11–25: Machine Learning Algorithms

11. What is logistic regression?
Logistic regression is a classification algorithm used to predict binary outcomes. It models the probability of a class using the sigmoid function. The output is between 0 and 1. It's widely used for tasks like spam detection and medical diagnosis. Despite its name, it's a classifier, not a regressor.

12. What is the difference between classification and regression?
Classification predicts categorical labels, while regression predicts continuous values. For example, classifying emails as spam vs. not spam is classification. Predicting house prices is regression. Different metrics are used: accuracy/F1 for classification, MSE/RMSE for regression. Both are forms of supervised learning.

13. What is the bias-variance trade-off?
Bias is error from overly simplistic models; variance is error from overly complex models. High bias leads to underfitting, high variance to overfitting. The trade-off is about balancing both for optimal performance. Cross-validation helps manage this balance. Ensemble methods often improve this trade-off.

14. What is a confusion matrix?
A confusion matrix is a table used to evaluate classification performance. It includes true positives, false positives, true negatives, and false negatives. From it, metrics like precision, recall, and F1-score can be derived. It's especially useful for imbalanced datasets. It provides a full picture of classification results.

15. What is precision and recall?
Precision is the ratio of true positives to predicted positives. Recall is the ratio of true positives to actual positives. Precision answers "How many selected items are relevant?" Recall answers "How many relevant items are selected?" F1-score balances both. These are crucial in imbalanced classification.

16. What is cross-validation?
Cross-validation splits data into multiple train-test folds to evaluate model performance. The most common type is k-fold cross-validation. It reduces overfitting by using different data splits for training and validation. The model's average performance across folds gives a better generalization estimate. It's a robust evaluation technique.

17. What is regularization?
Regularization penalizes complex models to reduce overfitting. It adds a term to the loss function: L1 (Lasso) or L2 (Ridge). L1 encourages sparsity; L2 penalizes large weights. It controls model complexity and improves generalization. Choosing the right λ value is key.

18. What is a decision tree?
A decision tree is a flowchart-like structure for classification or regression. It splits data based on feature values to maximize information gain. Nodes represent decisions, and leaves represent outputs. It's easy to interpret but prone to overfitting. Pruning and ensemble methods like random forests help.

19. What is a random forest?
Random forest is an ensemble of decision trees trained on bootstrapped data. Each tree votes, and the majority class (or average) is the final output. It reduces overfitting and improves accuracy. It handles high-dimensional and missing data well. It's robust and widely used in practice.

20. What is gradient boosting?
Gradient boosting builds models sequentially, each correcting the previous one's errors. It uses decision trees as weak learners. Each new tree fits the residual errors. Algorithms include XGBoost, LightGBM, and CatBoost. It's powerful but sensitive to overfitting if not tuned properly.

21. What is k-nearest neighbors (KNN)?
KNN is a non-parametric algorithm that classifies based on the closest k neighbors. It uses a distance metric (e.g., Euclidean). It's simple and effective for small datasets. No training is needed, but inference can be slow. It's sensitive to the choice of k and feature scaling.

22. What is support vector machine (SVM)?
SVM finds the optimal hyperplane that separates classes with the maximum margin. It uses kernels to handle non-linear data. It's effective for high-dimensional spaces. It can be computationally intensive. It's used in text classification and image recognition.

23. What is PCA (Principal Component Analysis)?
PCA reduces dimensionality by projecting data onto principal components. These components maximize variance and are orthogonal. It helps in visualization, noise reduction, and speeding up training. PCA is unsupervised. The number of components is a tunable parameter.

24. What is a learning curve?
A learning curve plots training and validation performance over time or dataset size. It helps identify underfitting, overfitting, and whether more data would help. If training and validation errors converge, the model is well-generalized. It's a valuable diagnostic tool. It guides model improvement decisions.

25. What is ensemble learning?
Ensemble learning combines multiple models to improve performance. Techniques include bagging, boosting, and stacking. It reduces variance (bagging), bias (boosting), or leverages meta-models (stacking). Random forests and XGBoost are popular examples. Ensembles often outperform single models.

Would you like me to continue with **26–50: Deep Learning** next?

