



Procesamiento de Lenguaje Natural

Naive Bayes

Mauricio Toledo-Acosta
mauricio.toledo@unison.mx

Departamento de Matemáticas
Universidad de Sonora



Section 1

Introducción



Referencias

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes



- **Chapter 4.** Jurafsky, D., Martin, J. H. (2019). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.*



Section 2

Clasificación en el PLN



Clasificación en PLN

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes

Consideremos la tarea de clasificación en PLN, es decir, asignar categorías a textos.

- **Análisis de sentimientos:** La extracción del sentimiento, es decir, la orientación positiva o negativa que el escritor expresa hacia algún objeto.

Reseña de una película, ↔ El sentimiento del autor hacia el producto

Editorial o un texto político ↔ El sentimiento hacia un candidato o una acción política.

Textos en redes sociales ↔ El estado de animo.



Clasificación en PLN

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naïve Bayes

- Detección de SPAM
- Identificación de idioma
- Atribución de autoría
- Detección de tópicos o temática
- Predicción de la siguiente palabra



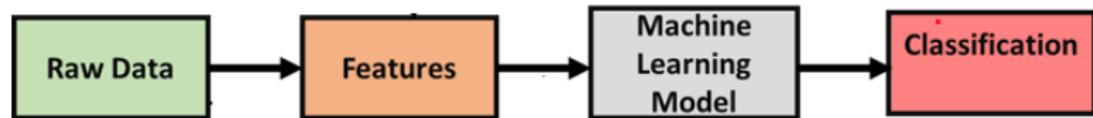
Clasificación en el Machine Learning

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes



¿Quienes son las *features* en el NLP?

- Métodos clásicos: BOW, TF-IDF, ...
- Embeddings: Redes Neuronales, LLMs.



Section 3

Clasificador Naive Bayes



Clasificación

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes

Clasificador Naive Bayes

Naive Bayes es un algoritmo de clasificación binaria y multiclase. Se llama *Naive Bayes* o Bayes ingenuo porque se hacen suposiciones para simplificar los cálculos de probabilidades por cada clase.



Enfoque probabilístico

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes

A diferencia de los clasificadores lineales que buscan una frontera de decisión que separe los datos en el espacio, un clasificador probabilístico busca estimar

$$P(\text{clase}_j|d)$$

Es decir, ¿cuál es la probabilidad de que estemos en la clase j si observamos el documento d ?



Enfoque probabilístico

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes

A diferencia de los clasificadores lineales que buscan una frontera de decisión que separe los datos en el espacio, un clasificador probabilístico busca estimar

$$P(\text{clase}_j|d)$$

Es decir, ¿cuál es la probabilidad de que estemos en la clase j si observamos el documento d ?

En un problema de clasificación binaria, predecimos que un documento d pertenece a la clase 0 si

$$P(\text{clase}_0|d) > P(\text{clase}_1|d).$$



Enfoque probabilístico

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes

A diferencia de los clasificadores lineales que buscan una frontera de decisión que separe los datos en el espacio, un clasificador probabilístico busca estimar

$$P(\text{clase}_j|d)$$

Es decir, ¿cuál es la probabilidad de que estemos en la clase j si observamos el documento d ?

En un problema de clasificación binaria, predecimos que un documento d pertenece a la clase 1 si

$$P(\text{clase}_1|x) > P(\text{clase}_0|x).$$



Enfoque probabilístico

A diferencia de los clasificadores lineales que buscan una frontera de decisión que separe los datos en el espacio, un clasificador probabilístico busca estimar

$$P(\text{clase}_j|d)$$

Es decir, ¿cuál es la probabilidad de que estemos en la clase j si observamos el documento d ?

En general, en un problema de clasificación multiclas con clases $C = \{c_0, \dots, c_m\}$, la clase predicha es

$$\hat{c} = \operatorname{argmax}_{c \in C} P(c|d).$$



Formulación

El objetivo de la clasificación Bayesiana es determinar

$$\hat{c} = \operatorname{argmax}_{c \in C} P(c|d)$$

transformando esta expresión en algo más fácil de calcular. Para esto, usamos la regla de Bayes:

$$\hat{c} = \operatorname{argmax}_{c \in C} \frac{P(d|c)P(c)}{P(d)}.$$

En todas las clases $c \in C$, el término $P(d)$ no cambia, entonces:

$$\hat{c} = \operatorname{argmax}_{c \in C} P(d|c)P(c).$$



Formulación

$$\hat{c} = \operatorname{argmax}_{c \in C} \overbrace{P(d|c)}^{\text{likelihood}} \overbrace{P(c)}^{\text{prior}}.$$

