Top 20 AI/ML Interview Questions with Answers for 6+ Months Experience

1. **Explain the difference between supervised and unsupervised learning with one real-world example for each.**
   Supervised learning uses labeled data to predict outcomes (e.g., spam email detection using labeled emails). Unsupervised learning finds patterns in unlabeled data (e.g., customer segmentation based on purchase behavior).
2. **How does a decision tree work, and how do you handle overfitting in it?**
   A decision tree splits data into branches based on feature values to make decisions. Overfitting is handled by pruning (removing branches), limiting tree depth, or setting a minimum number of samples per leaf.
3. **What is the bias-variance tradeoff, and how does it impact model performance?**
   Bias-variance tradeoff balances model complexity. High bias (underfitting) causes poor training performance; high variance (overfitting) causes poor generalization. Optimal models minimize both for better test performance.
4. **Describe how gradient descent optimizes a machine learning model.**
   Gradient descent iteratively updates model parameters by minimizing a loss function, moving in the direction of the steepest negative gradient using a learning rate.
5. **What are precision, recall, and F1-score? When would you prioritize one over the others?**
   Precision is true positives/(true positives + false positives); recall is true positives/(true positives + false negatives); F1-score is their harmonic mean. Prioritize precision for low false positives (e.g., spam detection), recall for low false negatives (e.g., disease detection).
6. **Explain the role of activation functions in neural networks. Compare ReLU and sigmoid.**
   Activation functions introduce non-linearity. ReLU (max(0,x)) is computationally efficient, avoids vanishing gradients, but can cause dead neurons. Sigmoid (1/(1+e^-x)) outputs 0-1, suitable for binary classification, but prone to vanishing gradients.
7. **How would you preprocess text data for an NLP task?**
   Tokenize text, convert to lowercase, remove stop words/punctuation, apply stemming/lemmatization, and encode (e.g., TF-IDF or word embeddings) for model input.
8. **What is the vanishing gradient problem, and how can it be mitigated?**
   Vanishing gradients occur when gradients become too small during backpropagation, slowing learning. Mitigate with ReLU activation, batch normalization, or architectures like LSTM/GRU.
9. **Describe the k-means clustering algorithm and its limitations.**

K-means assigns data points to k clusters by minimizing within-cluster variance. Limitations: assumes spherical clusters, sensitive to outliers, and requires predefined k.

10. **How do you select features for a machine learning model? Name two techniques.**
    Feature selection reduces irrelevant features. Techniques: filter methods (e.g., correlation analysis) and wrapper methods (e.g., recursive feature elimination).

11. **What is regularization, and how do L1 and L2 regularization differ?**
    Regularization adds a penalty to the loss function to prevent overfitting. L1 (Lasso) adds absolute weight penalties, promoting sparsity; L2 (Ridge) adds squared weight penalties, reducing weight magnitude.

12. **Explain how a convolutional neural network (CNN) processes image data.**
    CNNs use convolutional layers to extract features (e.g., edges), pooling layers to reduce dimensions, and fully connected layers for classification, leveraging spatial hierarchies in images.

13. **What is cross-validation, and why is k-fold cross-validation preferred?**
    Cross-validation evaluates model performance by splitting data into training/test sets. K-fold splits data into k subsets, training on k-1 and testing on 1, averaging results for robust performance estimation.

14. **How would you handle imbalanced datasets in a classification problem?**
    Use oversampling (e.g., SMOTE), undersampling, class-weighted loss functions, or ensemble methods to balance model focus on minority classes.

15. **Describe the architecture of a recurrent neural network (RNN) and its use cases.**
    RNNs process sequential data with loops, retaining memory of previous inputs. Use cases: time-series prediction, speech recognition. Variants like LSTM handle long-term dependencies.

16. **What is transfer learning, and when would you use it in a deep learning project?**
    Transfer learning uses pre-trained models (e.g., BERT, VGG) fine-tuned for specific tasks. Use when data is limited or training from scratch is computationally expensive.

17. **How do you evaluate the performance of a regression model? Name three metrics.**
    Metrics: Mean Squared Error (MSE) for average squared errors, Mean Absolute Error (MAE) for average absolute errors, and R-squared for variance explained.

18. **Explain the concept of word embeddings in NLP. How does Word2Vec work?**
    Word embeddings map words to dense vectors capturing semantic mean-

ing. Word2Vec uses neural networks (CBOW or skip-gram) to predict words from context or vice versa.

19. **What are hyperparameters, and how do you tune them (e.g., grid search vs. random search)?**
    Hyperparameters are model settings (e.g., learning rate). Grid search tests all combinations; random search samples randomly, often faster for large spaces.

20. **Describe a recent AI/ML project you worked on, including the problem, approach, and results.**
    Example: Built a churn prediction model using logistic regression on customer data. Preprocessed with one-hot encoding, used grid search for hyperparameter tuning, and achieved 85% accuracy on test data.