

Campus: Ciudad Universitaria
Facultad: Ingeniería
Materia : Inteligencia Artificial
Semestre: 2022-2
Equipo: 1
Clave: 0406
Participantes:
- Barrera Peña Víctor Miguel
- Espino De Horta Joaquín Gustavo

Profesor: Dr. Ismael Everardo Barcenás Patiño
Título : Proyecto
Subtítulo : Inferencia bayesiana
Fecha entrega: 03/05/2022

Capítulo 0 Estructura del repositorio

```
.  
├──Codigo  
├──Diapositivas  
├──Documentacion  
├──Material_ayuda  
├──LICENSE (Archivo)  
└──README.md (Archivo)
```

Aquí mostramos la estructura de los archivos contenidos en el repositorio para que puedas navegar dentro.

Código

Aquí se encuentra el código en `C++` y el `.exe` para poder ejecutarlo, esta compilado para maquina de `64 bits` en sistema Windows pero esta el código disponible para su compilación.

Diapositivas

Sólo se encuentra el PDF de las diapositivas y el código `.tex`.

Documentación

Se encuentra una documentación del programa escrito en C++ y por tanto es posible exportar un `.html` que explica todas las funcioens

Material de ayuda

Aquí se recopila, pdf´s dados por el profesor y externos con los cuales se basó para crear el proyecto, se acumulan aquí, ya que la duración de un archivo en la red tiene un tiempo limitado de vida y es posible que si se desea mejorar este proyecto a futuro o hacerle un fork, sería buena idea contar con el.

Archivos

- License (Licencia) la licencia es GNU.
- README.md Es el mismo archivo que esta leyendo, solo que en formato `.md` para poder leerlo desde Github.

Capítulo 1 Introducción

La probabilidad es una rama de las matemáticas surgido en 1553 de la mano de Gerolamo Cardano (1501-1576). Por otra parte << Pierre Fermat (1601-1665) y Blaise Pascal (1623-1662) son conocidos como los padres de la teoría de la probabilidad debido las grandes aportaciones que realizaron sobre este campo>>

<<Andréi Kolmogorov. Fue el creador de la obra «Los fundamentos de la Teoría de la Probabilidad» en la que expuso la axiomática de Kolmogorov y le hizo ser reconocido como una eminencia de la probabilidad>>.

La probabilidad busca encontrar el nivel de certeza de que ocurra un evento dado, por lo cual existe un porcentaje asociado a ello, lo cual puede ir desde un 0% hasta un 100%. Cuando el evento se aproxima a la cantidad más alta, significa que es muy posible que suceda el evento, por otro lado, cuando es cercano a 0 significa que es probable que el evento no suceda.

Ahora un concepto más avanzado es el cálculo de probabilidades dado por un suceso anterior, es decir que tan probable es que suceda un evento dado por que ocurra haya ocurrido otro evento. Para calcular dicha probabilidad utilizamos el **teorema de Bayes** el cual nos proporciona una forma fácil de calcular dicha probabilidad.

Conceptos

Definición (Regla de la adición).

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1)$$

Definición 9 (Probabilidad condicional) . La probabilidad condicional de un evento B dado otro evento A , escrita $P(B|A)$, se define

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (2)$$

Definición 13 (Clasificación Bayesiana). Considere el espacio muestra compuesto por los siguientes vectores:

x_1	x_2	\dots	x_n	y
$a_{1,1}$	$a_{1,2}$	\dots	$a_{1,n}$	b_1
$a_{2,1}$	$a_{2,2}$	\dots	$a_{2,n}$	b_2
\vdots	\vdots	\vdots	\vdots	\vdots
$a_{m,1}$	$a_{m,2}$	\dots	$a_{m,n}$	b_m

El modelo de clasificación Bayesiana se define como sigue:

$$\hat{y} = \max \{P(y_i)P(x_1 | y_i)P(x_2 | y_i) \dots P(x_n | y_i) \mid i = 1, \dots, m\} \quad (3)$$

Problema

Entrada: un espacio muestra, nuevos datos para clasificar.

Salida: la clasificación de los datos.

