ENTRI

100 Machine Learning Interview Questions and Answers

1. What is Machine Learning?

Machine Learning (ML) is a subset of artificial intelligence where machines learn patterns from data without being explicitly programmed.

2. Differentiate between Supervised, Unsupervised, and Semi-Supervised Learning.

- Supervised Learning uses labeled data for training.
- Unsupervised Learning uses unlabeled data to find patterns.
- Semi-Supervised Learning uses both labeled and unlabeled data.

3. What is Reinforcement Learning?

Learning through trial and error, where an agent receives rewards or penalties to learn optimal behavior.

4. What are the different types of data used in Machine Learning?

Structured, unstructured, and semi-structured data.

5. Difference between Regression and Classification?

Regression predicts continuous outcomes, classification predicts discrete classes.

6. Define Features and Labels.

Features are input variables; labels are the output or target.

7. What is Scikit-learn?

A Python library for ML algorithms and data processing.

ENTRI

8. What are Training Set and Test Set?

Training set is for fitting the model; test set evaluates its performance.

9. List the stages of building a Machine Learning model.

Data collection → Preprocessing → Model selection → Training → Evaluation → Deployment.

10. What is a Confusion Matrix?

A table used to evaluate classification models showing TP, TN, FP, FN.

11. What are Type I and Type II errors?

Type I is false positive; Type II is false negative.

12. Define Precision, Recall, Accuracy, and F1 Score.

Precision is TP/(TP+FP), recall is TP/(TP+FN), accuracy is correct predictions/total

predictions, F1 score is harmonic mean of precision and recall.

13. What is the P-value?

Probability that observed results are due to chance under null hypothesis.

14. Explain ROC Curve.

Plot of true positive rate vs false positive rate at various thresholds.

15. How is KNN different from k-means clustering?

KNN is supervised classification; k-means is unsupervised clustering.

16. What does ‘Naive’ in Naive Bayes mean?

Assumes independent features.

17. What is Overfitting?

When the model performs well on training data but poorly on unseen data.

18. What is Underfitting?

When a model is too simple to capture the data pattern.

19. How to prevent overfitting?

Use techniques like cross-validation, regularization, pruning, early stopping.

20. What is Cross-Validation?

A technique to evaluate model's generalization by partitioning data into folds.

21. Define Bias and Variance.

Bias is error from erroneous assumptions; variance is error from sensitivity to data fluctuations.

22. What is Regularization?

Technique to reduce overfitting by adding penalty terms to loss functions.

23. Difference between L1 and L2 regularization?

L1 adds absolute weights penalty (sparsity), L2 adds squared weights penalty.

24. What is Gradient Descent?

Optimization algorithm to minimize the loss by iteratively adjusting parameters.

25. Different types of Gradient Descent?

Batch, Stochastic, Mini-batch.

26. What is a Decision Tree?

A tree-like model used for classification and regression.

27. Explain Random Forest.

An ensemble of decision trees to improve accuracy and reduce overfitting.

28. What is Ensemble Learning?

Combining multiple models to improve performance.

29. Difference between Bagging and Boosting?

Bagging reduces variance by averaging; Boosting reduces bias by sequentially correcting errors.

30. What is PCA (Principal Component Analysis)?

A technique to reduce dimensionality by projecting data onto principal components.

31. What is Clustering?

Grouping similar data points together.

32. Explain K-means Clustering.

An algorithm that partitions data into k clusters by minimizing distances.

33. What is Hierarchical Clustering?

Cluster data into a tree of clusters.

34. What is the Curse of Dimensionality?

High-dimensional data becomes sparse and affects model performance.

35. Types of Machine Learning Algorithms?

Supervised, Unsupervised, Reinforcement Learning algorithms.

36. What is a Neural Network?

A network of interconnected nodes inspired by the human brain.

37. What is Deep Learning?

A subset of ML using multi-layered neural networks.

38. What are Activation Functions?

Functions like ReLU, sigmoid used to introduce non-linearity in networks.

39. What is Backpropagation?

An algorithm for training neural networks by updating weights based on error gradient.

40. Difference between Batch and Online Learning?

Batch learns from all data at once; online learns incrementally.

41. What is a Confusion Matrix used for?

Evaluating classification model effectiveness.

42. What is Data Preprocessing?

Cleaning and transforming raw data before model training.

43. Why is Feature Scaling important?

Ensures features contribute equally to distance calculations in models.

44. What are Hyperparameters?

Settings used to control the learning process.

