

1. Estrategia Tf-IDF

- Se utiliza en procesamiento de texto que asigna pesos según su importancia en un texto.

- Cálculo

$$Tf(t, d) = \frac{\text{Num. que aparece } t \text{ en } d}{\text{Num. términos en } d}$$

$$IDF(t, D) = \log \left(\frac{\text{Num. total de documentos en corpus}}{\text{Num. documentos que contienen } t} \right)$$

$$Tf-IDF = Tf(t, d) \cdot IDF(t, D)$$

- Cuando es efectivo

Es útil en tareas donde se quiere dar un peso a términos que son importantes para un documento pero que no son muy comunes en todo el texto.

2. Laplace smoothing

- Resuelve el problema de la probabilidad de 0 asignando una probabilidad mínima positiva a todos los N-grams posibles.

- Como trabaja

Añade 1 a la frecuencia de cada N-gram durante el entrenamiento para evitar la posibilidad de 0.

- Impacto Modelo NLP

Mejora la generalización del modelo al manejar mejor N-grams no vistos durante el entrenamiento.

3: ¿Que pasa cuando una palabra no se encuentra en el vocabulario del modelo de los N-gram?

- se considera una palabra out-of-vocabulary.

Modelación OOV

- Se asigna una probabilidad mínima a las palabras utilizando Laplace smoothing, o asignar una Probabilidad Uniforme a todas las palabras.
- se implementa ajustando los conteos de palabras durante el entrenamiento para incluir un valor a las palabras