Dado un vector de condiciones \vec{Q} que contiene los valores $[q_1, q_2, \dots, q_j]$ para j condiciones, a los cuales debe igualarse Am_i , obtener $Y_{max}(\vec{Q})$ que es la probabilidad más grande para dicho vector.

Ejemplo. Considere la siguiente base de datos.

#	Usuario	Género	Calificación
1	F	Terror	1
2	M	Acción	3
3	F	Drama	2
4	M	Drama	2
5	F	Acción	2
6	M	Terror	3
7	F	Terror	3
8	M	Drama	1
9	F	Acción	2

Calcule la calificación que le pondría un usuario M a una película de Drama.

$$\begin{aligned} P(1)P(M | 1)P(\text{Drama} | 1) &= 2/9 * 1/2 * 1/2 = 1/18 \\ P(2)P(M | 2)P(\text{Drama} | 2) &= 4/9 * 1/4 * 1/2 = 1/18 \\ P(3)P(M | 3)P(\text{Drama} | 3) &= 1/3 * 2/3 * 0 = 0 \end{aligned} \quad (4)$$

En este caso $Y_{max}(\vec{Q}) = 1/18$, ya que es la mayor probabilidad, con las condiciones que se le dierron

Capítulo 2 Desarrollo

Solución

Pseudocódigo

```

inicio main():
    Datos= cargarDatos(nombre)

    condiciones <= input()
    Cuestion    <= input()

    real    probabilidad = 1.0, masProbable = 0.0
    cadena  Argumento = "No hay coincidencias", Objetivo, Regla

```

```

por_cada Objetivo en obten_coleccion(Cuestion) realiza:

    probabilidad = obten_Probabilidad(Datos,Objetivo)

    por_cada Regla en obten_coleccion(Condiciones) realiza:

        probabilidad = probabilidad *
obten_Probabilidad(Datos,Objetivo,Regla)

    fin_bucle

    si (probabilidad > masProbable) entonces:

        mas probable <= probabilidad
        Argumento      <= Objetivo

    fin_condicion

fin_bucle

imprimir(Argumento)

fin_main

inicio obten_Probabilidad(Datos,Objetivo):
    retorna veces_En_Categoria(Datos,Objetivo) / (Datos.Altura - 1)
fin_obten_Probabilidad

inicio obten_Probabilidad(Datos,Objetivo,Regla):
    retorna coincidencia(Datos,Objetivo,Regla) /
veces_En_Categoria(Datos,Objetivo)
fin_obten_Probabilidad

```

Experimentos

Por cada nivel de dificultad usaremos un conjunto de datos (data-set) y el problema cambiará de lo que se está preguntado, en este caso serán 3 preguntas diferentes acerca de la probabilidad.

Baja dificultad

Considere la siguiente base de datos (tomada de la pág. 8 del texto proporcionado por el profesor).

Puedes encontrar el archivo como data-set1.csv en la carpeta de código

Género	Calificación
Terror	1
Acción	3
Drama	2
Drama	2
Acción	2
Terror	3
Teror	3
Drama	1
Acción	2