45. Difference between Parametric and Non-parametric Models?

Parametric assume fixed number of parameters; non-parametric grow with data.

46. What is ROC AUC?

Area under the ROC curve measuring classification performance.

47. Define Precision-Recall Curve.

Graph showing trade-off between precision and recall for different thresholds.

48. What is an Epoch?

One pass over entire training data in neural networks.

49. What is Over-sampling and Under-sampling?

Techniques to balance imbalanced datasets.

50. What is Feature Engineering?

Creating new features from raw data to improve model performance.

51. What is a Confounder?

A variable influencing both independent and dependent variables.

52. What is Concept Drift?

When the statistical properties of target variable change over time.

53. What is a Kernel in SVM?

Function that transforms data to higher dimension for linear separability.

54. Difference between Parametric and Non-Parametric SVM?

Parametric uses fixed parameters; non-parametric depends on the data points (support vectors).

55. What is Bias-Variance Tradeoff?

Balancing underfitting (bias) and overfitting (variance) in model.

56. What is Early Stopping?

Halting model training when performance on validation set stops improving.

57. What is Data Leakage?

When information from outside training data influences the model.

58. What is the purpose of Softmax?

Converts outputs into probability distribution for multi-class classification.

59. What is a Confusion Matrix?

See Question 10.

60. What is Clustering?

See Question 31.

61. What is the difference between Logistic Regression and Linear Regression?

Logistic for classification; linear for regression problems.

62. What is Gradient Boosting?

An ensemble technique combining weak learners to create strong prediction.

63. What is AdaBoost?

Adaptive boosting algorithm that adjusts weights of training examples.

64. What is XGBoost?

An optimized gradient boosting framework.

65. What is Bagging?

Bootstrap aggregating that trains multiple models on random subsets.

66. What are Outliers?

Data points distant from others, potentially skewing models.

67. What is the Central Limit Theorem?

Sum of distributions tends toward normal distribution as sample size increases.

68. What is Cross-Entropy Loss?

Loss function for classification measuring difference between two probability distributions.

69. What is RMSE?

Root Mean Square Error, a metric for regression accuracy.

70. What is Mean Absolute Error?

Average of absolute errors between predicted and actual values.

71. What is Hyperparameter Tuning?

Optimizing model parameters for best performance.

72. What is Grid Search?

Systematic way of tuning hyperparameters.

73. What is Random Search?

Selecting random combinations of hyperparameters to tune.

74. What is the difference between Generative and Discriminative models?

Generative models learn joint probability; discriminative learn decision boundaries.

75. What is a Markov Chain?

Model describing sequence of events with probabilities based on current state.

76. What is the Curse of Dimensionality?

See Question 34.

77. What is Bootstrapping?

Sampling method to estimate statistics from data.

78. What is the No-Free-Lunch Theorem?

No one model works best for all problems.

79. What is Transfer Learning?

Using knowledge from one task to improve performance on another.

80. What is the difference between ANN and CNN?

ANNs are general; CNNs specialize in spatial hierarchies like images.

81. What is Dropout in Neural Networks?

Technique to prevent overfitting by randomly disabling neurons.

82. What is Batch Normalization?

Method to stabilize and accelerate training deep networks.

83. What is Loss Function?

Function that quantifies error during training.

84. What is Epoch?

See Question 48.

85. What is Learning Rate?

Step size in updating model parameters during training.

86. What is Gradient Vanishing Problem?

When gradients become too small for parameters to update.

87. What is Early Stopping?

See Question 56.

88. What is LSTM?

Long Short-Term Memory networks good for sequential data.

89. What is Reinforcement Learning?

See Question 3.

90. What is Deep Learning?

Subset of ML using deep neural networks.

91. What are Convolutional Layers?

Layers in CNN that detect spatial features.

92. What are Recurrent Neural Networks?

Networks designed for sequence data processing.

93. What is Overfitting?

See Question 17.

94. What is the difference between Epoch and Batch?

Epoch is full pass over data; batch is subset passed to the model.

95. What is a Learning Curve?

Graph of model performance vs number of training samples.

96. What is Feature Selection?

Choosing most relevant features for better model.

ENTRI

97. What are Dimensionality Reduction techniques?

Methods like PCA, t-SNE to reduce data features.

98. What is a Decision Boundary?

The line/surface separating classes in classification tasks.

99. What are Support Vectors?

Data points closest to decision boundary in SVM.

100. What is Model Evaluation?

Process of assessing model’s predictive performance using metrics.