

Deep Learning Interview Questions

Beginner

1. What is Deep Learning, and how does it relate to Machine Learning?
2. What is an artificial neuron in a neural network?
3. Why do neural networks use activation functions like sigmoid, ReLU, or tanh?
4. What is overfitting in a machine learning model?
5. Explain the difference between Deep Learning and traditional Machine Learning, with the perspective of feature engineering.
6. What is a Softmax function and when is it used?
7. What are GPUs and TPUs, and why are they used in deep learning?
8. What is an epoch, a batch, and an iteration in model training?
9. Why do Neural Networks need a bias term?
10. What is a PyTorch Tensor and how does it differ from a NumPy array?

Intermediate

1. What is Autograd? Elaborate w.r.t Pytorch.
2. What is Gradient Descent and how does backpropagation work in training neural networks?
3. Compare and contrast Batch Gradient Descent, Stochastic Gradient Descent, and Mini-Batch Gradient Descent.
4. Explain the vanishing gradient problem and one common way to mitigate it.
5. What is the purpose of Batch Normalization in deep learning?
6. How do L1 and L2 regularization differ in their effect on model coefficients?
7. What are the key components of a Convolutional Neural Network (CNN)? Explain convolution, kernels, pooling, padding, and stride.
8. What is Data Augmentation and why is it used in CNN training?
9. What is Transfer Learning in the context of CNNs, and how are pre-trained models (ResNet, EfficientNet, MobileNet, etc.) used?
10. Explain RNNs (Recurrent Neural Networks) and how LSTM/GRU cells work.

Advanced

1. What is the Transformer architecture? Describe embeddings, positional encoding, attention, multi-head attention, encoder, and decoder.
2. Between ReLU and Sigmoid, which activation would you prefer for a hidden layer in a large MLP? Why?
3. You're building a model to detect fraudulent transactions. How would you handle imbalanced class labels during training?
4. You're tasked with building a real-time object detection system for autonomous drones. What trade-offs would you consider between accuracy and latency?
5. How does attention in transformers help capture long-range dependencies better than RNNs or CNNs in NLP? Give a use-case example.
6. Design a model pipeline to classify medical X-rays, and include steps for ensuring regulatory compliance, reproducibility, and fairness.
7. You're seeing overfitting despite using dropout and L2 regularization. What advanced techniques would you explore next?
8. Why is positional encoding essential in transformer architectures, and how does it affect context awareness in sequence models?
9. Describe how you'd evaluate the performance of a binary classification model in a high-stakes healthcare setting.
10. Suppose you're training a CNN on CIFAR-10 and your training accuracy improves, but test accuracy drops. What could be going wrong?