

Processamento de Linguagem Natural

Prova 1

Aluno(a): _____

1. Cite três vantagens da representação distribuída para palavras, ao invés da representação *one-hot*.
2. Sobre o *Skip-Gram*, marque as alternativas corretas.
 - a. O algoritmo prediz a palavra central a partir das palavras que formam o contexto.
 - b. O vetor final é dado pela média dos vetores de entrada.
 - c. Seu desempenho é pior do que o algoritmo CBOW, quando o corpus é relativamente pequeno.

3. Suponha que você queira classificar comentários sobre filmes em positivos e negativos. Proponha um algoritmo para realizar essa tarefa. Explique suas escolhas em termos de evitar overfitting e justifique que essas escolhas irão levar a bons resultados.

4. Suponha que você produziu, com o algoritmo Skip-Gram, vetores semânticos de palavras utilizando textos de artigos do Wikipedia. Agora você tem uma tarefa específica, para a qual você tem um pequeno corpus, e você se depara com a seguinte questão:
 - a. Utilizar os vetores da forma como eles estão.
 - b. Re-treinar os vetores no corpus específico, mas ao invés de iniciar os vetores aleatoriamente, usa-se os vetores pré-treinados.

Qual a escolha correta? Justifique.