50 ML INTERVIEW QUESTIONS

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

- **Supervised Learning:** Uses labeled data to train models for predictive tasks.
    - Example: Regression (predicting continuous values) and Classification (predicting categories).
- **Unsupervised Learning:** Uses unlabeled data to find hidden patterns or groupings.
    - Example: Clustering, Dimensionality Reduction.

---

2. What is the difference between classification and regression?

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

---

3. What is the bias-variance tradeoff?

- **Bias:** Error from overly simplistic models → causes underfitting.
- **Variance:** Error from overly complex models → causes overfitting.
- Goal: Find balance to minimize total error.

---

4. How to deal with overfitting and underfitting?

- **Overfitting:** Too complex → use regularization, cross-validation, reduce complexity, early stopping.
- **Underfitting:** Too simple → add features, increase model complexity, reduce regularization.

---

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

- **Definition:** Splitting data into multiple folds to test model performance reliably (e.g., k-fold CV).
- **Importance:** Prevents overfitting, gives robust performance estimates, helps tune hyperparameters.

---

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

- **Precision:** TP / (TP + FP) → How many predicted positives are correct.
- **Recall:** TP / (TP + FN) → How many actual positives are detected.
- **F1-Score:** Harmonic mean of precision and recall — balances both.

---

7. How to choose the right evaluation metric?

- Based on **problem type**, **data imbalance**, and **business goal**.
- **Classification:** Accuracy, F1, ROC-AUC.
- **Regression:** MAE, MSE, R^2.
- Example: Fraud detection → Recall or AUC.

---

8. Difference between accuracy, precision, and recall?

- **Accuracy:** (TP + TN) / Total → Overall correctness.
- **Precision:** TP / (TP + FP).
- **Recall:** TP / (TP + FN).
- Use Precision when false positives are costly; Recall when false negatives are costly.

---

9. What is a confusion matrix?

- Table comparing predictions vs actual outcomes.
    - **Metrics:** Accuracy, Precision, Recall, F1-score.
    - **Shows:** TP, TN, FP, FN.

---

10. How do you handle missing or corrupted data?

- **Methods:**
    - Deletion (if small % missing).
    - Imputation (mean, median, mode, KNN).
    - Flag missing values.
    - Use models handling missing data (e.g., XGBoost).

---

11. How to handle categorical variables?

- **One-hot encoding:** For non-ordered categories.
- **Ordinal encoding:** For ordered categories.
- **Target encoding:** Replace with mean target value (for high cardinality).

---

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

- Process of creating, transforming, or selecting features to improve model performance.
- Includes scaling, encoding, interactions, and polynomial features.

---

13. Difference between parametric and non-parametric models?

- **Parametric:** Fixed number of parameters (e.g., Linear Regression). Fast but less flexible.
- **Non-parametric:** Grows with data (e.g., Decision Trees, KNN). More flexible but slower.

---

14. What is the curse of dimensionality?

- Too many features → data becomes sparse → harder learning, overfitting.
- **Fix:** Use PCA, feature selection, or collect more data.

---

15. What is regularization?

- Penalizes large coefficients → reduces overfitting.
- **Types:**

- L1 (Lasso): Shrinks some coefficients to zero.
- L2 (Ridge): Shrinks all coefficients slightly.

---

16. What are assumptions of linear regression?

1. Linearity.
2. Independence.
3. Homoscedasticity.
4. Normality of residuals.
5. No multicollinearity.

---

17. Role of activation functions in logistic regression?

- Sigmoid maps linear output into probability (0–1) → allows binary classification.

---

18. How to interpret coefficients in logistic regression?

- Each coefficient represents the log-odds change in the dependent variable for a one-unit change in the predictor.

---

19. How do decision trees work?

- Recursive splits using best feature (Gini, Entropy).
- **Pros:** Simple, interpretable.
- **Cons:** Can overfit → use pruning or ensembles.

---

20. What is random forest motivation?

- Combines many decision trees (bagging) → reduces variance, improves accuracy and robustness.

---

21. Difference between bagging and boosting?

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

---

22. Hard vs Soft Voting?

- **Hard Voting:** Majority class wins.
- **Soft Voting:** Average probabilities — usually better accuracy.

