

Cracking the Code: 200+ Interview MCQs

(on Data Science, Machine Learning and Deep Learning)



1. In machine learning, what term is used to describe the problem of a model performing well on the training data but poorly on unseen data?

- A) Stochastic Gradient Descent (SGD)
- B) Adam Optimizer
- C) Levenberg-Marquardt Algorithm

2. Which of the following algorithms is best suited for anomaly detection in unstructured data like images or audio?

- A) K-means clustering
- B) Support Vector Machines (SVM)
- C) Autoencoders

3. In a neural network, what is the purpose of the activation function?

- A) Normalize the input data
- B) Regularize the model to prevent overfitting
- C) Introduce non-linearity to the model

4. Which evaluation metric is commonly used for imbalanced classification problems to assess model performance?

- A) Mean Squared Error (MSE)
- B) Receiver Operating Characteristic (ROC) AUC
- C) R-squared (R²) score

5. In the context of Natural Language Processing (NLP), what does TF-IDF stand for?

- A) Term Frequency-Inverse Document Frequency
- B) Text Frequency-Inverse Data Filter
- C) Total Frequency-Inverse Document Filtering

6. Which technique is used to reduce the dimensionality of data while preserving its variance in multivariate analysis?

- A) Principal Component Analysis (PCA)
- B) Support Vector Machine (SVM)
- C) Total Frequency-Inverse Document Filtering

7. In Reinforcement Learning, what is the name of the technique used to estimate the expected long-term reward from a given state?

- A) Q-learning
- B) Policy Gradient
- C) Value Function

8. Which method is used for handling missing data by creating multiple imputations based on the observed data?

- A) K-nearest neighbors (KNN)
- B) Mean substitution
- C) Multiple Imputation by Chained Equations (MICE)

9. Which algorithm is a variant of the Random Forest algorithm and uses gradient boosting for improved performance?

- A) XGBoost
- B) Naive Bayes
- C) K-means

10. In deep learning, which layer type is responsible for reducing the spatial dimensions of the input data?

- A) Dense layer
- B) Convolutional layer
- C) Pooling layer

11. Which optimization algorithm updates the model parameters with a learning rate that adapts based on the historical gradients?

- A) Stochastic Gradient Descent (SGD)
- B) Adam Optimizer
- C) Levenberg-Marquardt Algorithm

12. In a time series forecasting problem, what is the term used to describe the pattern where the data exhibits upward or downward trends over a long period?

- A) Seasonality
- B) Cyclical pattern
- C) Trend

13. Which statistical test is used to determine if there is a significant relationship between two categorical variables in a dataset?

- A) Pearson correlation coefficient
- B) Chi-Square test
- C) t-test

14. When building a recommendation system, which method uses collaborative filtering based on user-item interactions?

- A) Content-based filtering
- B) Singular Value Decomposition (SVD)
- C) Apriori algorithm

15. Which algorithm is commonly used for semi-supervised learning tasks, where the training data has both labeled and unlabeled examples?

- A) K-nearest neighbors (KNN)
- B) Decision Tree
- C) Expectation-Maximization (EM)

16. In time series analysis, what is the purpose of differencing in the Box-Jenkins method?

- A) Remove seasonality
- B) Make the series stationary
- C) Calculate autocorrelation

17. What is the primary advantage of using an ensemble of models, such as Random Forest or Gradient Boosting Machines?

- A) Faster training time
- B) Simpler model interpretation
- C) Improved prediction accuracy

18. Which technique is used for text data preprocessing in NLP, where words are converted to their base or root form?

- A) Tokenization
- B) Lemmatization
- C) Bag-of-Words

19. Which regularization term is added to the cost function during training to penalize large weights in a neural network?

- A) L1 regularization (Lasso)
- B) L2 regularization (Ridge)
- C) ElasticNet regularization

20. Which algorithm is commonly used for finding the optimal number of clusters in a dataset when the number of clusters is unknown?

- A) K-means clustering
- B) Hierarchical clustering
- C) Silhouette analysis

21. In time series analysis, what is the autocorrelation of a time series at lag k ?

- A) Time series lagged correlation.
- B) Cross-series lagged correlation.
- C) Time series moving correlation.

22. Which method is used for hyperparameter tuning by exhaustively trying all possible combinations of hyperparameter values?

- A) Grid Search
- B) Random Search
- C) Bayesian Optimization

23. In a classification problem, which metric can be used to evaluate the model's ability to distinguish between positive and negative samples?

- A) F1 score
- B) Mean Absolute Error (MAE)
- C) R-squared (R^2) score

24. Which technique is used to prevent data leakage when scaling numerical features before feeding them into a machine learning model?

- A) Min-Max scaling
- B) Standardization
- C) Scaling after train-test split

25. In natural language processing, what does NER stand for?

- A) Named Entity Recognition
- B) Natural Entity Retrieval
- C) Normalized Entity Resolution

26. Which algorithm is commonly used for document clustering in unsupervised learning?

- A) Decision Tree
- B) K-means clustering
- C) Random Forest

26. Which algorithm is commonly used for document clustering in unsupervised learning?

- A) Decision Tree
- B) K-means clustering
- C) Random Forest

27. In a decision tree, which attribute selection measure aims to maximize the information gain?

- A) Gini impurity
- B) Entropy
- C) Variance

28. Which method is used for reducing the number of features in a dataset while preserving the most important information?