4o

You said:

yes

ChatGPT said:

Great! Here are **questions 26–50**, focusing on **Deep Learning**, with 5-line perfect answers:

26–50: Deep Learning

26. What is deep learning?
Deep learning is a subfield of machine learning based on neural networks with many layers. It automatically learns features from raw data using hierarchical representation. It excels in image, audio, and natural language processing tasks. Common architectures include CNNs, RNNs, and transformers. Training requires large datasets and high computational power.

27. What is a neural network?
A neural network is a model inspired by the human brain. It consists of layers of interconnected neurons, where each neuron applies an activation function to weighted inputs. The output is propagated forward, and errors are backpropagated for learning. It learns patterns in data. Neural networks are the foundation of deep learning.

28. What is forward propagation?
Forward propagation is the process of computing the output of a neural network. Inputs pass through each layer using learned weights and activation functions. The final output is used for prediction. Loss is computed based on this output. It's the first step before backpropagation.

29. What is backpropagation?
Backpropagation is the algorithm used to update neural network weights. It calculates the gradient of the loss function with respect to each weight. Using the chain rule, gradients flow from output to input layers. Optimizers use these gradients to update weights. It's key to model learning.

30. What are activation functions?
Activation functions introduce non-linearity into neural networks. Common types include ReLU, sigmoid, and tanh. ReLU is widely used due to its simplicity and effectiveness. Activation functions allow networks to learn complex functions. Without them, the model would behave like linear regression.

31. What is ReLU and why is it used?
ReLU (Rectified Linear Unit) is defined as `f(x) = max(0, x)`. It introduces sparsity and helps prevent vanishing gradients. It's computationally efficient and accelerates convergence. However, it can suffer from the "dying ReLU" problem. Variants like Leaky ReLU and ELU address this.

32. What is a convolutional neural network (CNN)?
CNNs are specialized for processing grid-like data, such as images. They use convolutional layers to extract features like edges, textures, and shapes. Pooling layers reduce spatial dimensions and control overfitting. CNNs are widely used in computer vision. Famous architectures include VGG, ResNet, and Inception.

33. What is a recurrent neural network (RNN)?

RNNs are designed for sequence data, with feedback loops in their architecture. They maintain a memory of previous inputs through hidden states. This enables modeling of temporal dependencies. However, they suffer from vanishing gradients. LSTMs and GRUs are improved variants.

34. What are LSTM and GRU?

LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) are RNN variants. They use gates to control the flow of information, addressing long-term dependency issues. LSTMs have input, forget, and output gates, while GRUs combine them. GRUs are faster, LSTMs are more expressive. Both are widely used in NLP and time series.

35. What is a transformer model?

Transformers rely on self-attention instead of recurrence. They model relationships between words regardless of distance. The architecture includes encoders and decoders with multi-head attention. It powers models like BERT, GPT, and T5. Transformers dominate modern NLP and vision tasks.

36. What is self-attention?

Self-attention computes the relevance of each element in a sequence to all others. It creates weighted representations using queries, keys, and values. This enables capturing contextual relationships. It's the backbone of transformers. Multi-head attention improves learning capacity.

37. What is transfer learning in deep learning?

Transfer learning uses a pre-trained model on a new, related task. It reduces training time and data requirements. Layers are reused, and only final layers may be fine-tuned. It's effective in vision and NLP tasks. Examples include using ResNet for medical imaging or BERT for sentiment analysis.

38. What is batch normalization?

Batch normalization normalizes layer inputs to have zero mean and unit variance. It stabilizes training, allows higher learning rates, and reduces dependence on initialization. It helps prevent internal covariate shift. It's applied after linear transformations and before activation. It often improves performance.

39. What is dropout?

Dropout is a regularization technique that randomly disables neurons during training. It prevents overfitting by ensuring the network doesn't rely too heavily on specific nodes. At inference, all neurons are used with scaled weights. Dropout is typically used in fully connected layers. It's simple and effective.

40. What is the vanishing gradient problem?

In deep networks, gradients can become very small during backpropagation. This makes early layers learn slowly or not at all. It's common with sigmoid or tanh activations. Solutions include ReLU, batch normalization, and residual connections. It was a major issue in early deep learning.