Representamos cada documento d con las *features* w_1, \dots, w_n (las palabras del documento).

$$\hat{c} = \operatorname{argmax}_{c \in C} P(w_1, w_2, \dots, w_n | c) P(c).$$

Usando la suposición *naive* de que los eventos w_1, \dots, w_n son independientes:

$$\hat{c} = \operatorname{argmax}_{c \in C} P(w_1 | c) P(w_2 | c) \cdots P(w_n | c) P(c).$$



Formulación

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes

También estamos haciendo la suposición que la posición de la palabra no importa, sólo su aparición.

$$\hat{c} = \operatorname{argmax}_{c \in C} P(w_1|c)P(w_2|c) \cdots P(w_n|c)P(c).$$



Aspectos técnicos

Procesamiento
de Lenguaje
Natural

Introducción

Clasificación
en el PLN

Clasificador
Naive Bayes

Para evitar el underflowing, usando logaritmos:

$$\hat{c} = \operatorname{argmax}_{c \in C} (\log P(w_1|c) + \log P(w_2|c) + \dots + \log P(w_n|c) + P(c))$$



Entrenamiento

Entrenar el modelo Naive Bayes quiere decir estimar las probabilidades $P(c)$ y $P(w_j|c)$. Esto lo hacemos contando:

$$P(c) = \frac{N_c}{N_{\text{total}}}$$

$$P(w_i|c) = \frac{\text{count}(w_i, c) + 1}{\sum_{w \in V} (\text{count}(w, c) + 1)}$$

- V : Vocabulario del corpus, es el conjunto de todas las palabras que aparecen en el corpus de entrenamiento (en todas las clases).
- N_c : Número de documentos en la clase c .
- N_t : Número de documentos en el corpus.



Consideraciones y estrategias adicionales

- ¿Qué hacemos con las palabras desconocidas en el conjunto de prueba?
 - ¿Qué hacemos con las stopwords?
 - ¿Cómo manejamos las negaciones?

*didn't like this movie , → didn't NOT-like
but I ... NOT_this NOT_movie ,
but I ...*

- Podemos incluir n -gramas, palabras de acuerdo al dominio de la tarea.
 - Podemos incluir conteos binarios (presencia o ausencia de palabras). En este caso, usar BernoulliNB.



Ejemplo

	Cat	Documents
Training	-	just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
Test	?	predictable with no fun

Determinar la clase del documento *predictable with no fun*.

Clase MultinomialNB en Scikit-learn



¿Por qué Naive Bayes Multinomial?

Supongamos que se realiza un experimento consistente en **extraer n bolas de k colores diferentes de una bolsa**, sustituyendo las bolas extraídas después de cada extracción. Las bolas del mismo color son equivalentes.

Denotemos por **X_i la variable que denota el número de bolas extraídas de color i** , y como p_i la probabilidad de que una extracción dada sea de color i . La función de masa de probabilidad de esta distribución multinomial es:

$$f(x_1, \dots, x_k; n, p_1, \dots, p_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$

para enteros no negativos x_1, \dots, x_k .



La distribución multinomial

Es decir:

n : longitud del documento (en palabras)

k : número de palabras posibles

p_i : probabilidad de que la palabra aparezca en la clase.

Supongamos que se realiza un experimento consistente en **extraer n bolas de k colores diferentes de una bolsa**, sustituyendo las bolas extraídas después de cada extracción. Las bolas del mismo color son equivalentes.



Ejercicio 1

Considera los siguientes reviews cortos y sus sentimientos:

Texto	Sentimiento
me encanta esta película	+
odio esta película	-
qué gran película	+
no me gustó para nada	-
es increíble	+

Usando Naive Bayes con suavizado *add-1* y un vocabulario restringido a las siguientes 7 palabras:

me, encanta, película, odio, no, gustó, increíble

Predice el sentimiento del siguiente review:

no me encanta la película



Ejercicio 2

Considera los siguientes reviews cortos de películas:

fun, copouple, love, love	comedy
fast, furious, shoot	action
copouple, fly, fast, fun, fun	comedy
furious, shoot, shoot, fun	action
fly, fast, shoot, love	action

Usando Naive Bayes y un suavizado *add-1*, predice el género de la película del review que contiene las siguientes palabras:

fast, couple, shoot, fly, live