

Trabajo Práctico Nro. 1: Método de Bayes

Ezequiel Agustin Perez, Gonzalo Nahuel Baliarda, Lucas
Agustin Vittor

Instituto Tecnológico de Buenos Aires

{eperez, gbaliarda, lvittor}@itba.edu.ar

30 de agosto de 2023

Ejercicio 1

●
○○○○○
○○○

Ejercicio 2

○
○○○
○○○

Ejercicio 3

○
○
○
○
○○○○
○○○○○
○

Ejercicio 1



Enunciado

Dado $(scones, cerveza, whisky, avena, fútbol) \in \{0, 1\}^5$:

1. Implementar un clasificador de Naive Bayes
2. Clasificar $x_1 = (1, 0, 1, 1, 0)$
3. Clasificar $x_2 = (0, 1, 1, 0, 1)$



Training Set

Preferencias británicos:

- ▶ Gustos de comida/deporte según nacionalidad (Escocés o Inglés).
- ▶ 1 = gusta la comida/deporte.
- ▶ 0 = no gusta la comida/deporte.

scones	cerveza	whiskey	avena	futbol	Nacionalidad
0	0	1	1	1	I
1	0	1	1	0	I
1	1	0	0	1	I
1	1	0	0	0	I
0	1	0	0	1	I
0	0	0	1	0	I
1	0	0	1	1	E
1	1	0	0	1	E
1	1	1	1	0	E
1	1	0	1	0	E
1	1	0	1	1	E
1	0	1	1	0	E
1	0	1	0	0	E



Testing Set

	scones	cerveza	whiskey	avena	futbol	Nacionalidad
x1	1	0	1	1	0	?????
x2	0	1	1	0	1	?????



Modelo

Clasificador Naive Bayes

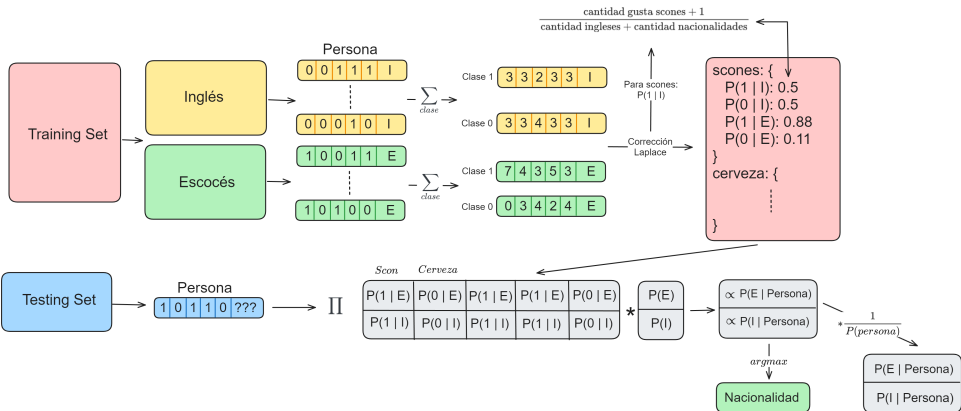
$$P(a_1, \dots, a_n | v_j) = \prod_i P(a_i | v_j)$$

$$v_{opt} = \arg \max_{v_j \in V} P(a_1, \dots, a_n | v_j) P(v_j)$$



$$v_{NB} = \arg \max_{v_j \in V} \prod_{i=1}^n P(a_i | v_j) * P(v_j)$$

Implementación



Resultados

$$P(I \mid scon, !cerveza, !whiskey, avena, !futbol) \propto$$

$$P(I) * P(scon|I) * P(!cerveza|I) * P(!whiskey|I) * P(avena|I) * P(!futbol|I) =$$

$$\frac{6}{13} * \frac{3+1}{6+2} * \dots * \dots * \dots * \dots \approx 0,011$$

Análogamente,

$$P(E \mid scon, !cerveza, !whiskey, avena, !futbol) \propto 0,035$$

Por lo que clasificamos:

	scones	cerveza	whiskey	avena	futbol	Nacionalidad
x1	1	0	1	1	0	E
x2	0	1	1	0	1	?????



Resultados

Repitiendo el proceso para x_2 , obtenemos:

	scones	cerveza	wiskey	avena	futbol	Nacionalidad
x_1	1	0	1	1	0	E
x_2	0	1	1	0	1	I

Pasando los resultados a probabilidades,

$$P(E|x_1) = 76 \%$$

$$P(I|x_2) = 83 \%$$



Conclusiones

- ▶ Se obtuvieron resultados contundentes para ambos ejemplos de prueba.
- ▶ No podemos aplicar métricas de evaluación con el conjunto de test dado, ya que no conocemos las etiquetas reales para los datos de prueba.

Ejercicio 1

○
○○○○○
○○○

Ejercicio 2

●
○○○
○○○

Ejercicio 3

○
○
○
○
○○○○
○○○○○
○

Ejercicio 2

Enunciado

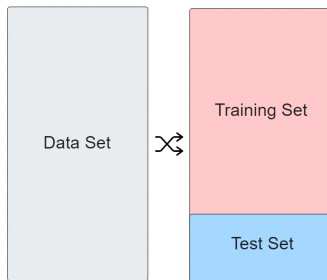
Implementar un clasificador de texto con Naive Bayes para clasificar noticias según su tipo.