---

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

- Finds k nearest neighbors → predicts based on majority vote (classification) or average (regression).
- Requires feature scaling.

---

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

1. Choose k centroids.
2. Assign points to nearest centroid.
3. Update centroids until stable.

- Sensitive to initial placement and requires k beforehand.

---

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

- **Elbow Method, Silhouette Score, Gap Statistic.**
- Silhouette is most interpretable.

---

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

- Density-based clustering → forms clusters of arbitrary shape, no need to specify k, robust to noise.

---

27. Feature selection vs feature extraction

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

---

28. Feature importance in tree-based models

- Based on **reduction in impurity** or **split frequency**.
- Helps in interpretability and feature selection.

---

29. What is PCA and when to use it?

- Dimensionality reduction by transforming data into uncorrelated components capturing most variance.
- Useful for visualization and high-dimensional data.

---

30. What is LDA and when to use it?

- Supervised dimensionality reduction → maximizes class separability.
- Use when improving classification accuracy with labeled data.

---

31. How to handle multicollinearity?

- Remove correlated variables, use regularization, or apply PCA.

---

32. How to make models robust to outliers?

- Use robust algorithms (trees), detect/remove outliers (IQR, Z-score), use Huber loss, normalize data.

---

33. Difference between generative and discriminative models?

- **Generative:** Learn P(X, Y), can generate data (e.g., Naive Bayes, GANs).
- **Discriminative:** Learn P(Y|X), focus on boundaries (e.g., Logistic Regression, SVM).

---

34. How to choose which algorithm to use?

- Based on problem type, data size, interpretability, and resources.
- Use EDA and experiments to decide.

---

35. L1 vs L2 Regularization

- **L1 (Lasso):** Shrinks some weights to zero → feature selection.
- **L2 (Ridge):** Shrinks all weights slightly → stabilizes model.

---

36. What is the kernel trick in SVM?

- Allows linear separation in higher dimensions using kernel functions (RBF, Polynomial) without explicit mapping.

---

37. Batch, Mini-batch, and Stochastic Gradient Descent

- **Batch:** Uses all data → stable but slow.
- **Mini-batch:** Uses small groups → fast and stable (common in DL).
- **SGD:** Uses one sample → fast but noisy.

---

38. How does gradient descent work?

- Iteratively updates parameters opposite to gradient direction to minimize loss.
- Controlled by **learning rate**.

---

39. What is the learning rate?

- Step size for parameter updates.
    - Too high → diverge.
    - Too low → slow convergence.

---

40. What are hyperparameters and tuning methods?

- **Hyperparameters:** Control training (e.g., learning rate, tree depth).
- **Tuning Methods:** Grid Search, Random Search, Bayesian Optimization, AutoML.

---

41. How to prevent overfitting during hyperparameter tuning?

- Use cross-validation, early stopping, and regularization.
- Avoid over-tuning to validation set.

---

42. Grid Search vs Random Search

- **Grid Search:** Exhaustive, best for small spaces.
- **Random Search:** Faster, better for large spaces.

---

43. What is ROC curve and AUC?

- **ROC Curve:** Plots TPR vs FPR at different thresholds.
- **AUC:** Area under curve — higher = better discrimination.

---

44. What is Silhouette Score?

- Measures how well each point fits in its cluster.
    - +1 → well clustered.
    - 0 → boundary.
    - -1 → wrong cluster.

---

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

- **Filter:** Stats tests (correlation, chi-square).
- **Wrapper:** Model-based selection (RFE).
- **Embedded:** During training (Lasso, Trees).
- **PCA:** Dimensionality reduction.

---

46. What is R^2 and Adjusted R^2?

- **R^2:** % variance explained.
- **Adjusted R^2:** Penalizes irrelevant features — decreases if useless features added.

---

47. Difference between feature selection and extraction?

- **Selection:** Choose best original features.

- **Extraction:** Create new features from old ones (e.g., PCA).

---

48. A/B Testing vs Model Deployment

- **A/B Testing:** Compare versions to choose best.
- **Deployment:** Put trained model into production for real-time use.

---

49. Purpose of Test Set vs Validation Set

- **Validation:** For tuning during training.
- **Test:** For final evaluation on unseen data.
- **Key:** Test set used only once.

---

50. Stages in a Machine Learning Project

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