Problema 1

Problema 2

Problema 3

Media dificultad

order_no	order_date	buyer	ship_city	ship_state	sku	description	quantity	item_total	shipping_fee	cod	order_status
405-9763961-5211537	Sun, 18 Jul, 2021, 10:38 pm IST	Mr.	CHANDIGARH,	CHANDIGARH	SKU: ZX-3C0F-KNJJE	100% Leather Elephant Shaped Piggy Coin Bank Block Printed West Bengal Handicrafts (Shantiniketan Art) Money Bank for Kids Children's Gift Ideas	1	₹449.00			Delivered to buyer
404-3964908-7850720	Tue, 19 Oct, 2021, 6:05 pm IST	Minam	PASIGHAT,	ARUNACHAL PRADESH	SKU: DN-0WDX-VYOT	Women's Set of 5 Multicolor Pure Leather Single Lipstick Cases with Mirror, Handy and Compact Handcrafted Shantiniketan Block Printed Jewelry Boxes	1	₹449.00	₹60.18		Delivered to buyer
171-8103182-4289117	Sun, 28 Nov, 2021, 10:20 pm IST	yatipertin	PASIGHAT,	ARUNACHAL PRADESH	SKU: DN-0WDX-VYOT	Women's Set of 5 Multicolor Pure Leather Single Lipstick Cases with Mirror, Handy and Compact Handcrafted Shantiniketan Block Printed Jewelry Boxes	1	₹449.00	₹60.18		Delivered to buyer
405-3171677-9557154	Wed, 28 Jul, 2021, 4:06 am IST	aciya	DEVARAKONDA,	TELANGANA	SKU: AH-J3AO-R7DN	Pure 100% Leather Block Print Rectangular Jewelry Box with Mirror Button Closure Multiple Utility Case (Shantiniketan Handicrafts) (Yellow)	1			Cash On Delivery	Delivered to buyer
402-8910771-1215552	Tue, 28 Sept, 2021, 2:50 pm IST	Susmita	MUMBAI,	MAHARASHTRA	SKU: KL-7WAA-Z8ZI	Pure Leather Sling Bag with Multiple Pockets and Adjustable Strap Shantiniketan Block Print Cross-Body Bags for Women (1 pc) (Brown)	1	₹1,099.00	₹84.96		Delivered to buyer
406-9292208-6725123	Thu, 17 Jun, 2021, 9:12 pm IST	Subinita	HOWRAH,	WEST BENGAL	SKU: HH-F0WW-5WVO	Women's Trendy Pure Leather Clutch Purse Leather Zipper Wallet	1	₹200.00			Delivered to buyer

Context

- Title: Amazon Seller - Order Status Prediction
- About this file: BL is a small leather products business which has recently started selling its products on Amazon. Currently, it has around 40 SKUs registered in the Indian Marketplace. Over the past few months, it has incurred some loss due to return orders. Now, BL seeks help to predict the likelihood of a new order being rejected. This would help them to take necessary actions and subsequently reduce the loss.

Contexto

- Título: Vendedor de Amazon - Predicción del estado del pedido
- Acerca de este archivo: BL es un pequeño negocio de productos de cuero que recientemente ha comenzado a vender sus productos en Amazon. Actualmente, tiene alrededor de 40 SKU registrados en el mercado indio. En los últimos meses, ha incurrido en algunas pérdidas debido a los pedidos de devolución. Ahora, BL busca ayuda para predecir la probabilidad de que una nueva orden sea rechazada. Esto les ayudaría a tomar las medidas necesarias y, posteriormente, a reducir la pérdida
- Son 172 registros por 12 columnas

[Link Kaggle](#)

Problema 1

Problema 2

Problema 3

Alta dificultad

title	artist	top genre	year released	added	bpm	nrngy	dnce	dB	live	val	dur	acous	spch	pop	top year	artist type
STARSTRUKK (feat. Katy Perry)	3OH!3	dance pop	2009	2022-02-17	140	81	61	-6	23	23	203	0	6	70	2010	Duo
My First Kiss (feat. Ke\$ha)	3OH!3	dance pop	2010	2022-02-17	138	89	68	-4	36	83	192	1	8	68	2010	Duo
I Need A Dollar	Aloe Blacc	pop soul	2010	2022-02-17	95	48	84	-7	9	96	243	20	3	72	2010	Solo
Airplanes (feat. Hayley Williams of Paramore)	B.o.B	atl hip hop	2010	2022-02-17	93	87	66	-4	4	38	180	11	12	80	2010	Solo
Nothin' on You (feat. Bruno Mars)	B.o.B	atl hip hop	2010	2022-02-17	104	85	69	-6	9	74	268	39	5	79	2010	Solo

Context

- Title: Spotify Top 100 Songs of 2010-2019
- About this file: Top 100 songs of each year on Spotify from 2010 to 2019.