A) Principal Component Analysis (PCA)

B) Forward Feature Selection

C) Recursive Feature Elimination (RFE)

29. In deep learning, which activation function is often used in the output layer of a binary classification neural network?

A) Sigmoid

B) Tanh

C) ReLU

30. Which machine learning algorithm is used for generating synthetic samples to balance class distribution in an imbalanced dataset?

A) Decision Tree

B) SMOTE (Synthetic Minority Over-sampling Technique)

C) AdaBoost

31. In Bayesian inference, what term is used to represent the prior belief about a model parameter before seeing the data?

A) Posterior

B) Likelihood

C) Prior

32. Which algorithm is used to detect patterns or associations between items in a transactional database?

A) K-means clustering

B) Apriori algorithm

C) Support Vector Machines (SVM)

33. In time series analysis, what is the difference between seasonal and cyclic patterns?

A) Seasonal: Fixed period, cyclic: Variable period.

B) Seasonal: Short duration, cyclic: Longer duration.

C) Seasonal: More predictable, cyclic: Less predictable.

34. Which algorithm is an extension of linear regression that can handle multiple outputs, making it suitable for multi-output regression tasks?

A) Logistic Regression

B) Lasso Regression

C) Multivariate Linear Regression

35. In unsupervised learning, what is the objective of clustering algorithms?

- A) Predict a target variable based on input features.
- B) Group similar data points together based on their characteristics.
- C) Minimize the error between predicted and actual values.

36. Which regularization term adds a penalty based on the absolute value of the model coefficients in linear regression?

- A) L1 regularization (Lasso)
- B) L2 regularization (Ridge)
- C) ElasticNet regularization

37. In Natural Language Processing (NLP), what is the purpose of tokenization?

- A) Preprocess text data by converting words to lowercase.
- B) Split the text into individual words or tokens.
- C) Remove punctuation and special characters from the text.

38. Which ensemble learning method combines predictions from multiple models through a weighted voting system to make final predictions?

- A) Bagging
- B) Boosting
- C) Stacking

39. In time series analysis, what is the order of differencing used to make a non-stationary time series stationary?

- A) First-order differencing
- B) Second-order differencing
- C) It depends on the series characteristics

40. Which machine learning algorithm is suitable for solving multi-class classification problems by reducing them into multiple binary classification tasks?

- A) K-nearest neighbors (KNN)
- B) Logistic Regression
- C) Decision Tree

41. In deep learning, what is the role of the learning rate in the training process?

- A) Hidden layer control.
- B) Output layer size.
- C) Weight update step size.

42. Which method is used to measure the similarity between two documents in natural language processing?

- A) Euclidean distance
- B) Jaccard similarity
- C) Pearson correlation coefficient

43. In reinforcement learning, what is the term used for the agent's strategy for selecting actions to maximize the cumulative reward over time?

- A) Policy
- B) Value Function
- C) Q-function

44. Which algorithm is commonly used for regression tasks when the target variable exhibits a non-linear relationship with the input features?

- A) Linear Regression
- B) Decision Tree
- C) K-means clustering

45. In machine learning, what is the purpose of cross-validation?

- A) Multi-dataset training.
- B) Independent dataset assessment.
- C) Train-test data division.

46. Which method is used for dimensionality reduction in feature space by transforming data to a lower-dimensional representation that preserves pairwise distances?

- A) t-SNE (t-distributed Stochastic Neighbor Embedding)
- B) LDA (Linear Discriminant Analysis)
- C) Isomap (Isometric Mapping)

47. In a classification problem, which metric is used to evaluate the model's ability to correctly classify positive samples?

- A) Mean Squared Error (MSE)
- B) Precision
- C) R-squared (R²) score

48. Which method is used for feature selection by recursively removing the least important features based on model performance?

- A) Recursive Feature Elimination (RFE)
- B) Principal Component Analysis (PCA)
- C) Variance Thresholding

49. In natural language processing, what is the term for reducing words to their base or root form to capture their core meaning?

- A) Tokenization
- B) Lemmatization
- C) Stopword Removal

50. Which algorithm is commonly used for time series forecasting and can handle both trend and seasonality in the data?

- A) Support Vector Machines (SVM)
- B) K-nearest neighbors (KNN)
- C) Autoregressive Integrated Moving Average (ARIMA)

51. In ensemble learning, which method uses multiple weak learners sequentially, where each learner corrects the errors of its predecessor?

- A) Bagging
- B) Boosting
- C) Stacking

52. Which technique is used to preprocess text data by converting all words to lowercase to ensure consistent comparisons ?

- A) Tokenization
- B) Lemmatization
- C) Lowercasing

53. In deep learning, what is the role of the loss function in training a neural network?

- A) Regularize the model to prevent overfitting.
- B) Measure the model's prediction error during training.
- C) Control the learning rate for gradient updates.

54. Which machine learning algorithm is well-suited for outlier detection and can be used for one-class classification?

- A) Random Forest
- B) k-Nearest Neighbors (k-NN)
- C) Support Vector Machines (SVM)

55. In time series analysis, what is the term for patterns that repeat at regular intervals, such as daily, weekly, or yearly patterns?

