

Cuantificación: Estimación de Prevalencias por Clase

Taller práctico

[Pablo González, Olaya Pérez-Mon]

[Universidad de Oviedo]

10 de julio de 2025



Universidad de Oviedo



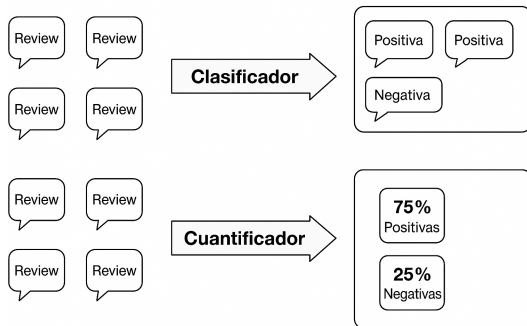
Objetivos del taller

- Poner en práctica los conceptos vistos en la sesión teórica:
Cuantificación, estimación de prevalencias por clase mediante aprendizaje supervisado.
- Entender cuando necesitamos la cuantificación desde un enfoque práctico.
- Introducir herramientas para implementar cuantificadores de forma sencilla (quantificationlib).
- Aprender a entrenar cuantificadores y evaluarlos correctamente.



¿Qué es la cuantificación?

- En lugar de predecir etiquetas individuales, queremos estimar **la proporción de ejemplos por clase** en una **bag** de ejemplos no etiquetada.
- Ejemplo: estimar qué porcentaje de opiniones sobre un producto en Amazon son positivas, sin etiquetarlas una a una.
- Otras aplicaciones: control de calidad, análisis de opinión, medicina, biología, etc.



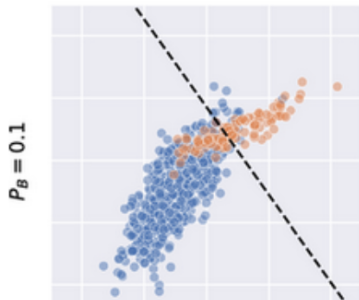
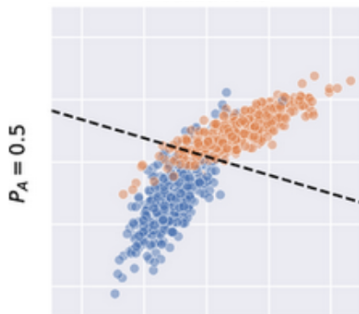
Asunción de cambio de distribución

- Suponemos que existe un **cambio de distribución** entre los datos de entrenamiento y los de test, concretamente **Prior Probability Shift**:

$$p_{\text{train}}(x | y) = p_{\text{test}}(x | y) \quad (1)$$

$$p_{\text{train}}(y) \neq p_{\text{test}}(y) \quad (2)$$

- Es decir: cambia la prevalencia de las clases, pero no el concepto de cómo son los ejemplos de cada clase.



Dataset: Amazon Reviews (LeQua)

- Opiniones de productos en Amazon.
- Dos clases: **positivas** y **negativas**.
- Se han preprocesado y convertido a vectores con 250 características por cada opinión (bag-of-words o embeddings).
- Conjunto de entrenamiento etiquetado individualmente con **5000 opiniones**.
- Conjunto de validación para probar nuestros cuantificadores con **1000 bolsas** con **250 opiniones** cada una.
 - Cada bolsa contiene un conjunto de opiniones con una distribución concreta de clases.
 - No disponemos de las etiquetas individuales de cada opinión.

Opinión positiva

"Me encanta este producto, llegó rápido y es de muy buena calidad. Lo recomiendo sin duda."

Opinión negativa

"El producto no funcionaba, muy mala experiencia. No lo volveré a comprar."

¿Qué es quantificationlib?

- Librería Python para cuantificación:
<https://github.com/AICGijon/quantificationlib>
- Proporciona implementaciones de los cuantificadores más relevantes:
 - **CC**: Classify & Count (solución trivial).
 - **AC**: Adjusted Count.
 - **DFy**: Métodos basados en ajuste de distribuciones.
 - **EMQ**: Expectation Maximization.
 - ...
- Permite entrenar y evaluar cuantificadores de forma sencilla.



¡Manos a la obra!

<http://bit.ly/44uRwag>