Contexto

- Título: Las 100 mejores canciones de Spotify de 2010-2019
- Acerca de este archivo: Las 100 mejores canciones de cada año en Spotify de 2010 a 2019.
- Son registros 1000 por 17 columnas

[Link Kaggle](#)

Problema 1

Problema 2

Problema 3

Sin solución

Context

- Title: Walmart Sales Forecast
- About this file: This file contains additional data related to the store, department, and regional activity for the given dates. It contains the following fields:

Contexto

- Título: Pronóstico de ventas de Walmart
- Acerca de este archivo: este archivo contiene datos adicionales relacionados con la actividad de la tienda, el departamento y la región para las fechas indicadas. Contiene los siguientes campos:

Store	Dept	Date	Weekly_Sales	IsHoliday
1	1	05/02/2010	24924.5	FALSE
1	1	12/02/2010	46039.49	TRUE
1	1	19/02/2010	41595.55	FALSE
1	1	26/02/2010	19403.54	FALSE
1	1	05/03/2010	21827.9	FALSE
1	1	12/03/2010	21043.39	FALSE
1	1	19/03/2010	22136.64	FALSE
1	1	26/03/2010	26229.21	FALSE
1	1	02/04/2010	57258.43	FALSE
1	1	09/04/2010	42960.91	FALSE
1	1	16/04/2010	17596.96	FALSE
1	1	23/04/2010	16145.35	FALSE
1	1	30/04/2010	16555.11	FALSE
1	1	07/05/2010	17413.94	FALSE
1	1	14/05/2010	18926.74	FALSE
...

- Son 421,571 registros por 5 columnas

[Link Kaggle](#)

Capítulo 3 Conclusión

Barrera Peña Víctor Miguel

- Se cumplió de acuerdo a lo esperado el proyecto, presento una dificultad considerable, sobre todo la teoría matemática, sin embargo fue posible la creación del programa utilizando la probabilidad condicional usando conjuntos de datos reales, con lo cual apenas se puede vislumbrar la verdadera introducción a la inteligencia artificial, mediante la selección de "comportamiento inteligentes", como podría ser, que acción realizar de acuerdo a eventos pasados sucedidos, constituye un avance que se verá incrementado con el siguiente proyecto.

Espino de Horta Joaquín Gustavo

Anexo (teoría)

Definición 1 (Experimentos aleatorios). Un experimento con diferentes resultados, incluso si es repetido en la misma manera, se llama un experimento aleatorio.

Definición 2 (Espacio muestra). El conjunto de todos los posibles resultados de un experimento aleatorio, se llama el espacio muestra del experimento.

Ejemplo:

$$S = \mathbb{R}^+ = \{x \in \mathbb{R} | x > 0\} \quad (5)$$

$$S = \{x \in \mathbb{N} | 10 < x < 20\} \quad (6)$$

$$S = \{low, medium, light\} \quad (7)$$

$$S = \{yes, not\} \quad (8)$$

Definición 3 (Espacio muestra discreto). Un espacio muestra es discreto si y sólo si, es contable.

Ejemplo:

$$\begin{aligned} S &= \{x \in \mathbb{N} | 10 < x < 20\} \\ S &= \{low, medium, light\} \\ S &= \{yes, not\} \\ S &= \{x \in \mathbb{N} | \exists y \in \mathbb{N} : 2y + 1 = x\} \\ S &= \{0, 1\}^* \end{aligned} \quad (9)$$

Definición 4 (Espacio muestra continuo). Un espacio muestra S es continuo si y sólo si, existe una biyección $f : S \mapsto \mathbb{R}$.

Ejemplos:

$$\begin{aligned} S &= \{x \in \mathbb{R} | 0 < x < 2\} \\ S &= \mathbb{R}^+ \times \mathbb{R}^+ \\ S &= \{x | x = |S'|, S' \subseteq \mathbb{N}\} \end{aligned} \quad (10)$$

Definición 5 (Evento). Un evento S , es decir, $E \subseteq S$.

- La unión de dos eventos E_1, E_2 se define $E_1 \cup E_2 = \{x | x \in E_1 \text{ or } x \in E_2\}$.
- La intersección de dos eventos E_1, E_2 , se define $E_1 \cap E_2 = \{x | x \in E_1 \text{ and } x \in E_2\}$.
- El complemento de un evento E en el espacio muestral S se define $E^c = \{x \in S | x \notin E\}$.

