
- **Flag:** Add indicator for missingness.
- **Advanced:** Use algorithms like XGBoost that handle missing data.

---

11. How to handle categorical variables?

Answer:

- **One-hot encoding:** For nominal data.
- **Ordinal encoding:** For ordered categories.
- **Target encoding:** For high-cardinality features.

---

12. What is feature engineering and why is it important?

Answer:
Creating or transforming features to improve model performance.
Examples: Scaling, encoding, creating interactions.
Importance: Good features = better accuracy and interpretability.

---

13. What is the difference between parametric and non-parametric models?

Answer:

- **Parametric:** Fixed parameters (e.g., Linear Regression).

    - Fast, simple, less flexible.

- **Non-parametric:** Grows with data (e.g., Decision Trees, k-NN).
    - Flexible, may overfit.

---

14. What is the curse of dimensionality?

Answer:
Too many features cause sparse data, meaningless distances, and overfitting.
Solution: PCA or feature selection.

---

15. What is regularization?

Answer:
Adds a penalty to reduce model complexity.

- **L1 (Lasso):** Shrinks some coefficients to zero (feature selection).
- **L2 (Ridge):** Shrinks coefficients but keeps all features.

---

16. Core assumptions of linear regression

Answer:

1. Linearity
2. Independence
3. Homoscedasticity
4. Normality of residuals
5. No multicollinearity
   Violations lead to biased or invalid results.

---

17. Role of activation function in logistic regression

Answer:
Uses **sigmoid** to map linear output (z = wx + b) to a probability between 0 and 1 for binary classification.

---

18. How to interpret logistic regression coefficients?

Answer:
Coefficients represent the log odds of the outcome. Used to calculate metrics like accuracy, precision, recall, and F1.

---

19. How do decision trees work?

Answer:

- Splits data by conditions using metrics like Gini or information gain.
- **Nodes:** Features
- **Branches:** Decisions
- **Leaves:** Predictions
  Note: Simple but can overfit; use pruning or Random Forest.

---

20. What is the motivation behind Random Forests?

Answer:
Combines many decision trees trained on random subsets of data and features.
Benefits: Reduces variance, improves accuracy, resists overfitting.

---

21. Difference between bagging and boosting

Answer:

- **Bagging:** Parallel training, reduces variance (e.g., Random Forest).
- **Boosting:** Sequential training, reduces bias (e.g., AdaBoost, XGBoost).

---

22. Hard vs. soft voting in ensemble methods

Answer:

- **Hard Voting:** Majority vote of predicted classes.
- **Soft Voting:** Averages predicted probabilities.
  Soft voting is generally more accurate.

---

23. What is k-NN and how does it work?

Answer:
Finds the *k* nearest points to predict class or value.

- **Classification:** Majority vote.
- **Regression:** Average value.
  Needs: Feature scaling and proper k selection.

---

24. What is k-Means and how does it work?

Answer:
Groups data into *k* clusters based on distance to centroids.
Steps:

1. Choose k centroids.
2. Assign points.
3. Recalculate centroids.

4. Repeat.
   Drawbacks: Sensitive to initial points, requires k.

---

25. How to select the best k in k-Means?

Answer:

- **Elbow Method:** Look for sharp bend in inertia plot.
- **Silhouette Score:** Higher = better clusters.
- **Gap Statistic:** Compares within-cluster variation.

---

26. What is DBSCAN and why is it better than k-Means?

Answer:
Density-based clustering that detects arbitrary shapes and outliers.
Advantage: No need for k; handles noise; detects non-linear clusters.

---

27. Feature selection vs. feature extraction

Answer:

- **Selection:** Choose most important features (e.g., correlation, RFE).
- **Extraction:** Transform features (e.g., PCA, autoencoders).

---

28. How to interpret feature importance in trees?

Answer:
Measured by how much a feature reduces impurity (Gini, entropy).
Higher importance → stronger impact on prediction.

---

29. What is PCA and when is it used?

Answer:
Reduces dimensionality by creating new uncorrelated components capturing max variance.
Used for noise reduction, visualization, and faster training.

---

30. What is LDA and when is it used?

Answer:
Supervised dimensionality reduction maximizing class separability.
Used in classification when data follows normal distribution.

---

31. How to handle multicollinearity?

Answer:

- Remove correlated variables.
- Use L1/L2 regularization.
- Apply PCA for uncorrelated components.