1. Utilizar al menos 4 categorías.
2. Hacer matriz de confusión.
3. Calcular Accuracy, Precisión, tasa verdaderos positivos, tasa de falsos positivos y f1-score
4. Calcular la curva ROC

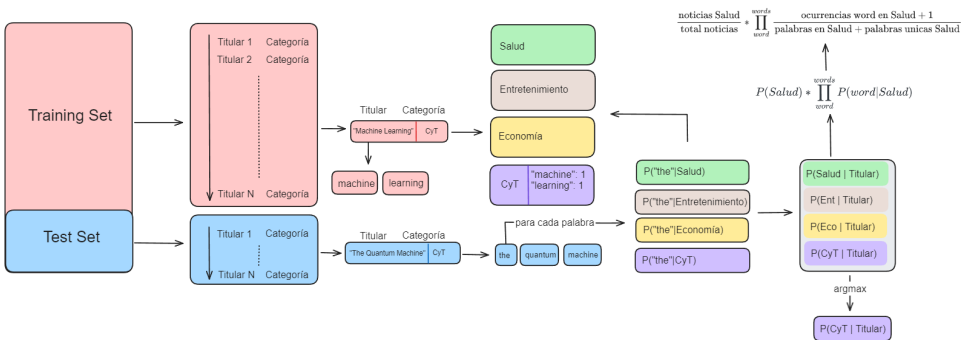
Dataset

Noticias argentinas

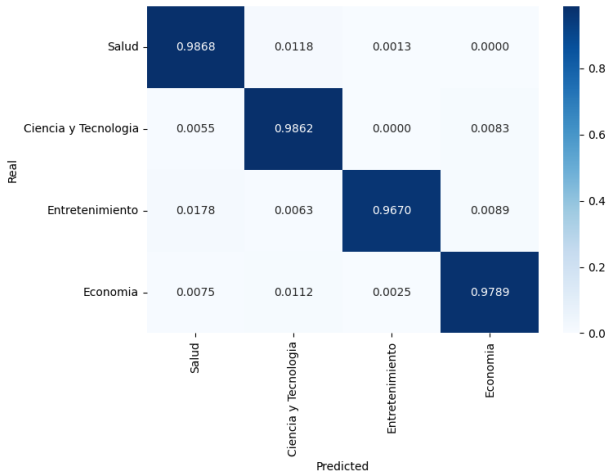
- ▶ Multiclase con clases: Salud, Entretenimiento, Economía y CyT
- ▶ Columna relevante titular
- ▶ 80-20 división de entrenamiento y testeo



Implementación



Matriz de confusión





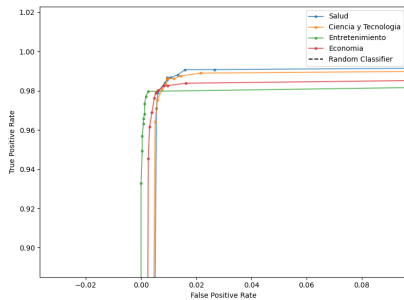
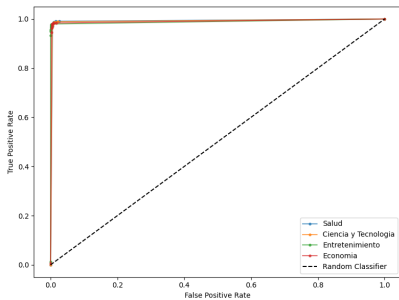
Medidas de evaluación

Categoría	Accuracy	Precisión	Tasa FP	Recall	F1-Score
Salud	0,989	0,969	0,01	0,987	0,978
Ciencia y Tecnología	0,989	0,969	0,01	0,986	0,978
Entretenimiento	0,99	0,996	0,001	0,967	0,981
Economía	0,99	0,984	0,006	0,979	0,981



Curva ROC

- Umbral $\in [0, 1]$
- Paso del umbral 0.1





Conclusiones

- ▶ Todas las categorías muestran valores muy elevados de precisión, accuracy, recall y f1-score, y muy bajos valores de tasa de falsos positivos.
- ▶ Todas las curvas ROC dan muy similares a un clasificador perfecto, con diferencias mínimas entre ellas.