41. What is the exploding gradient problem?
Exploding gradients occur when gradients grow exponentially during backpropagation. It causes numerical instability and poor convergence. It's common in deep or recurrent networks. Gradient clipping is a common solution. Proper initialization and architecture choice also help.

42. What is an optimizer in deep learning?
An optimizer updates model weights to minimize loss. It uses gradients calculated by backpropagation. Common optimizers include SGD, Adam, RMSprop, and Adagrad. Adam is widely used due to its adaptive learning rate. The optimizer choice can greatly affect convergence.

43. What is the difference between SGD and Adam?
SGD updates weights using a fixed learning rate. Adam combines momentum and adaptive learning rates for each parameter. Adam often converges faster and handles sparse gradients better. SGD with momentum can generalize better in some cases. Choice depends on the task.

44. What is an epoch, batch, and iteration?
An epoch is one complete pass through the training dataset. A batch is a subset of the data used in one forward/backward pass. An iteration is one weight update step. If you have 1000 samples and batch size 100, one epoch has 10 iterations. These control the training process.

45. What is model capacity?
Model capacity refers to a model's ability to learn complex patterns. Low-capacity models underfit, high-capacity ones overfit. Capacity depends on architecture depth, width, and parameters. Regularization helps control excessive capacity. A balanced model generalizes well.

46. What is early stopping?
Early stopping halts training when validation performance stops improving. It prevents overfitting and reduces training time. It's a form of regularization. Training is monitored using a patience parameter. It's commonly used with deep networks.

47. What is a loss function?
A loss function quantifies how far predictions are from actual values. It guides the optimizer during training. Common types include MSE (for regression), cross-entropy (for classification), and hinge loss (for SVMs). The choice depends on the task. Lower loss implies better performance.

48. What is cross-entropy loss?
Cross-entropy loss measures the difference between predicted probabilities and actual class labels. It penalizes incorrect predictions with higher probability. It's widely used in classification tasks. It combines log loss and likelihood. It's ideal when outputs are probabilities.

49. What is the purpose of a softmax function?
Softmax converts logits into probabilities that sum to 1. It's used in the output layer for

multi-class classification. It highlights the most likely class while suppressing others. It enables interpretation of model confidence. The input with the highest softmax score is chosen.

50. What is a residual connection in deep learning?
A residual connection adds the input of a layer to its output, bypassing intermediate operations. It helps with training very deep networks by solving vanishing gradient issues. ResNets introduced this technique. It allows gradients to flow more easily. It enables deeper, more stable models.

Would you like me to continue with **51–75: Natural Language Processing (NLP)** next?

4o

You said:

yes

ChatGPT said:

Perfect! Here's the next set of **AI interview questions 51–75**, focused on **Natural Language Processing (NLP)** with crisp, 5-line answers:

51–75: Natural Language Processing (NLP)

51. What is Natural Language Processing (NLP)?
NLP is a field of AI that enables machines to understand, interpret, and generate human language. It combines linguistics with machine learning. Applications include sentiment analysis, translation, and chatbots. Key tasks are tokenization, parsing, POS tagging, and named entity recognition. NLP powers tools like Siri, GPT, and Google Translate.

52. What is tokenization in NLP?
Tokenization splits text into smaller units like words or subwords. It's the first step in many NLP pipelines. Word-based, character-based, and subword-based tokenization are common types. Modern models often use Byte Pair Encoding (BPE). Tokenization standardizes inputs for downstream models.

53. What is stemming and lemmatization?
Stemming reduces words to their base form by chopping off suffixes (e.g., "running" → "run"). Lemmatization maps words to their dictionary root (e.g., "better" → "good"). Lemmatization uses vocabulary and grammar, while stemming is rule-based and crude. Lemmatization is more accurate. Both help reduce vocabulary size.

54. What is part-of-speech (POS) tagging?
POS tagging labels words with their grammatical roles (e.g., noun, verb, adjective). It helps understand sentence structure and meaning. POS tags improve downstream tasks like

parsing and entity recognition. Algorithms include HMMs and CRFs. Transformers also perform POS tagging via sequence labeling.

55. What is named entity recognition (NER)?

NER identifies and classifies named entities in text, such as people, locations, and organizations. It helps extract structured data from unstructured text. Common applications are in legal, medical, and financial domains. Models label tokens with BIO (Begin-Inside-Outside) schemes. BERT-based models achieve strong NER performance.