- A) Seasonality
- B) Trend
- C) Cyclical pattern

56. Which method is used for hyperparameter tuning that employs probabilistic models to model the objective function and decide which hyperparameters to try next?

- A) Grid Search
- B) Random Search
- C) Bayesian Optimization

57. In a classification problem, which metric is used to evaluate the model's ability to correctly identify negative samples?

- A) Mean Absolute Error (MAE)
- B) Recall
- C) R-squared (R2) score

58. Which preprocessing technique is used to scale numerical features to a specific range, such as [0, 1]?

- A) Min-Max scaling
- B) Standardization
- C) Normalization

59. Which technique is used to handle missing data in a dataset?

- A) Feature scaling
- B) Data imputation
- C) Principal Component Analysis (PCA) answer

60. Which ensemble learning method combines multiple models through averaging or voting to make final predictions?

- A) Bagging
- B) Boosting
- C) Stacking

61. In a time series, what is the difference between lagged variables and moving averages?

- A) Lagged variables represent past.
- B) Lagged variables represent future.
- C) Different concepts, not interchangeable.

62. Which technique is used to prevent data leakage during feature engineering by fitting transformations based only on the training data?

- A) Min-Max scaling
- B) Standardization
- C) Scaling after train-test split

63. In deep learning, what is the primary function of a dropout layer?

- A) Introduce noise into the data to prevent overfitting.
- B) Regularize the model by randomly dropping some neurons during training.
- C) Reduce the dimensionality of the data.

64. Which algorithm is commonly used for text classification tasks when the text data is represented as word frequencies or TF-IDF vectors?

- A) Decision Tree
- B) k-Nearest Neighbors (k-NN)
- C) Naive Bayes

65. In Bayesian inference, what is the term used for updating the prior belief with new data to obtain the posterior distribution?

- A) Maximum Likelihood Estimation (MLE)
- B) Maximum a Posteriori (MAP) estimation
- C) Bayes' Theorem

66. Which algorithm is commonly used for finding the optimal number of clusters in a dataset when the number of clusters is unknown?

- A) K-means clustering
- B) Hierarchical clustering
- C) Silhouette analysis

67. In time series analysis, what is the term for a model that combines autoregressive (AR), moving average (MA), and differencing components?

- A) Autoregressive Integrated Moving Average (ARIMA)
- B) Exponential Smoothing (ETS)
- C) Seasonal Autoregressive Integrated Moving-Average (SARIMA)

68. Which machine learning algorithm is used for predicting continuous numeric values, such as house prices or stock prices?

- A) Logistic Regression
- B) Decision Tree
- C) Linear Regression

69. In ensemble learning, what is the purpose of bagging techniques, such as Random Forest?

- A) Ensemble model decision combining.
- B) Model averaging variance reduction.
- C) Sequential boosting weak learners.

70. Which method is used for data preprocessing to handle categorical variables by converting them into numerical form?

- A) One-Hot Encoding
- B) Standardization
- C) Feature Scaling

71. In deep learning, which activation function is commonly used in the hidden layers of a neural network to introduce non-linearity?

- A) Sigmoid
- B) Tanh
- C) ReLU (Rectified Linear Unit)

72. Which machine learning algorithm is commonly used for binary classification problems when the data is linearly separable?

- A) Decision Tree
- B) Naive Bayes
- C) Logistic Regression

73. In reinforcement learning, what is the term used for the process of exploring the environment and collecting experiences to improve the agent's policy?

- A) Q-learning
- B) Policy Gradient
- C) Exploration-Exploitation trade-off

74. Which method is used to handle missing data by filling in missing values based on the average of the non-missing values in the same column?

- A) K-nearest neighbors (KNN)
- B) Mean substitution
- C) Multiple Imputation by Chained Equations (MICE)

75. In Natural Language Processing (NLP), which technique is used to convert words into their base form by removing suffixes or prefixes?

- A) Tokenization
- B) Lemmatization
- C) Stemming

76. Which evaluation metric is commonly used for regression tasks to assess the variance explained by the model's predictions?

- A) F1 score
- B) Mean Absolute Error (MAE)
- C) R-squared (R2) score

77. In deep learning, what is the purpose of the optimizer in the training process of a neural network?

- A) Initialize the model weights randomly.
- B) Reduce the learning rate as the training progresses.
- C) Update the model weights to minimize the loss function.

78. Which machine learning algorithm is suitable for solving multi-class classification problems by transforming them into multiple binary classification tasks?

- A) K-nearest neighbors (KNN)
- B) Logistic Regression
- C) Decision Tree

79. In ensemble learning, which technique is used to combine multiple models into one weighted model based on their performance?

- A) Bagging
- B) Boosting
- C) Stacking

80. Which method is used for dimensionality reduction by mapping data points into a lower-dimensional space while preserving pairwise distances?

- A) Principal Component Analysis (PCA)
- B) Linear Discriminant Analysis (LDA)
- C) t-distributed Stochastic Neighbor Embedding (t-SNE)

81. In deep learning, which activation function is commonly used in the output layer of a multi-class classification neural network?

