

Pergunta: Qual dos seguintes não é um tipo de Learning Rate Scheduler?

Resposta: ➔ Aumento Exponencial (não é usado, pois instabiliza o treino).

Pergunta: No contexto de Processamento de Linguagem Natural (PLN), qual das seguintes opções é uma aplicação de Deep Learning?

Resposta: ➔ Tradução automática.

Pergunta: O underfitting ocorre quando um modelo é complexo demais, aprendendo não apenas os padrões subjacentes dos dados, mas também o ruído presente neles.

Resposta: ➔ Falso. Isso descreve overfitting; underfitting ocorre quando o modelo é simples demais.

Pergunta: As Redes Adversariais Generativas (GANs) operam em um modelo de cooperação entre o gerador e o discriminador?

Resposta: ➔ Falso. Elas operam em competição.

Pergunta: Como o IDF ajusta o peso das palavras no cálculo do TF-IDF?

Resposta: ➔ Diminuindo o peso de palavras muito frequentes no corpus.

Pergunta: A regularização pode ser implementada de diversas formas... Verdadeiro ou Falso?

Resposta: ➔ Verdadeiro. Inclui L1/L2, Dropout, Early Stopping.

Pergunta: Qual das seguintes não é uma aplicação de TF-IDF?

Resposta: ➔ Detecção de spam em e-mails (não é aplicação típica).

Pergunta: Qual das seguintes opções NÃO é uma aplicação comum para funções de perda?

Resposta: ➔ Aumentar deliberadamente o erro (não faz parte do uso).

Pergunta: Qual dos seguintes não é um otimizador em redes neurais?

Resposta: ➔ Backtracking Line Search.

Pergunta: A taxa de aprendizagem determina o tamanho do passo na direção do gradiente... Verdadeiro ou Falso?

Resposta: ➔ Verdadeiro.

Pergunta: O processo de otimização em Deep Learning utiliza exclusivamente o Gradiente Descendente e variantes... Verdadeiro ou Falso?

Resposta: ➔ Falso. Há outros métodos também.

Pergunta: Qual é a principal limitação da vetorização com TF-IDF?

Resposta: ➔ Não capta ordem das palavras nem semântica contextual.

Pergunta: Quais são as principais desvantagens da Sigmoid e Tanh?

Resposta: ➔ Causam desaparecimento do gradiente em redes profundas.

Pergunta: Para problemas de regressão, quais funções de perda são mais comuns?

Resposta: ➔ MSE e MAE.

Pergunta: Qual é o propósito de um Learning Rate Scheduler?

Resposta: ➔ Ajustar a taxa de aprendizado ao longo do tempo, geralmente diminuindo-a.

Pergunta: Dropout e Early Stopping são exclusivos do Deep Learning?

Resposta: ➔ Falso. Podem ser usados em outros modelos.

Pergunta: Durante backpropagation, calcula-se o gradiente antes do erro?

Resposta: ➔ Falso. Primeiro calcula-se o erro, depois os gradientes.

Pergunta: Dropout e BatchNorm podem ser usados juntos?

Resposta: →■ Verdadeiro. São complementares.

Pergunta: O treinamento com backpropagation é feito em mini-lotes para estabilidade?

Resposta: →■ Verdadeiro.

Pergunta: A regularização pode aumentar o overfitting?

Resposta: →■ Falso. Ela previne overfitting.

Pergunta: Modelos autorregressivos como PixelRNN geram dados sequencialmente?

Resposta: →■ Verdadeiro.

Pergunta: Qual a importância do Deep Learning para veículos autônomos?

Resposta: →■ Reconhecimento de objetos, pessoas e sinais para navegação segura.

Pergunta: O que faz o processo de otimização após o backpropagation?

Resposta: →■ Ajusta os pesos na direção oposta ao gradiente.

Pergunta: Qual o papel das funções de ativação?

Resposta: →■ Introduzir não-linearidade para modelar relações complexas.

Pergunta: Qual a principal função do backpropagation?

Resposta: →■ Calcular gradiente da perda em relação a cada peso.

Pergunta: A função de perda apenas mede o erro, sem considerar generalização. Verdadeiro ou Falso?

Resposta: →■ Verdadeiro.

Pergunta: Batch Normalization ajuda contra internal covariate shift?

Resposta: →■ Verdadeiro.

Pergunta: VAEs geram dados sem espaço latente?

Resposta: →■ Falso. Dependem de espaço latente.

