Top 100 Machine Learning & AI Interview Questions

1. Fundamentals of Machine Learning

1. What is the difference between supervised, unsupervised, and reinforcement learning?
2. What is overfitting and underfitting?
3. How do you prevent overfitting?
4. Explain bias–variance tradeoff.
5. What is cross-validation and why is it used?
6. What is regularization?
7. Compare L1 and L2 regularization.
8. What is the curse of dimensionality?
9. What is feature scaling and why is it important?
10. What are hyperparameters vs parameters?

2. Regression & Classification

1. Explain linear regression and its assumptions.
2. What is multicollinearity and how do you handle it?
3. What is logistic regression and how does it differ from linear regression?
4. What is the ROC curve and AUC score?
5. What is precision, recall, and F1-score?
6. How do you handle imbalanced datasets?
7. What is the confusion matrix?
8. What is regularized regression (Ridge/Lasso)?
9. What are residuals in regression?
10. What metrics would you use to evaluate a regression model?

3. Decision Trees & Ensemble Methods

1. How does a decision tree decide where to split?
2. What is Gini impurity vs entropy?
3. What is pruning in decision trees?
4. Explain bagging and boosting.
5. How does Random Forest work?
6. What is feature importance in Random Forest?
7. How does Gradient Boosting work?
8. What is the difference between AdaBoost, XGBoost, LightGBM, and CatBoost?

9. Why can ensemble methods outperform single models?
10. How do you avoid overfitting in ensemble models?

4. Clustering & Dimensionality Reduction

1. Explain K-Means clustering.
2. How do you choose the number of clusters (k)?
3. What are the limitations of K-Means?
4. What is hierarchical clustering?
5. What is DBSCAN?
6. Explain PCA and its mathematical intuition.
7. What is the difference between PCA and t-SNE?
8. What are eigenvalues and eigenvectors in PCA?
9. How do you interpret principal components?
10. What is the purpose of dimensionality reduction?

5. Time Series & Forecasting

1. What is stationarity?
2. How do you check if a time series is stationary?
3. What are AR, MA, ARMA, and ARIMA models?
4. What is seasonal decomposition?
5. How do you handle missing values in time series?
6. What is autocorrelation and partial autocorrelation?
7. Explain exponential smoothing.
8. What is differencing?
9. What metrics do you use to evaluate forecasting models?
10. What is drift in time series?

6. Neural Networks & Deep Learning

1. What is a perceptron?
2. Explain forward and backward propagation.
3. What is gradient descent?
4. What are vanishing and exploding gradients?
5. What is ReLU and why is it popular?
6. What are CNNs and their use cases?
7. What are RNNs and LSTMs?
8. What is the attention mechanism?
9. What is transfer learning?

10. What are dropout and batch normalization?

7. Generative AI & LLMs

1. What is a Transformer architecture?
2. How do self-attention and multi-head attention work?
3. What is positional encoding?
4. How do models like GPT or BERT differ?
5. What is fine-tuning vs pre-training?
6. What is prompt engineering?
7. What are embeddings in LLMs?
8. What is Retrieval-Augmented Generation (RAG)?
9. What is quantization and why is it used?
10. What is model distillation?

8. Reinforcement Learning

1. What is the difference between policy and value function?
2. Explain Q-learning.
3. What is exploration vs exploitation?
4. What are rewards and returns?
5. What is the Bellman equation?
6. What is a Markov Decision Process (MDP)?
7. What is Deep Q-Network (DQN)?
8. What are actor–critic methods?
9. What is reward shaping?
10. Explain policy gradient methods.

9. Data Engineering & Feature Engineering

1. What are missing-value handling techniques?
2. What is feature selection vs feature extraction?
3. What is one-hot encoding vs label encoding?
4. What is normalization vs standardization?
5. How do you handle categorical and numerical features together?
6. What are outliers and how do you detect them?
7. How do you handle high cardinality categorical variables?
8. What is target encoding?
9. How do you select important features?
10. What is data leakage and how do you prevent it?

10. Model Evaluation, Deployment & Ethics

1. What are common evaluation metrics for ML models?
2. What is cross-entropy loss?
3. What is early stopping?
4. How do you deploy ML models to production?
5. What are model drift and concept drift?
6. What is A/B testing in model evaluation?
7. How do you ensure model explainability?
8. What is SHAP vs LIME?
9. How do you ensure fairness and avoid bias in ML?
10. How do you monitor ML models post-deployment?