- A) Sigmoid
- B) Tanh
- C) Softmax

82. Which algorithm is commonly used for document clustering in unsupervised learning?

- A) Decision Tree
- B) K-means clustering
- C) Random Forest

83. In a decision tree, which attribute selection measure aims to minimize the Gini impurity? 83. In a decision tree, which attribute selection measure aims to minimize the Gini impurity?

- A) Gini impurity
- B) Entropy
- C) Variance

84. Which method is used for reducing the number of features in a dataset by selecting the most important ones based on statistical tests or scoring functions?

- A) Principal Component Analysis (PCA)
- B) Forward Feature Selection
- C) Recursive Feature Elimination (RFE)

85. In deep learning, what is the role of the dropout technique during training a neural network?

- A) Data noise for overfitting.
- B) Dropout regularization for neurons.
- C) Learning rate control gradient.

86. Which machine learning algorithm is used for predicting continuous numeric values, such as house prices or stock prices?

- A) Logistic Regression
- B) Decision Tree
- C) Linear Regression

87. In natural language processing, what is the term for reducing words to their base or root form to capture their core meaning?

- A) Tokenization
- B) Lemmatization
- C) Stopword Removal

89. In a time series, what is the difference between lagged variables and moving averages?

- A) Lagged: Past, Moving: Window.
- B) Lagged: Future, Moving: Window.
- C) Different concepts, not interchangeable.

90. Which technique is used to prevent data leakage during feature engineering by fitting transformations based only on the training data?

- A) Min-Max scaling
- B) Standardization
- C) Scaling after train-test split

91. Which algorithm is a variant of the Random Forest algorithm and uses gradient boosting for improved performance?

- A) XGBoost
- B) Naive Bayes
- C) K-means

92. Which algorithm is commonly used for text classification tasks when the text data is represented as word frequencies or TF-IDF vectors?

- A) Decision Tree
- B) k-Nearest Neighbors (k-NN)
- C) Naive Bayes

93. In Bayesian inference, what is the term used for updating the prior belief with new data to obtain the posterior distribution?

- A) Maximum Likelihood Estimation (MLE)
- B) Maximum a Posteriori (MAP) estimation
- C) Bayes' Theorem

94. Which algorithm is commonly used for finding the optimal number of clusters in a dataset when the number of clusters is unknown?

- A) K-means clustering
- B) Hierarchical clustering
- C) Silhouette analysis

95. In time series analysis, what is the term for a model that combines autoregressive (AR), moving average (MA), and differencing components?

- A) Autoregressive Integrated Moving Average (ARIMA)
- B) Exponential Smoothing (ETS)
- C) Seasonal Autoregressive Integrated Moving-Average (SARIMA)

96. Which machine learning algorithm is commonly used for binary classification problems when the data is linearly separable?

- A) Decision Tree
- B) Naive Bayes
- C) Logistic Regression

97. In reinforcement learning, what is the term used for the process of exploring the environment and collecting experiences to improve the agent's policy?

- A) Q-learning
- B) Policy Gradient
- C) Exploration-Exploitation trade-off

98. Which method is used to handle missing data by filling in missing values based on the average of the non-missing values in the same column?

- A) K-nearest neighbors (KNN)
- B) Mean substitution
- C) Multiple Imputation by Chained Equations (MICE)

99. In Natural Language Processing (NLP), which technique is used to convert words into their base form by removing suffixes or prefixes?

- A) Tokenization
- B) Lemmatization
- C) Stemming

100. Which evaluation metric is commonly used for regression tasks to assess the variance explained by the model's predictions?

- A) R-squared (R²) score
- B) Mean Absolute Error (MAE)
- C) F1 score

101. In deep learning, what is the purpose of the optimizer in the training process of a neural network?

- A) Initialize the model weights randomly.
- B) Update the model weights to minimize the loss function.
- C) Reduce the learning rate as the training progresses.

102. Which machine learning algorithm is suitable for solving multi-class classification problems by transforming them into multiple binary classification tasks?

- A) K-nearest neighbors (KNN)
- B) Logistic Regression
- C) Decision Tree

103. In ensemble learning, which technique is used to combine multiple models into one weighted model based on their performance?

- A) Bagging
- B) Boosting
- C) Stacking

104. Which method is used for dimensionality reduction by mapping data points into a lower-dimensional space while preserving pairwise distances?

- A) Principal Component Analysis (PCA)
- B) Linear Discriminant Analysis (LDA)
- C) t-distributed Stochastic Neighbor Embedding (t-SNE)

105. Which method is used for handling missing data by creating multiple imputations based on the observed data?

- A) K-nearest neighbors (KNN)
- B) Mean substitution
- C) Multiple Imputation by Chained Equations (MICE)

106. Which algorithm is commonly used for document clustering in unsupervised learning?

- A) Decision Tree
- B) K-means clustering
- C) Random Forest

107. In a decision tree, which attribute selection measure aims to minimize the Gini impurity?

- A) Gini impurity
- B) Entropy
- C) Variance

108. Which method is used for reducing the number of features in a dataset by selecting the most important ones based on statistical tests or scoring functions?