Ejemplos: Considere el espacio muestra $S = yy, yn, ny, nn$. Los siguientes son eventos S .

$$\begin{aligned} E_1 &= \{yy, yn, ny\} \\ E_2 &= \{nn\} \\ E_3 &= \emptyset \\ E_4 &= S \\ E_5 &= \{yn, ny, nn\} \\ E_1 \cup E_2 &= S \quad E_1 \cap E_5 = \{yn, ny\} \quad E_1 \cup E_5 = S, \quad E_1 \cap E_5 = \{yn, ny\} \end{aligned} \quad (11)$$

Mas ejemplos. Considere $S = \mathbb{R}^+, E_1 = \{x | 1 \leq x < 10\}$ y $E_2 = \{x | 3 < x < 118\}$, entonces

$$\begin{aligned} E_1 \cup E_2 &= \{x | 1 \leq x < 118\} \\ E_1 \cap E_2 &= \{x | 3 < x < 10\} \\ E_1^c &= \{x | x \geq 10\} \\ E_1^c \cap E_2 &= \{x | 10 \leq x < 118\} \end{aligned} \quad (13)$$

Algunas propiedades de los eventos.

Dos eventos A y B se dicen mutuamente exclusivos si y sólo si, su intersección es vacía, es decir, $A \cap B = \emptyset$

$$\begin{aligned}
(E^c)^c &= E \\
(A \cup B) \cap C &= (A \cap C) \cup (B \cap C) \\
(A \cap B) \cup C &= (A \cup C) \cap (B \cup C) \\
(A \cup B)^c &= A^c \cap B^c \\
(A \cap B)^c &= A^c \cup B^c \\
A \cup B &= B \cup A \\
A \cap B &= B \cap A
\end{aligned} \tag{14}$$

$$\begin{aligned}
(E^c)^c &= E \\
(A \cup B) \cap C &= (A \cap C) \cup (B \cap C) \\
(A \cap B) \cup C &= (A \cup C) \cap (B \cup C) \\
(A \cup B)^c &= A^c \cap B^c \\
(A \cap B)^c &= A^c \cup B^c \\
A \cup B &= B \cup A \\
A \cap B &= B \cap A
\end{aligned} \tag{15}$$

Definición 6 (Probabilidad). En un espacio muestra discreto, la probabilidad de un evento E , escrito $P(E)$, es igual a la suma de las probabilidades de sus resultados en E .

Ejemplo.

Un experimento aleatorio puede resultar en a, b, c, d con probabilidades 0.1, 0.3, 0.5 y 0.1, respectivamente. Considere el evento A como $\{a, b\}$, $B = \{b, c, d\}$ y $C = d$. Entonces,

$$\begin{aligned}
P(A) &= 0.1 + 0.3 = 0.4 \\
P(B) &= 0.3 + 0.5 + 0.1 = 0.9 \\
P(C) &= 0.1
\end{aligned} \tag{16}$$

También, $P(A^c) = 0.6$, $P(B^c) = 0.1$, $P(C^c) = 0.9$. Más aún, debido a $A \cap B = \{b\}$, entonces $P(A \cap B) = 0.3$. Debido a $A \cup B = \{a, b, c, d\}$, $P(A \cup B) = 0.1 + 0.3 + 0.5 + 0.1 = 1$. Y debido a $A \cap C = \emptyset$, entonces $P(A \cap C) = 0$.

Definición 7 (Axiomas de la probabilidad). Considere los eventos E , E_1 y E del espacio muestra S de un experimento aleatorio.

- $P(S) = 1$
- $0 \leq P(E) \leq 1$
- Si $E_1 \cap E_2 = \emptyset$, entonces $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

Algunas propiedades.

$$P(\emptyset) = 0 \quad P(E^c) = 1 - P(E) \quad \text{Si } E_1 \subseteq E_2, \text{ entonces } P(E_1) \leq P(E_2) \tag{17}$$

Ejercicios

Problema 1

Suponga que las placas de los vehículos están compuestas inicialmente por tres dígitos (0–9), seguidas de tres letras (A–Z). Calcule la probabilidad de una determinada placa.

Problema 2

Un mensaje puede seguir diferentes rutas a través de una red de servidores. En el primer paso, el mensaje puede llegar a cinco servidores, a partir de cada uno de estos servidores, el mensaje puede llegar a cinco servidores más, desde los cuales puede acceder a otros cuatro servidores.

