Murillo Blancas Sergio G

# MLE Modelo de estimación de máxima verosimilitud

## Objetivo

Que el estudiante comprenda y aplique el cálculo de **probabilidades condicionales** en modelos de lenguaje basados en n-gramas, utilizando el enfoque de **Estimación de Máxima Verosimilitud (MLE)**, considerando la incorporación de **fronteras de oración** para un modelo más realista.

## 1. Corpus tokenizado

Oraciones originales:

el gato duerme en la cama

el perro corre en el parque

Tokenización por oración:

['el', 'gato', 'duerme', 'en', 'la', 'cama']

['el', 'perro', 'corre', 'en', 'el', 'parque']

## 2. Tabla de frecuencias y probabilidades condicionales (sin fronteras)

| w_{i-1} | w_i | C(w_{i-1},w_i) | C(w_{i-1}) | P(w_i\|w_{i-1}) |
|---|---|---|---|---|
| el | gato | 1 | 3 | 0.3333333333333333 |
| gato | duerme | 1 | 1 | 1.0 |
| duerme | en | 1 | 1 | 1.0 |
| en | la | 1 | 2 | 0.5 |
| la | cama | 1 | 1 | 1.0 |
| el | perro | 1 | 3 | 0.3333333333333333 |
| perro | corre | 1 | 1 | 1.0 |
| corre | en | 1 | 1 | 1.0 |
| en | el | 1 | 2 | 0.5 |

| | | | | |
|---|---|---|---|---|
| el | parque | 1 | 3 | 0.3333333333333333 |

## 3. Tabla de frecuencias y probabilidades condicionales (con fronteras <s> y </s>)

| w_{i-1} | w_i | C(w_{i-1},w_i) | C(w_{i-1}) | P(w_i\|w_{i-1}) |
|---|---|---|---|---|
| <s> | el | 2 | 2 | 1.0 |
| el | gato | 1 | 3 | 0.3333333333333333 |
| gato | duerme | 1 | 1 | 1.0 |
| duerme | en | 1 | 1 | 1.0 |
| en | la | 1 | 2 | 0.5 |
| la | cama | 1 | 1 | 1.0 |
| cama | </s> | 1 | 1 | 1.0 |
| el | perro | 1 | 3 | 0.3333333333333333 |
| perro | corre | 1 | 1 | 1.0 |
| corre | en | 1 | 1 | 1.0 |
| en | el | 1 | 2 | 0.5 |
| el | parque | 1 | 3 | 0.3333333333333333 |
| parque | </s> | 1 | 1 | 1.0 |

## 4. Conclusiones

Incluir las fronteras de oración <s> y </s> en un modelo de n-gramas permite que el sistema aprenda explícitamente cómo empiezan y terminan las oraciones. Esto ayuda a mejorar las predicciones, porque evita mezclar palabras del final de una oración con las del inicio de otra, y permite que el modelo aprenda patrones típicos de apertura o cierre. En otras palabras, hace el modelo más realista al reflejar los límites naturales de las oraciones.