56. What is word embedding?

Word embeddings are vector representations of words capturing semantic meaning. Similar words have similar embeddings. Popular methods include Word2Vec, GloVe, and FastText. Embeddings are dense and low-dimensional. They enable neural networks to process language numerically.

57. What is Word2Vec?

Word2Vec creates embeddings using a shallow neural network. Two main methods are CBOW (predict center word from context) and Skip-Gram (predict context from center word). It captures semantic and syntactic relationships. Trained on large corpora, it produces meaningful word vectors. "King - Man + Woman ≈ Queen" is a famous result.

58. What is GloVe?

GloVe (Global Vectors) uses matrix factorization of co-occurrence statistics to generate word vectors. It blends global matrix factorization and local context windowing. It captures semantic similarities well. Unlike Word2Vec, it uses entire corpus statistics. It’s trained using unsupervised learning.

59. What is the difference between Bag of Words and TF-IDF?

Bag of Words (BoW) counts word frequency in a document. TF-IDF weighs words by importance, reducing the effect of common terms. BoW is simpler but ignores importance. TF-IDF improves relevance in information retrieval. Both result in sparse, high-dimensional vectors.

60. What is cosine similarity?

Cosine similarity measures the angle between two vectors. It’s used to compare document or word embeddings. Values range from -1 (opposite) to 1 (identical). It’s scale-invariant, unlike Euclidean distance. Widely used in recommendation and NLP tasks.

61. What is BLEU score?

BLEU (Bilingual Evaluation Understudy) is used to evaluate machine translation. It measures overlap between machine-generated and reference translations. Precision-based, it penalizes short outputs (brevity penalty). Scores range from 0 to 1. Higher BLEU indicates better translation quality.

62. What is perplexity in NLP?

Perplexity measures how well a language model predicts a sample. Lower perplexity indicates better performance. It is the exponentiated average negative log-likelihood of the words. High perplexity suggests the model is uncertain. It’s used in evaluating models like GPT.

63. What is a language model?
A language model predicts the next word in a sequence. It captures grammar, context, and semantics. Types include statistical (n-grams) and neural (RNNs, Transformers). Pre-trained LMs like BERT and GPT revolutionized NLP. They enable zero-shot and few-shot learning.

64. What is BERT?
BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based model trained on masked language modeling and next-sentence prediction. It captures bidirectional context. It's fine-tuned for tasks like QA, NER, and classification. BERT uses only the encoder stack. It sparked the transfer learning revolution in NLP.

65. What is GPT?
GPT (Generative Pre-trained Transformer) is a unidirectional transformer trained with causal language modeling. It generates coherent text and excels at few-shot learning. GPT-3 and GPT-4 support multi-task learning via prompts. GPT uses only the decoder stack. It powers tools like ChatGPT.

66. What is the difference between BERT and GPT?
BERT is bidirectional and optimized for understanding tasks. GPT is unidirectional and optimized for text generation. BERT uses masked tokens; GPT uses autoregression. BERT performs better in classification; GPT in generation. Both use transformer architectures but with different goals.

67. What is fine-tuning in NLP?
Fine-tuning adapts a pre-trained model to a specific task. The base layers are retained, and task-specific layers are trained. It requires less data than training from scratch. Examples include fine-tuning BERT for sentiment analysis. It's central to modern NLP.

68. What is a sequence-to-sequence model?
Seq2Seq models map an input sequence to an output sequence. They're used in translation, summarization, and question-answering. It consists of an encoder and a decoder, often with attention. RNNs were used initially; now transformers dominate. Seq2Seq enables variable-length input/output.

69. What is attention in NLP?
Attention lets models focus on relevant parts of the input sequence when generating output. It calculates weights between encoder and decoder tokens. Improves performance in translation and summarization. Scaled dot-product attention is common in transformers. It enables long-range dependency modeling.

70. What is masked language modeling?
Masked language modeling randomly masks words in input text and trains the model to predict them. Used by BERT during pre-training. It allows learning bidirectional context. Unlike autoregressive models, it doesn't predict the next word. It's unsupervised and powerful.