- A) Principal Component Analysis (PCA)
- B) Forward Feature Selection
- C) Recursive Feature Elimination (RFE)

109. In Reinforcement Learning, what is the name of the technique used to estimate the expected long-term reward from a given state?

- A) Q-learning
- B) Policy Gradient
- C) Value Function

110. Which machine learning algorithm is used for predicting continuous numeric values, such as house prices or stock prices?

- A) Logistic Regression
- B) Decision Tree
- C) Linear Regression

111. What are autoencoders in deep learning?

Models used for reinforcement learning

Networks designed to reconstruct input data

Neural networks used for unsupervised learning

112. Which layers are typically present in autoencoders?

Encoder and Classifier layers

Input and Output layers

Encoder and Decoder layers

113. Two practical usages of autoencoders are:

Image denoising, Anomaly detection

Image classification, Speech recognition

Text generation, Language translation

114. What is an activation function in deep learning?

A function to convert numeric data to categorical data

A function used to define the output of a neuron

A function used to define the hidden layers

115. Three different types of activation functions are:

Linear, ReLU, and Sigmoid

Softmax, Sigmoid, and Step

Sigmoid, Tanh, and Leaky ReLU

116. To reduce overfitting in a deep neural network, you can:

Increase the number of layers and neurons

Use dropout, regularization, and early stopping

Train the model for more epochs without validation

117. Why should we use Batch Normalization in deep learning?

To stabilize and accelerate the training process

To increase the complexity of the model

To reduce the training time of the model

118. If your model is suffering from the problem of Exploding Gradients, what might happen during training?

Loss and gradients become very large

Loss and gradients become very small

The model converges quickly to the global minimum

119. A few hyperparameters used for training a neural network are:

Activation function, Dropout, Regularization strength

Learning rate, Batch size, and Number of epochs

Momentum, Learning rate, and Hidden layer sizes

120. Parameter sharing in deep learning refers to:

Sharing parameters between input and output layers

Sharing parameters between different types of layers

Sharing parameters between layers of the same type

121. The architecture of a typical Convolutional Neural Network (CNN) includes:

Convolutional, Pooling layers, Fully connected layers

Only fully connected layers

Only pooling layers

122. What is the Vanishing Gradient Problem in Artificial Neural Networks?

When gradients in backpropagation become too small

When gradients in backpropagation become too large

When the learning rate becomes very small

123. If the loss doesn't decrease in a few epochs while training an artificial neural network, it could be due to:

An issue with the data preprocessing or normalization

A large learning rate causing divergence

Overfitting on the training data

124. Why Sigmoid or Tanh is not preferred in the hidden layer of a neural network?

They are computationally expensive

They are not differentiable

They cause the vanishing gradient problem

125. Transfer learning is recommended when:

The source and target have different input domains

The source and target tasks are unrelated

The target task has a large labeled dataset

126. What is the vanishing gradient problem in RNN, and how can it be solved?

The gradients become too large during training

The model is unable to learn new patterns

The gradients become too small during training

127. The main gates in LSTM (Long Short-Term Memory) are:

- Input gate, Output gate, and Forget gate
- Input gate, Output gate, and Update gate
- Input gate, Forget gate, and Update gate

128. Using a CNN to classify 1D signals:

- Is not recommended, as CNNs are designed for image
- Is suitable only for binary classification tasks
- Is effective if the 1D signals represent sequential data

129. L1/L2 regularization affects a neural network by:

- Penalizing large weights to avoid overfitting
- Reducing the model size
- Increasing the learning rate during training

130. How to change a pre-trained neural network from classification to regression?

- Change the loss function
- Replace the activation function with a linear activation
- Add more hidden layers to the network

131. Setting the momentum hyperparameter too close to 1 (e.g., 0.9999) when using an SGD optimizer can lead to:

- Overshooting the global minimum and oscillations
- Faster convergence of the model
- Slower learning and longer training times

132. Hyperparameters that can be optimized for the batch normalization layer are:

- Learning rate and Dropout rate
- Scale factor and Center factor
- Activation function and Optimizer

133. Dropout affects the training and prediction speed of a deep learning model by:

- Decreasing training and increasing prediction speed
- Increasing training and decreasing prediction speed
- Having no impact on either

134. The advantage of deep learning over traditional machine learning is:

- The ability to learn hierarchical features from data
- Simplicity of model architecture
- Lower computational requirements

135. A depthwise Separable layer in a CNN:

- Is used for global feature extraction
- Combines depthwise and pointwise convolutions
- Performs pooling operations for downsampling

136. Siamese Network is beneficial in addressing the one-shot learning problem because it:

- Uses multiple GPUs for parallel processing
- Utilizes transfer learning to improve model
- Encodes similarity between inputs like shot learning

137. Grayscale serves the purpose of:

- Converting colored images to grayscale
- Reducing image dimensionality
- Enhancing the contrast of images

138. Translational equivariance means that a model:

- Produces same output for translated versions of input
- Can translate inputs to different languages
- Is invariant to translation invariance

139. YOLO (You Only Look Once) is an object detection algorithm that:

- Divide the image into grid and predicts bounding boxes
- Processes the image in multiple passes
- Uses a sliding window to detect objects in the image

140. Dropouts in neural networks are used for:

- Removing irrelevant features from the input data
- Preventing overfitting by deactivating neurons
- Reducing the learning rate during training

141. Exploding gradients occur when:

- The model fails to learn the training data
- The model underfits the training data
- The gradients during backpropagation become large

142. According to you, which one is more powerful — a two-layer neural network without any activation function or a two-layer decision tree?