- Calcule la cantidad de rutas.
- Si todas las rutas son igualmente probables, calcule la probabilidad de que el mensaje llegue a alguno de los cuatro servidores del tercer bloque.

Demuestre lo siguiente

$$\begin{array}{ll}
 P(\emptyset) = 0 & (18) \\
 S^c = \emptyset & \text{Definimos el conjunto} \quad (19) \\
 P(S \cup S^c) = P(S) + P(S^c) - P(S \cap S^c) & \text{Usamos la propiedad} \quad (20) \\
 P(S \cup S^c) = 1 + P(S^c) - 0 & \text{Sustituimos} \quad (21) \\
 P(S \cup S^c) = 1 & \text{Declaramos una propiedad} \quad (22) \\
 1 - 1 = P(S^c) & \text{Sustituimos lo anterior} \quad (23) \\
 0 = P(S^c) & (24) \\
 P(S^c) = 0 & (25) \\
 \therefore P(\emptyset) = 0 & (26)
 \end{array}$$

$$P(E^c) = 1 - P(E) \quad (27)$$

$$\text{Para cualquier evento } A, P(A) + P(A') = 1, \text{ a partir de la cual } P(A) = 1 - P(A'). \quad (28)$$

Comprobación En el axioma 3, sea $k = 2$, $A_1 = A$ y $A_2 = A'$. Como por definición de A' , $A \cup A' = S$ en tanto A y A' sean eventos disjuntos, $1 = P(S) = P(A \cup A') = P(A) + P(A')$

$$\text{Si } E_1 \subseteq E_2 \text{ entonces } P(E_1) \leq P(E_2) \quad (29)$$

- Si tienen los mismos elementos, entonces la probabilidad de obtener un evento E_1, E_2 , es la misma, pero y si E_2 Tiene un elemento más que E_1 su probabilidad aumenta, esto se puede demostrar por ordinales de conjuntos.

$$\frac{\#E_1}{Total} \leq \frac{\#E_2}{Total} \quad (30)$$

E_1, E_2 es el número de elementos a favor.

Mas ejemplos

Problema 1

Una mezcla química es preparada correctamente por el 25% de los técnicos de un laboratorio, 70% de los técnicos la preparan con un error mínimo, y 5% con un error mayor.

- Si un técnico es elegido aleatoriamente, ¿Cuál es la probabilidad de que prepare la mezcla sin error alguno?

$$P(E) = 25\% \quad (31)$$

- Calcule la probabilidad de que el técnico la prepare con cualquier tipo de error.

$$P(E) = 70\% + 5\% = 75\% \quad (32)$$

Problema 2

Considere las emisiones de tres fabricas clasificadas por su calidad. De la primera fábrica 22 muestras de emisiones cumplen con el mínimo, y 8 no lo hacen; 25 cumplen con el mínimo y 5 no, en el caso de la segunda fábrica; en cuanto a la tercera, 30 cumplen y 10 no. Considere A denota el evento de las muestras de emisiones de la primera fábrica, y B como el evento de una muestra cumple con el mínimo. Calcule las siguientes probabilidades.

$$P(A) \quad P(B) \quad P(A^c) \quad P(A \cap B) \quad P(A \cup B) \quad P(A^c \cup B) \quad (33)$$

Solución:

$$P(A) = \frac{22 + 8}{100} = 0.3 \quad (34)$$

$$P(B) = \frac{22 + 25 + 30}{100} \quad (35)$$

$$P(A^c) = \frac{70}{100} \quad (36)$$

$$P(A \cap B) = \frac{22}{100} \quad (37)$$

$$P(A \cup B) = \frac{22 + 25 + 30}{100} \quad (38)$$

$$P(A^c \cup B) = \frac{70 + 8}{100} \quad (39)$$

Definición 8 (regla de la adición).

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (40)$$

Definición (Regla de la adición).

