

**D S T Q Q S S**

- vantagem n pequeno: maior controle sobre crescimento exponencial

6) Perplexidade serve para medir o quão bom é um modelo probabilístico. Uma baixa perplexidade indica que a distribuição de probabilidade é boa para prever a amostra. A perplexidade é a probabilidade inversa do conjunto de teste, normalizada pelo número de palavras.

$$\text{Perplexidade de } P(W) = P(W_1, \dots, W_n)^{-\frac{1}{N}}$$

$$= N \sqrt[N]{\frac{1}{P(W_1, \dots, W_n)}}$$

$$= N \sqrt[N]{\prod_{i=1}^N \frac{1}{P(W_i | W_1, \dots, W_{i-1})}}$$

7) Consiste em escolher aleatoriamente 2 bigramas de acordo com a sua probabilidade sucessivamente e depois juntá-los.

8) A suavização de Laplace também conhecida como Add-one estimation consiste em add um(1) a cada contagem de palavras com o intuito de resolver o problema de frequência de uma palavra ser igual a zero(0) quando esta não aparece no conjunto de treino e aparece no conjunto de teste, o que faz com que apareça zero(0) na probabilidade. Assim, não é possível calcular a perplexidade. A suavização de Laplace é usada em NLP modelos para classificação de texto no domínio onde o nº de zeros(0) não é tão grande.