

# Deep Learning for Natural Language Processing

1) Respostas curtas:

- a) Suponha que você tenha uma rede neural que está operando em overfitting em relação aos dados de treinamento. Por que essa não é uma situação desejável? Descreva duas maneiras de corrigir esta situação.
  
  
  
  
  
  
- b) Por que o skip-gram com amostragem negativa (negative sampling) é mais rápido de treinar do que o modelo skip-gram original (sem negative sampling)?
  
  
  
  
  
  
- c) Suponha que a função de perda de uma arquitetura neuronal esteja sendo avaliada de maneira que a perda aumenta com o número de épocas. Sugira duas opções para contornar essa situação. Justifique as sugestões.
  
  
  
  
  
  
- d) Vimos que vetores densos de palavra têm muitas vantagens sobre o uso de vetores esparsos, como one-hot-encoding. Qual das alternativas a seguir NÃO é uma vantagem que os vetores densos têm sobre os vetores esparsos?
  - i) Modelos que usam vetores densos de palavras generalizam melhor para palavras novas do que aqueles que usam vetores esparsos.
  - ii) Modelos que usam vetores densos de palavras generalizam melhor para palavras raras do que aqueles que usam vetores esparsos.
  - iii) Vetores densos de palavras codificam semelhanças entre as palavras, enquanto os vetores esparsos não.

- e) Um modelo de rede neural multi-camadas treinado usando-se descida de gradiente no mesmo conjunto de dados, porém com inicializações diferentes para seus parâmetros, irá garantidamente chegar aos mesmos parâmetros ao fim do treinamento? Justifique sua resposta.
- f) Qual das afirmações a seguir sobre Skip-gram está correta?
- i) Ele prevê a palavra central a partir das palavras do contexto circundante.
  - ii) O vetor final para uma palavra é a média ou soma do vetor de entrada  $v$  e vetor de saída  $u$  correspondendo a essa palavra.
  - iii) Faz uso de estatísticas globais de co-ocorrência.
  - iv) Nenhuma dessas.
- 2) Discuta os problemas/limitações inerentes a arquiteturas sequence-to-sequence e como a arquitetura Transformers lida com essas limitações.

- 3) A tarefa de “Reconhecimento de Entidades Nomeadas” consiste em classificar nomes em categorias pré-estabelecidas, como Pessoas, Localidades etc. Proponha uma solução completa para o problema, incluindo o que espera-se de dados de treinamento. A solução deve usar uma arquitetura sequence-to-sequence.