Ejercicio 1

○
○
○○○○○
○○○

Ejercicio 2

○
○
○○○
○○○○

Ejercicio 3

●
○
○
○
○○○○
○○○○○
○○○○○
○

Ejercicio 3

Enunciado

Dadas las variables:

- ▶ $admit \in \{0, 1\}$
- ▶ $GRE \in \mathbb{R}$
- ▶ $GPA \in \mathbb{R}$
- ▶ $rank \in \{1, 2, 3, 4\}$

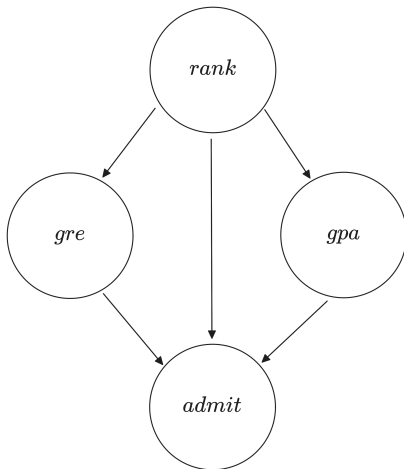
Discretizar GRE y GPA tal que:

1. $gre_d \in \{GRE \geq 500, GRE < 500\}$
2. $gpa_d \in \{GPA \geq 3, GPA < 3\}$

Luego,

1. Calcular probabilidad de que una persona que proviene de una escuela con rango 1 no haya sido admitida en la universidad
2. Calcular la probabilidad de que una persona que fue a una escuela de rango 2, tenga $GRE = 450$ y $GPA = 3.5$ sea admitida en la universidad.
3. ¿Cuál es el proceso de aprendizaje?

Estructura de la Red





$P(admit|rank, gre_d, gpa_d)$

rank	gre_d	gpa_d	admit_no	admit_yes	All
1	gre < 500	gpa < 3	0.714286	0.285714	1.0
1	gre < 500	gpa ≥ 3	0.500000	0.500000	1.0
1	gre ≥ 500	gpa < 3	0.200000	0.800000	1.0
1	gre ≥ 500	gpa ≥ 3	0.448980	0.551020	1.0
2	gre < 500	gpa < 3	0.545455	0.454545	1.0
2	gre < 500	gpa ≥ 3	0.791667	0.208333	1.0
2	gre ≥ 500	gpa < 3	0.761905	0.238095	1.0
2	gre ≥ 500	gpa ≥ 3	0.582524	0.417476	1.0
3	gre < 500	gpa < 3	0.928571	0.071429	1.0
3	gre < 500	gpa ≥ 3	0.791667	0.208333	1.0
3	gre ≥ 500	gpa < 3	0.545455	0.454545	1.0
3	gre ≥ 500	gpa ≥ 3	0.737500	0.262500	1.0
4	gre < 500	gpa < 3	0.900000	0.100000	1.0
4	gre < 500	gpa ≥ 3	0.823529	0.176471	1.0
4	gre ≥ 500	gpa < 3	0.625000	0.375000	1.0
4	gre ≥ 500	gpa ≥ 3	0.775000	0.225000	1.0
All			0.666971	0.333028	1.0

$$P(gpa_d|rank)$$

rank	$gpa < 3$	$gpa \geq 3$	All
1	0.131148	0.868852	1.0
2	0.185430	0.814570	1.0
3	0.173554	0.826446	1.0
4	0.208955	0.791045	1.0
All	0.177500	0.822500	1.0

$$P(gre_d|rank)$$

rank	gre < 500	gre ≥ 500	All
1	0.180328	0.819672	1.0
2	0.205298	0.794702	1.0
3	0.280992	0.719008	1.0
4	0.343284	0.656716	1.0
All	0.247500	0.752500	1.0

$P(rank)$

rank	proba
1	0.1525
2	0.3775
3	0.3025
4	0.1675
All	1.0

Inferencia

$$P(\text{admit} = 0 | \text{rank} = 1) = \frac{P(\text{admit} = 0, \text{rank} = 1)}{P(\text{rank} = 1)}$$

Donde,

$$P(\text{admit} = 0, \text{rank} = 1) = \sum_{\text{gre_d} \in \{\geq 500, < 500\}} \left(\sum_{\text{gpa_d} \in \{\geq 3, < 3\}} P(\text{admit} = 0, \text{rank} = 1, \text{gre_d}, \text{gpa_d}) \right)$$

Donde,

$$P(\text{admit} = 0, \text{rank} = 1, \text{gre_d}, \text{gpa_d}) =$$

$$P(\text{rank} = 1) \cdot P(\text{gre_d} | \text{rank} = 1) \cdot P(\text{gpa_d} | \text{rank} = 1) \cdot P(\text{admit} = 0 | \text{rank} = 1, \text{gre_d}, \text{gpa_d})$$



Inferencia

Reemplazando los valores presentes en las tablas, podemos calcular:

$$P(\text{admit} = 0 | \text{rank} = 1) = 0,465816$$



Inferencia

Viendo las tablas de probabilidad condicional armadas, podemos obtener directo:

$$P(admit = 1 | rank = 2, gre < 500, gpa \geq 3) = 0,208333$$



Proceso de aprendizaje

- ▶ El aprendizaje es paramétrico dado que el conocimiento adquirido son las probabilidades a priori y condicionales, plasmadas en las tablas de probabilidad.
- ▶ La estructura de la red es conocida de antemano.



Conclusiones

- ▶ Fue necesario aplicar corrección de Laplace para evitar que ciertas probabilidades se anulen.
- ▶ La inferencia es costosa dado que requiere iterar las tablas de probabilidad condicional.

Ejercicio 1

○
○○○○○
○○○

Ejercicio 2

○
○○○
○○○○

Ejercicio 3

○
○
○
○
○○○○
○○○○○
●

Muchas gracias