71. What is next sentence prediction?
Next sentence prediction trains a model to determine whether one sentence logically follows another. Used in BERT's pretraining. It improves understanding of sentence-level

relationships. It enables downstream tasks like QA and sentence similarity. It's binary classification (is_next or not_next).

72. What is zero-shot learning in NLP?
Zero-shot learning enables models to perform tasks without task-specific training. It uses general knowledge and prompt engineering. GPT models perform well in this setup. Useful for scaling to many tasks. It requires strong language understanding and generalization.

73. What is prompt engineering?
Prompt engineering is the art of crafting inputs to steer language models. Good prompts can improve task performance without fine-tuning. It includes few-shot, zero-shot, and chain-of-thought prompting. It's essential in LLM deployment. It bridges model capabilities and user intent.

74. What is chain-of-thought prompting?
Chain-of-thought prompting asks the model to reason step by step before answering. It improves accuracy on logic and math problems. Example: "Let's think step-by-step." Used with models like GPT-4. It mimics human reasoning and improves interpretability.

75. What is retrieval-augmented generation (RAG)?
RAG combines language generation with information retrieval. It retrieves relevant documents from a knowledge base and conditions generation on them. It improves factual correctness and reduces hallucination. Useful in QA, chatbots, and search. It integrates search engines with language models.

Would you like me to continue with **76–100: Computer Vision & AI System Design**?

4o

You said:
yes

ChatGPT said:

Excellent! Here's the final set of **AI interview questions 76–100**, covering **Computer Vision and AI System Design**, with 5-line expert-level answers:

76–87: Computer Vision

76. What is computer vision?
Computer vision is a field of AI that enables machines to interpret and understand visual data. It includes tasks like image classification, object detection, segmentation, and tracking. It mimics human visual perception. Applications range from medical imaging to autonomous vehicles. Deep learning has revolutionized this field.

77. What is image classification?
Image classification assigns a label to an image based on its content. A model processes the image and predicts a class. CNNs are typically used due to their spatial feature learning. Datasets like CIFAR-10 and ImageNet are standard benchmarks. Accuracy and confusion matrix evaluate performance.

78. What is object detection?
Object detection locates and classifies objects in an image. It returns bounding boxes and class labels. Common models include YOLO, SSD, and Faster R-CNN. It's used in surveillance, self-driving cars, and AR. Evaluation metrics include mAP (mean Average Precision).

79. What is semantic segmentation?
Semantic segmentation classifies each pixel in an image into a predefined class. It provides detailed understanding of image regions. Models like U-Net and DeepLab are commonly used. It's used in medical imaging and autonomous navigation. Output is a labeled mask.

80. What is instance segmentation?
Instance segmentation identifies individual object instances and classifies each pixel. It combines object detection and semantic segmentation. Mask R-CNN is a leading model. Each object is uniquely labeled. It's critical in applications like robotics and scene understanding.

81. What is edge detection?
Edge detection finds boundaries in images by detecting intensity changes. Algorithms like Canny, Sobel, and Laplacian are used. It's often a preprocessing step in vision tasks. Helps in detecting shapes and contours. Deep models learn edge features implicitly.

82. What is image augmentation?
Image augmentation generates new training samples by applying transformations. Techniques include rotation, flipping, zooming, and noise addition. It helps prevent overfitting and improves generalization. It's essential in deep learning pipelines. Libraries like Albumentations and torchvision provide tools.

83. What is the role of CNN filters?
Filters (or kernels) slide over an image to extract local patterns like edges, textures, or shapes. Each filter learns a specific feature map. Lower layers detect simple patterns; deeper layers learn complex features. Filters are learned via backpropagation. They enable spatial feature extraction.

84. What is transfer learning in vision?
Transfer learning reuses a pre-trained CNN (e.g., ResNet, VGG) for a new vision task. Base layers are frozen or fine-tuned. It reduces training time and data needs. It's common in medical, satellite, and industrial vision applications. Fine-tuning the last few layers customizes the model.

85. What is ResNet?
ResNet (Residual Network) uses skip connections to address vanishing gradients. It enables training of very deep networks (e.g., ResNet-50, ResNet-101). It won the ImageNet

2015 challenge. The residual block adds identity mappings to ease optimization. ResNet remains a foundational model in CV.