---

32. How to make a model robust to outliers?

Answer:

- Use robust algorithms (tree-based).
- Remove outliers (Z-score, IQR).
- Use Huber loss.
- Normalize data.

---

33. Generative vs. Discriminative models

Answer:

- **Generative:** Learn joint distribution P(X, Y) (e.g., Naive Bayes, GANs).
- **Discriminative:** Learn conditional P(Y|X) (e.g., Logistic Regression, SVM).
  Generative models can create data; discriminative focus on classification.

---

34. How to choose the right algorithm?

Answer:
Depends on:

- Problem type (classification, regression, clustering)
- Data size and quality
- Interpretability
- Computational cost
- Domain knowledge
  Try multiple models + EDA for best fit.

---

35. L1 vs. L2 regularization

Answer:

- **L1 (Lasso):** Adds |w| penalty → sparse model, feature selection.
- **L2 (Ridge):** Adds w^2 penalty → smooth shrinkage, all features kept.

---

36. What is the kernel trick in SVM?

Answer:
Maps non-linear data into higher dimensions using kernel functions (RBF, polynomial) without

explicit transformation.
Use: Enables linear separation in complex spaces.

---

37. Batch, mini-batch, and stochastic gradient descent

Answer:

- **Batch:** Uses all data per update → stable but slow.
- **Mini-batch:** Uses small batches → balanced and common.
- **Stochastic (SGD):** One sample per update → fast but noisy.

---

38. How does gradient descent work?

Answer:
Iteratively updates parameters opposite to gradient of loss to minimize error.
Learning rate (α): Controls step size; too high → diverge, too low → slow.

---

39. What is learning rate and how does it affect convergence?

Answer:

- Too high → skips optimum.
- Too low → slow convergence.
  Goal: Balance speed and accuracy.

---

40. What are hyperparameters and how to tune them?

Answer:
Settings before training (e.g., learning rate, tree depth).
Tuning Methods:

- Grid Search

- Random Search
- Bayesian Optimization
- Auto tools (e.g., Hyperopt, Optuna)

---

41. How to prevent overfitting during hyperparameter tuning?

Answer:

- Use cross-validation.
- Avoid overusing validation set.
- Apply early stopping.
- Use regularization.

---

42. Grid search vs. random search

Answer:

- **Grid:** Exhaustive search → small parameter spaces.
- **Random:** Random combinations → faster for large spaces.

---

43. How to evaluate a model with ROC curve?

Answer:
Plots **TPR (Recall)** vs **FPR** across thresholds.
AUC (Area Under Curve):

- 1 → perfect model

- 0.5 → random
  Higher AUC = better model.

---

44. What is silhouette score in clustering?

Answer:
Measures how well a point fits its cluster:
Score ∈ [-1, 1]:

- 1 → good clustering
- 0 → boundary
- -1 → wrong cluster
  Used for: Evaluating clustering quality and optimal k.

---

45. How to select features in high-dimensional data?

Answer:

- **Filter:** Correlation, chi-square.
- **Wrapper:** RFE, model-based search.
- **Embedded:** Lasso, tree importance.
- **Dimensionality Reduction:** PCA.

---

46. Difference between R^2 and Adjusted R^2

Answer:

- **R^2:** Proportion of variance explained.

- **Adjusted R^2:** Penalizes irrelevant predictors.
  Note: Adjusted R^2 can decrease when unnecessary variables are added.

---

47. Difference between feature selection and feature extraction

Answer:

- **Selection:** Keep best original features.
- **Extraction:** Create new transformed features.
  Both reduce dimensionality and improve performance.

---

48. A/B testing vs. ML deployment

Answer:

- **A/B Testing:** Compare versions experimentally.
- **ML Deployment:** Use trained model for real-time predictions.
  A/B validates; deployment operationalizes.

---

49. Purpose of test set vs. validation set

Answer:

- **Training Set:** Train model.
- **Validation Set:** Tune hyperparameters.
- **Test Set:** Final evaluation on unseen data.
  Note: Test data must remain untouched during training.

---

50. Stages in a Machine Learning Project

Answer:

1. Problem Definition
2. Data Collection
3. Data Cleaning & Preprocessing
4. Exploratory Data Analysis (EDA)
5. Feature Engineering & Selection
6. Model Training
7. Hyperparameter Tuning
8. Model Evaluation
9. Deployment
10. Monitoring & Maintenance