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (41)$$

Ejemplo

Considere la tabla abajo con el historial de producción de 949 semiconductores. Suponga que un semiconductor es elegido aleatoriamente. Considere M denota el evento de que el semiconductor contiene niveles altos de contaminación. C es el evento cuando el semiconductor se encuentra en el centro de una herramienta de pulverización.

$$P(H \cup C) = P(H) + P(C) - P(H \cap C) = \frac{1}{940} (358 + 626 - 112) = \frac{872}{940} \quad (42)$$

Ejemplo: En el mismo contexto del ejemplo anterior, considere E_1 el evento de que un semiconductor contiene 4 o más partículas contaminantes, E_2 es el evento de que un semiconductor se encuentra en la orilla de la herramienta.

$$P(E_1 \cap E_2) = P(E_1) + P(E_2) - P(E_1 \cup E_2) = \quad (43)$$

Ejercicios

Si $P(A) = 0.3$, $P(B) = 0.2$, y $P(A \cap B) = 0.1$, determine las siguientes probabilidades:

$$\begin{array}{lll} P(A^c) & P(A \cup B) & P(A^c \cap B) \\ P(A \cap B^c) & P((A \cup B)^c) & P(A^c \cup B) \end{array} \quad (44)$$

Definición 9 (Probabilidad condicional). La probabilidad condicional de un evento B dado otro evento A , escrita $P(B|A)$, se define

$$P(A \cap B)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (45)$$

8

- Calcule la probabilidad de que el inspector detecte un objeto defectuoso.
- Si un objeto es clasificado libre de defectos, determine la probabilidad de que efectivamente lo esté.

Definición 13 (Clasificación Bayesiana). Considere el espacio muestra compuesto por los siguientes vectores:

x_1	x_2	\dots	x_n	y
$a_{1,1}$	$a_{1,2}$	\dots	$a_{1,n}$	b_1
$a_{2,1}$	$a_{2,2}$	\dots	$a_{2,n}$	b_2
\vdots	\vdots	\vdots	\vdots	\vdots
$a_{m,1}$	$a_{m,2}$	\dots	$a_{m,n}$	b_m

El modelo de clasificación Bayesiana se define como sigue:

$$\hat{y} = \max \{P(y_i)P(x_1 | y_i)P(x_2 | y_i) \dots P(x_n | y_i) \mid i = 1, \dots, m\} \quad (46)$$

Ejemplo

Considere la siguiente base de datos

Género	Calificación
Terror	1
Acción	3
Drama	2
Drama	2
Acción	2
Terror	3
Teror	3
Drama	1
Acción	2

- Calcule la probabilidad de que el inspector detecte un objeto defecuso

La calificación de una película de acción se puede predecir de la siguiente forma:

$$P(1)P(\text{Acción} \mid 1) = 2/9 * 0 = 0$$

$$\begin{aligned} P(2)P(\text{Acción} \mid 2) &= 4/9 * 1/2 = 2/9 \\ P(3)P(\text{Acción} \mid 3) &= 1/3 * 1/3 = 1/9 \end{aligned} \quad (47)$$

por lo tanto, la calificación de una película de acción será 2.

9

Ejemplo. Considere la siguiente base de datos.

#	Usuario	Género	Calificación
1	F	Terror	1
2	M	Acción	3
3	F	Drama	2
4	M	Drama	2
5	F	Acción	2
6	M	Terror	3
7	F	Terror	3
8	M	Drama	1
9	F	Acción	2

Calcule la calificación que le pondría un usuario M a una película de Drama.

$$\begin{aligned} P(1)P(M \mid 1)P(\text{Drama} \mid 1) &= 2/9 * 1/2 * 1/2 = 1/18 \\ P(2)P(M \mid 2)P(\text{Drama} \mid 2) &= 4/9 * 1/4 * 1/2 = 1/18 \\ P(3)P(M \mid 3)P(\text{Drama} \mid 3) &= 1/3 * 2/3 * 0 = 0 \end{aligned} \quad (48)$$

Ejercicios

Considere la siguiente base de datos.