86. What is YOLO in object detection?
YOLO (You Only Look Once) is a real-time object detection model. It predicts bounding boxes and class probabilities in a single pass. YOLO is fast and accurate, making it suitable for real-time applications. New versions (YOLOv4, YOLOv8) improve speed and accuracy. It trades off slightly on precision for speed.

87. What is mAP in object detection?
Mean Average Precision (mAP) is the standard metric for object detection. It measures precision across recall levels for each class and averages them. Higher mAP indicates better performance. It considers true/false positives based on IoU thresholds. It's crucial for comparing detection models.

88–100: AI System Design & Advanced Topics

88. What is model deployment?
Model deployment means integrating a trained model into a production environment. It includes packaging, serving, and monitoring the model. Tools like Docker, FastAPI, and Kubernetes are used. Deployment can be on-cloud, edge, or mobile. Monitoring performance drift is essential post-deployment.

89. What is model drift?
Model drift happens when the model's performance degrades over time due to changing data. Types include concept drift and data drift. It affects accuracy and reliability. Detecting drift requires monitoring predictions and retraining periodically. Tools like Evidently AI help in detection.

90. What is MLOps?
MLOps combines ML development and operations for scalable model lifecycle management. It covers versioning, reproducibility, deployment, and monitoring. Tools include MLflow, Kubeflow, and DVC. MLOps enables CI/CD for ML. It bridges the gap between data science and DevOps.

91. What is MLflow?
MLflow is an open-source platform for managing ML experiments. It offers tracking, projects, models, and registry components. It enables reproducibility and collaboration. It integrates with tools like PyTorch, TensorFlow, and scikit-learn. It simplifies the ML lifecycle from training to deployment.

92. What is an AI pipeline?
An AI pipeline is a sequence of stages from data ingestion to model deployment. It includes preprocessing, training, evaluation, and serving. Pipelines ensure modularity, reusability, and automation. Tools like Airflow, Kubeflow Pipelines, and Prefect are used. Pipelines are key to production-ready AI.

93. What is AutoML?

AutoML automates the process of model selection, hyperparameter tuning, and feature engineering. It democratizes ML by reducing manual effort. Tools include Google AutoML, H2O.ai, and AutoGluon. It's ideal for non-experts or rapid prototyping. Trade-off: less control but faster iteration.

94. What is explainable AI (XAI)?

XAI makes model predictions transparent and interpretable. Techniques include LIME, SHAP, and integrated gradients. It's critical in sensitive domains like healthcare and finance. It builds trust, ensures compliance, and uncovers bias. XAI balances performance with interpretability.

95. What is differential privacy in AI?

Differential privacy ensures that the inclusion or exclusion of a single data point doesn't significantly affect the model's output. It provides formal privacy guarantees. Techniques include noise addition and DP-SGD. Used in federated learning and sensitive data modeling. It protects user data during training.

96. What is federated learning?

Federated learning trains models across multiple decentralized devices without centralizing data. It enhances privacy and data locality. Devices share model updates, not raw data. Aggregation is done at a server. Applications include keyboard prediction and medical diagnostics.

97. What is reinforcement learning in real-world systems?

Reinforcement learning is used in recommendation systems, robotics, and finance. Agents interact with environments to learn optimal actions. Requires defining states, actions, and rewards. Challenges include sample efficiency and safety. It's powerful for dynamic decision-making.

98. What is multi-modal AI?

Multi-modal AI processes multiple input types (text, image, audio) together. It enables richer understanding and interaction. Models like CLIP and GPT-4V integrate multiple modalities. Applications include medical diagnosis and interactive assistants. It enhances robustness and context-awareness.

99. How would you handle class imbalance in training?

Use techniques like resampling (oversampling minority, undersampling majority), SMOTE, or class weights. Choose metrics like F1 or AUC-ROC instead of accuracy. Use data augmentation for rare classes. Ensemble methods often handle imbalance better. Evaluation should be thorough.

100. What are ethical concerns in AI?

Key concerns include bias, fairness, transparency, and misuse. Models may amplify discrimination or violate privacy. Addressing these involves auditing data, using XAI, and fairness constraints. Regulatory compliance like GDPR is essential. Responsible AI is a shared responsibility.