- The two-layer neural network
- Both have similar power, but depends on the problem
- The two-layer decision tree

143. While building a neural network architecture, how will you decide how many neurons and hidden layers the neural network should have?

- Add as many neurons and hidden layers as possible
- Use a trial-and-error method with various architecture
- Choose the neurons based on size of the training data

144. In the context of deep learning, what is a tensor?

- A type of activation function
- A vector representation of the weights in a NN
- A multi-dimensional array used to store data

145. What are the advantages of using transfer learning in deep learning?

- Faster training and better generalization
- Improved interpretability of the model
- Lower computational requirements

146. What is the role of multi-class classification problems?

- Multi-class involves more than two classes
- Multi-class assigns multiple labels to a single instance
- None of the above

147. What are the different techniques to achieve data normalization?

- Min-Max scaling, Standardization, and L1 normalization
- Feature scaling and L2 normalization
- Z-score normalization and L1 regularization

148. What is forward propagation in the context of deep learning?

- Forward propagation updates the weights,
- Forward propagation gives the gradients during training
- None of the above

149. Name the different types of deep neural networks.

- Support Vector Machines and Logistic Regression
- K-Means Clustering and Principal Component Analysis
- Feedforward Neural Networks, RNN, CNN and GAN

150. What is the benefit of using ReLU activation over Sigmoid or Tanh?

- ReLU has better numerical stability
- ReLU produces a continuous output in the range $[0, \infty]$
- ReLU doesn't suffer from vanishing gradient problem

151. Explain the concept of depthwise Separable layers and their advantages in CNNs.

They combine depthwise and pointwise convolutions

Depthwise Separable layers apply pooling operations

Depthwise Separable layers utilize weight sharing

152. How can Siamese Network be beneficial in addressing the one-shot learning problem?

It uses multiple GPUs for parallel processing

It utilizes transfer learning to improve performance

Siamese Network encodes similarity between inputs

153. Grayscale an image serves the purpose of:

Converting colored images to grayscale

Reducing image dimensionality

Enhancing the contrast of images

154. Explain the concept of translational equivariance in deep learning

The model can change inputs to different languages

It is the same as invariance to translation invariance

It refers to the ability to recognize translated versions

155. YOLO (You Only Look Once) is an object detection algorithm that:

Processes the image in multiple passes

Divides the image into a grid

Uses a sliding window to detect objects in the image

156. What do you know about dropouts in neural networks?

Dropouts are used for removing irrelevant features

Dropouts reduce the learning rate during training

Dropouts prevent overfitting by deactivating neurons

157. Explain the concept of exploding gradients in deep learning.

Occur when gradients become too small

Occur when gradients in backpropagation become large

Same as vanishing gradient

158. What is Bias Measure?

Bias measures the ability to generalize to new data

Bias measures the model's sensitivity to training data

Bias and variance are the same concepts

159. Which one is more powerful — a two-layer neural network without any activation function or a two-layer decision tree?

The two-layer neural network

The two-layer decision tree

Both have similar power, but it depends on problem

160. While building a neural network architecture, how will you decide how many neurons and hidden layers the neural network should have?