Pergunta: Função de ativação com saída entre 0 e ∞ é mais usada?

Resposta: →■ ReLU.

Pergunta: Qual o papel da função de perda?

Resposta: →■ Medir quanto bem previsões se alinharam com valores reais, visando minimização.

Pergunta: No attention do Transformer, Q, K, V vêm de onde?

Resposta: →■ Todos vêm da mesma entrada, via projeções lineares distintas.

Pergunta: Função de ativação mais usada em camadas ocultas para mitigar vanishing gradient?

Resposta: →■ ReLU.

Pergunta: Por que diminuir a learning rate ao longo do tempo?

Resposta: →■ Para ajustes mais sutis e reduzir overfitting.

Pergunta: Por que a não-linearidade é crucial?

Resposta: →■ Permite modelar relações complexas e não-lineares.

Pergunta: No mecanismo de atenção, como usar pontuações QK?

Resposta: →■ Ponderar os valores, criando combinação ponderada da entrada.

Pergunta: Problemas introduzidos por funções de ativação?

Resposta: →■ Vanishing ou exploding gradients.

Pergunta: No mecanismo de atenção, qual componente representa o foco da informação?
Resposta: ➔■ Query (Q).

Pergunta: Qual não é aplicação de visão computacional?
Resposta: ➔■ Previsão do tempo.

Pergunta: É possível usar GPU em Deep Learning?
Resposta: ➔■ Verdadeiro. GPUs são recomendadas.

Pergunta: Importância de versões de pacotes?
Resposta: ➔■ Verdadeiro. Fundamental para reproduzibilidade.

Pergunta: RMSprop usa média móvel de gradientes quadrados?
Resposta: ➔■ Verdadeiro.

Pergunta: ResNet é usada em tarefas de visão?
Resposta: ➔■ Verdadeiro.

Pergunta: Podemos usar só Lightning sem PyTorch?
Resposta: ➔■ Falso. Lightning depende de PyTorch.

Pergunta: Existem várias arquiteturas de GNNs (GCN, GraphSAGE, GAT, GGNN)?
Resposta: ➔■ Verdadeiro.

Pergunta: Normalização é parte do pré-processamento de imagens?
Resposta: ➔■ Verdadeiro.

Pergunta: O parâmetro epsilon define critério de parada?
Resposta: ➔■ Verdadeiro.

Pergunta: PyTorch pode ser usado só com GPU?
Resposta: ➔■ Falso. Também funciona em CPU.

Pergunta: Gradient clipping previne gradientes explosivos?
Resposta: ➔■ Verdadeiro.

Pergunta: Um tensor é um objeto geométrico arbitrariamente...?
Resposta: ➔■ Complexo.

Pergunta: Pesos de atenção não são aprendidos?
Resposta: ➔■ Falso. São aprendidos.

Pergunta: Lightning também só roda em GPU?
Resposta: ➔■ Falso. Roda em CPU e GPU.

Pergunta: Grafos modelam redes sociais, transporte, etc.?
Resposta: ➔■ Verdadeiro.

Pergunta: O DataLoader prepara dados em PyTorch?
Resposta: ➔■ Verdadeiro.

Pergunta: O modelo é treinado para minimizar uma função de...?
Resposta: ➔■ Custo.

Pergunta: Durante inferência não precisamos de gradientes?
Resposta: ➔■ Verdadeiro.

Pergunta: ResNet usa blocos residuais?

Resposta: ➔■ Verdadeiro.

Pergunta: Distância e diâmetro de um grafo?

Resposta: ➔■ Verdadeiro.

Pergunta: O _____ de um vértice é número de arestas incidentes?

Resposta: ➔■ Grau.

Pergunta: GNNs trabalham com dados representados por...?

Resposta: ➔■ Grafos.

Pergunta: Um grafo é denotado por...?

Resposta: ➔■ $G = (V, E)$.

Pergunta: Estamos escrevendo a história da evolução do Deep Learning?

Resposta: ➔■ Verdadeiro.

Pergunta: Grafos = nós + arestas?

Resposta: ➔■ Verdadeiro.

Pergunta: GAT usa atenção para ponderar vizinhos?

Resposta: ➔■ Verdadeiro.

Pergunta: Criamos DataLoader para treino e teste?

Resposta: ➔■ Verdadeiro.

Pergunta: CNN é inspirada no córtex visual?

Resposta: ➔■ Verdadeiro.