#	Productora	Usuario	Género	Calificación
1	Universal	F	Terror	1
2	Universal	M	Acción	3
3	Warner	F	Drama	2
4	Disney	M	Drama	2
5	Warner	F	Acción	2
6	Disney	M	Terror	3
7	Universal	F	Terror	3
8	Disney	M	Drama	1
9	Warney	F	Acción	2
10	Warner	M	Acción	1
11	Disney	F	Drama	2
12	Universal	F	Terror	3
13	Warner	F	Terror	3
14	Disney	M	Acción	2
15	Universal	M	Drama	1

Calcule la calificación que otorgará un usuario F a una película de Acción producida Warner.

$$\begin{aligned}
 & P(1) \cdot P(F|1) \cdot P(A|1) \cdot P(W|1) \\
 &= \frac{4}{15} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{4}{960} = 0.0041\bar{6}
 \end{aligned} \tag{49}$$

$$\begin{aligned}
 & P(2) \cdot P(F|2) \cdot P(A|2) \cdot P(W|2) \\
 &= \frac{6}{15} \cdot \frac{4}{6} \cdot \frac{3}{6} \cdot \frac{3}{6} = \frac{216}{3240} = 0.0666\bar{6}
 \end{aligned} \tag{50}$$

$$\begin{aligned}
 & P(3) \cdot P(F|3) \cdot P(A|3) \cdot P(W|3) \\
 &= \frac{5}{15} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{1}{5} = \frac{15}{1,875} = 0.0026\bar{6}
 \end{aligned} \tag{51}$$

10

Ejercicio

- Considere la siguiente base de datos.

#	Director	Productora	Usuario	Género	Calificación
1	Hnos. Coen	Universal	F	Terror	1
2	Del tor	Universal	M	Acción	3
3	Bañuel	Warner	F	Drama	2
4	Bañuel	Disney	M	Drama	2
5	Hnos. Coen	Warner	F	Acción	2
6	Del toro	Disney	M	Terror	3
7	Del toro	Universal	F	Terror	3
8	Hnos. Coen	Disney	M	Drama	1
9	Bañuel	Warner	F	Acción	2
10	Del toro	Warner	M	Acción	1
11	Bañuel	Disney	F	Drama	2
12	Hnos. Coen	Universal	F	Terror	3
13	Hnos. Coen	Warner	F	Terror	3
14	Del Toro	Disney	M	Acción	2
15	Bañuel	Universal	M	Drama	1
16	Bañuel	Warner	M	Acción	1
17	Del Toro	Warner	F	Acción	2
18	Bañuel	Disney	M	Drama	3
19	Hnos. Coen	Universal	M	Terror	1
20	Hnos. Coen	Warner	F	Terror	1
21	Del Toro	Disney	F	Acción	2
22	Bañuel	Uiversal	M	Drama	3

Calcule la calificación que otorgará un usuario M a una película de Terror, producida y dirigida por Bañuel.

- Implemente un algoritmo para el modelo de clasificación Bayesiana.

$$\begin{aligned}
 &P(1) \cdot P(M|1) \cdot P(Terror|1) \cdot P(Universal|1) \cdot P(Bañuel|1) \\
 &= \frac{7}{22} \cdot \frac{5}{7} \cdot \frac{3}{7} \cdot \frac{3}{7} \cdot \frac{2}{7} = \frac{45}{3773} = 0.0119268
 \end{aligned} \tag{52}$$

$$\begin{aligned}
 &P(2) \cdot P(M|2) \cdot P(Terror|2) \cdot P(Universal|2) \cdot P(Bañuel|2) \\
 &= \frac{8}{22} \cdot \frac{2}{7} \cdot 0 \dots = 0
 \end{aligned} \tag{53}$$

$$\begin{aligned}
 &P(3) \cdot P(M|3) \cdot P(Terror|3) \cdot P(Universal|3) \cdot P(Bañuel|3) \\
 &= \frac{7}{22} \cdot \frac{4}{7} \cdot \frac{4}{7} \cdot \frac{4}{7} \cdot \frac{2}{7} = \frac{896}{52,822} = 0.0169626
 \end{aligned} \tag{54}$$

Referencias

- Barcenás, E. (s. f.). *Introducción a la probabilidad*. classrom. Recuperado 3 de mayo de 2022, de <https://classroom.google.com/c/NDYyNTA3MDkxNDA2>
- Devore, J. L. *Probabilidad y estadística para ingeniería y ciencias*/por Jay L. Devore (No. 519 D4.).
- Westreicher, G. (2021, 14 julio). *Probabilidad*. Economipedia. <https://economipedia.com/definiciones/probabilidad.html>