Choose the number of neurons based on size of dataset

Add as many neurons and hidden layers as possible

Use a trial-and-error approach

161. In the context of deep learning, what is a tensor?

A multi-dimensional array used to store data

A type of activation function

A vector representation of the weights in a NN

162. What are the advantages of using transfer learning in deep learning?

Improved interpretability of the model

Faster training and better generalization

Lower computational requirements

163. What is NLG in NLP?

Natural Language Generation

Natural Language Graph

Natural Language Generator

164. What is the order of steps in natural language understanding?

Syntactic analysis, Signal analysis, Pragmatic analysis

Signal processing, Syntactic analysis, Semantic analysis

Pragmatic analysis, Signal analysis, Syntactic analysis

165. What is the purpose of signal processing in NLP?

To preprocess raw text data for analysis

To analyze sentiment in text

To convert audio signals to text

166. What does pragmatic analysis in NLP involve?

Analyzing the context and intention of the speaker

Extracting entities from text

Identifying the grammatical structure of sentences

167. What is the main focus of syntactic analysis in NLP?

Extracting named entities from text function

Analyzing the grammar and structure of sentences

Understanding the meaning of sentences

168. What is the goal of semantic analysis in NLP?

Determining the sentiment of a text

Identifying the subject and object in a sentence

Understanding the meaning and intent of sentences

169. What does sentiment analysis in NLP involve?

Analyzing the emotions expressed in a text

Detecting the grammatical errors in a sentence

Identifying the main topics in a text

170. What is discourse analysis in NLP?

Understanding the structure and flow of a conversation

Identifying coreference between pronouns & entities

Analyzing relationships between words in a sentence

171. What is pragmatic ambiguity in NLP?

Ambiguity related to context and speaker's intention

Ambiguity caused by multiple meanings of a word

Ambiguity due to errors in the text corpus

172. Which of the following is a major application of NLP?

Image recognition

Machine translation

Speech synthesis

173. Real-world applications of NLP include:

Object detection in images

Weather forecasting

Sentiment analysis of customer reviews

174. What are the common NLP techniques used for?

Image manipulation

Data visualization

Language understanding and text analysis

176. What are the main components of NLP?

Semantic analysis and Discourse analysis

Natural Language Understanding & NL Generation

Signal processing and Syntactic analysis

177. Which tools are commonly used for training NLP models?

TensorFlow and PyTorch

NumPy and SciPy

NLTK and spaCy

178. Which NLP technique uses a lexical knowledge base to obtain the correct base form of words?

WordNet-based normalization

Lemmatization

Stemming

179. Which models are commonly used to reduce the dimensionality of data in NLP?

SVM and Random Forest

LSA & LDA

Naive Bayes and Decision Trees

180. What does a masked language model do in NLP?

Fills in missing words in a sentence

Removes stop words from the text

Replaces ambiguous words with appropriate meaning

181. Which of the following is an open-source library for NLP?

Microsoft Word

Natural Language Toolkit (NLTK)

TensorFlow

182. What is the purpose of the bag of words model in NLP?

To represent data as frequency distribution of words

To represent the text data as a collection of words

To count the occurrence of each word in the text

183. What does CBOW stand for in NLP?

Continuous Bag of Words

Contextual Bag of Words

Comprehensive Bag of Words

184. What is TF-IDF used for in NLP?

- To represent text data as vectors
- To weigh the importance of words in a document
- To measure the similarity between documents

185. What do POS and tagging stand for in NLP?

- Position of Sentence and Text Analysis Grouping
- Part of Speech and Text Aggregation
- Part of Speech and Tagging

186. What is an n-gram in NLP?

- A combination of n words used for text analysis
- A type of language model
- A sequence of n words in a sentence

187. What is skip-gram in NLP?

- A word2vec model for predicting context words
- A technique to skip irrelevant words in a sentence
- A model used for skip-gram language learning

188. In NLP, what is a corpus?

- A collection of raw text documents used for analysis
- A statistical measure of word frequency
- A technique to convert speech to text

189. What are the features of the text corpus in NLP?

- File format, word length, and character encoding
- Word tokens and paragraph divisions
- POS tags, word frequency, and word embeddings

190. What is normalization used for in NLP?

- To bring data to a common format
- To standardize text documents for comparison
- To process speech data for analysis

191. What is keyword normalization in NLP?

- Converting keywords to lowercase for consistency
- Transforming keywords to their base forms
- Removing non-relevant words (stop words)

192. What is lemmatization in NLP?

- The process of replacing rare words with common
- The process of reducing a word to its base or root form
- The process of converting text into numerics

193. What is stemming in NLP?

- The process of finding the most relevant keywords
- The process of parsing sentences into grammar
- The process of grouping words with the same root

194. What does OpenCV stand for?

- Open Computer Vision
- Open Code Vision
- Optical Character Verification

195. What is OpenCV primarily used for?

- Image and video processing
- Audio processing
- Natural Language Processing

196. Which programming languages are officially supported by OpenCV?

- Python and C++
- Java and R
- C# and JavaScript

197. What is the role of OpenCV's "cv2" module in Python?

- Handling file I/O operations
- Image and video processing
- Mathematical computations

198. Which of the following data types is commonly used to represent images in OpenCV?

- Array of integers
- Dictionary of pixels
- Numpy array

199. What is the purpose of image thresholding in OpenCV?

- Converting an image to grayscale
- Applying a color map to the image
- Converting an image to binary

200. Which OpenCV function is used to detect edges in an image?

cv2.erode()

cv2.Canny()

cv2.dilate()

201. What is the "kernel" in OpenCV's image processing functions?

A part of an image

A function for color mapping

A small matrix used for convolution

202. What is histogram equalization used for in image processing?

Enhancing the contrast of an image

Removing noise from an image

Applying filters to an image

203. Which OpenCV function is used to apply a Gaussian blur to an image?

cv2.MedianBlur()

cv2.GaussianBlur()

cv2.boxFilter()

204. How does the Hough Line Transform work in OpenCV?

It detects straight lines in an image

It detects edges in an image

It detects circles in an image

205. Which OpenCV function is used to perform image resizing?

cv2.scale()

cv2.resize()

cv2.transform()

206. What is the role of the cv2.VideoCapture() function in OpenCV?

Capturing video from a camera or a file

Capturing images from a webcam

Recording audio from a microphone

207. What is the purpose of feature detection algorithms in OpenCV?

Detecting objects or keypoints in an image

Enhancing the color balance of an image

Removing noise from an image

208. Which OpenCV function is used for drawing shapes and text on an image?

cv2.drawContours()

cv2.rectangle()

cv2.putText()

209. What is the role of the cv2.matchTemplate() function in OpenCV?

Finding the similarity between two images

Calculating the histogram of an image

Template matching to locate objects in an image

210. How does the cv2.findContours() function work in OpenCV?

It finds and segments regions with similar intensity

It finds the edges of an image

It finds the minimum area bounding rectangle

211. What is the purpose of the cv2.CascadeClassifier() in OpenCV?

It classifies images into predefined categories

It detects faces or objects in an image or video

It performs morphological operations on images

212. What is the role of the cv2.goodFeaturesToTrack() function in OpenCV?

Detecting corners in an image

Tracking objects in a video stream

Calculating the optical flow between two frames

213. Which OpenCV function is used for image bitwise operations?

cv2.bitwise_and()

cv2.bitwise_xor()

cv2.bitwise_or()

214. What is the purpose of the cv2.SIFT() function in OpenCV?

Detecting and extracting SIFT keypoints from an image

Performing image rotation and scaling

Applying a Sobel filter to an image

215. Which OpenCV function is used to compute the perspective transformation of an image?

cv2.getPerspectiveTransform()

cv2.getAffineTransform()

cv2.getPerspective()

216. What is the role of the `cv2.warpAffine()` function in OpenCV?

- Applying an affine transformation to an image
- Resizing an image
- Converting an image to grayscale

217. What does the `cv2.fastNlMeansDenoising()` function do in OpenCV?

- Enhances the sharpness of an image
- Adjusts the brightness and contrast of an image
- Removes noise from an image

218. How does the `cv2.SimpleBlobDetector()` function work in OpenCV?

- It detects and extracts blobs or connected components
- It detects simple shapes like circles and squares
- It computes the center of mass of an object in an image

219. Which OpenCV function is used to apply a perspective transform to an image?

- `cv2.warpPerspective()`
- `cv2.warpAffine()`
- `cv2.perspectiveTransform()`

220. What is the purpose of the `cv2.inRange()` function in OpenCV?

- Creating a binary mask based on pixel intensity range
- Finding the range of pixel intensities in an image
- Detecting regions of interest in an image

221. How does the `cv2.SimpleBlobDetector_Params()` object help in OpenCV?

- It stores parameters for image rotation and scaling
- It stores parameters for the blob detection algorithm
- It stores parameters for image smoothing

222. What is the purpose of `cv2.getPerspectiveTransform()` in OpenCV?

- To find the perspective transformation matrix
- To perform an affine transformation on an image
- For perspective transformation matrix in image warping

223. How does the `cv2.createBackgroundSubtractorMOG2()` function work in OpenCV?

- It creates a background mask for object detection
- It creates a 3D model of the image foreground
- It creates a Gaussian Mixture-based background

224. What is the role of the `cv2.remap()` function in OpenCV?

- It remaps an image to a different color space
- It remaps an image from one coordinate to another
- It remaps an image using bilinear interpolation

225. Which OpenCV function is used to compute the optical flow of an image?

- `cv2.calcOpticalFlowFarneback()`
- `cv2.calcOpticalFlowSingleScale()`
- `cv2.calcOpticalFlowPyrLK()`

226. What is the purpose of the `cv2.HoughCircles()` function in OpenCV?

- Detecting circles in an image
- Detecting lines in an image
- Detecting corners in an image

227. How does the `cv2.KAZE()` feature detector work in OpenCV?

- It detects and extracts keypoints from an image
- It detects affine invariant keypoints in an image
- It detects edges in an image

228. What is the purpose of the `cv2.matchShapes()` function in OpenCV?

- It measures the similarity between two shapes
- It matches similar shapes in two images
- It finds the contour similarity between two images

229. What is the primary purpose of the Adam optimizer in deep learning?

- Regularization of the neural network
- Adaptive learning rate optimization
- Batch normalization of the input data

230. Which type of neural network architecture is best suited for sequential data and natural language processing tasks?

- Recurrent Neural Network (RNN)
- Feedforward Neural Network
- Convolutional Neural Network

231. Which technique is used to avoid the problem of exploding gradients in deep neural networks?

- Gradient clipping
- L1 regularization
- Dropout regularization

232. In a long short-term memory (LSTM) cell, which gate is responsible for deciding what information to forget from the previous time step?

Input gate

Forget gate

Output gate

233. In deep reinforcement learning, what is the function of the "Q-network"?

It calculates the gradients for updating the model

It determines the exploration rate during training

It maps states to actions and predicts future rewards

234. What is the "batch size" in stochastic gradient descent (SGD) optimization for neural networks?

Number of training samples before each weight update

The number of epochs during training

The learning rate used for weight updates

235. Which neural network architecture is designed to handle variable-length input sequences by applying the same weights at every time step?

Recurrent Neural Network

Feedforward Neural Network

Convolutional Neural Network

236. In deep reinforcement learning, what is the "policy network" responsible for?

It maps states to actions, specifying the agent's actions

It calculates discounted future rewards for each action

It determines the exploration rate during training

237. The technique of "weight initialization" in neural networks is crucial for:

Avoiding overfitting during training

Ensuring convergence during training

Preventing the vanishing gradient problem

238. What is the primary advantage of using batch normalization in deep neural networks?

It reduces the computation time during training

It increases the model's capacity

It improves the generalization of the model to new data

239. Which technique is used to reduce the dimensionality of the feature space in neural networks by projecting it into a lower-dimensional subspace?

Batch Normalization

Gradient Boosting

Principal Component Analysis (PCA)