

# Minería de Datos

## IIC2433

Naïve Bayes

Vicente Domínguez

Basado en diapositivas del prof. Denis Parra

# ¿Qué vimos las clases pasadas?

- KNN

# ¿Qué veremos esta clase?

- Otra forma de clasificar: Naïve Bayes

# Probabilidades condicionales y conjuntas

## *Ejemplo*



Al tirar una moneda

- ¿Cuál es la probabilidad de que salga sello
- ¿Cuál es la probabilidad de que salga dos veces sello al tirarla dos veces?

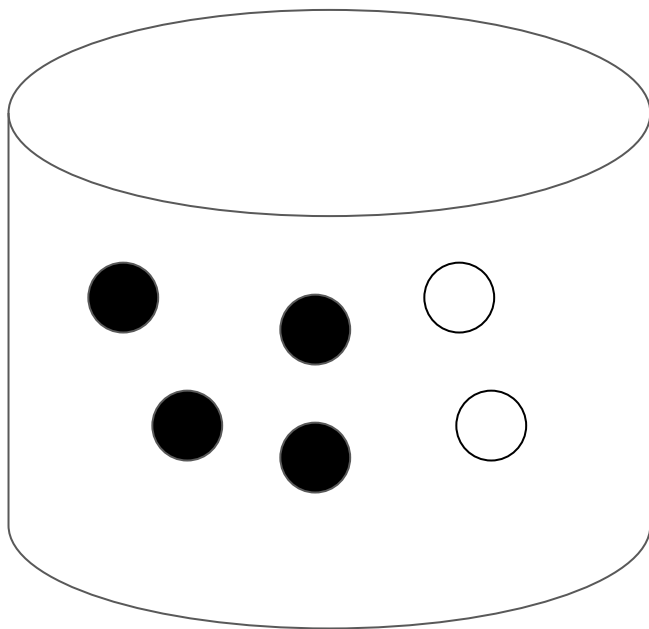
## Probabilidad conjunta con eventos **independientes**

$$P(A, B) = P(A) * P(B)$$

$$P(A, B, C) = P(A) * P(B) * P(C)$$

# Probabilidades condicionales y conjuntas

## *Ejemplo*



Tengo 4 bolitas negras y 2 blancas en una tómbola,

- Al sacar una bolita al azar, ¿Cuál es la probabilidad de que salga una blanca?
- Al sacar dos bolitas al azar, ¿Cuál es la probabilidad de que ambas salgan blancas?

## Probabilidad conjunta con eventos **dependientes**

$$P(A, B) = P(A|B)*P(B) = P(B|A)*P(A)$$

$$P(A, B, C) = P(A|B,C)*P(B|C)*P(C)$$

# Teorema de Bayes



Thomas Bayes (1701 –  
1761)

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

# Teorema de Bayes

- $P(A = \text{sí})$ : Probabilidad del evento A sea “sí”
- $P(A=\text{sí} | B=\text{sí})$ : Probabilidad de que el evento A sea “sí” DADO QUE el evento B fue “sí”
- Por simplicidad, usamos  $P(A) = P(A=\text{“sí”})$

$$P(A | B) = \frac{P(A, B)}{P(B)} \quad P(A | B) = \frac{P(B | A) * P(A)}{P(B)}$$



# Noción del Teorema de Bayes

- << La riqueza hace la felicidad >>
- ¿Son felices los ricos?  $P(\text{feliz} = \text{sí} \mid \text{rico} = \text{sí})$
- ... yo sé que de la gente feliz, 20% es rica.



- $$P(\text{feliz} \mid \text{rico}) = \frac{P(\text{rico} \mid \text{feliz}) * P(\text{feliz})}{P(\text{rico})}$$
- 20% no es tanto ... por lo cual podemos concluir que la riqueza no hace la felicidad. ¿o no?

# Noción del Teorema de Bayes

- “La riqueza hace la felicidad”
- ¿Son felices los ricos?  $P(\text{feliz} = \text{sí} \mid \text{rico} = \text{sí})$
- ... yo sé que de la gente feliz, 20% es rica.



## Supongamos:

A: gente feliz = 40%  
de la población

B: gente rica = 10% de  
la población

$$P(\text{feliz} \mid \text{rico}) = \frac{P(\text{rico} \mid \text{feliz}) * P(\text{feliz})}{P(\text{rico})}$$

$$P(\text{feliz} \mid \text{rico}) = \frac{P(\text{rico} \mid \text{feliz}) * P(\text{feliz})}{P(\text{rico})}$$

# Noción del Teorema de Bayes

- “La riqueza hace la felicidad”
- ¿Son felices los ricos?  $P(\text{feliz} = \text{sí} \mid \text{rico} = \text{sí})$
- ... yo sé que de la gente feliz, 20% es rica.



## Supongamos:

A: gente feliz = 40%  
de la población

B: gente rica = 10% de  
la población

C:  $P(\text{rico} \mid \text{feliz}) = 20\%$

$$P(\text{feliz} \mid \text{rico}) = \frac{P(\text{rico} \mid \text{feliz}) * P(\text{feliz})}{P(\text{rico})}$$

$$P(\text{feliz} \mid \text{rico}) = \frac{P(\text{rico} \mid \text{feliz}) * P(\text{feliz})}{P(\text{rico})}$$

$$P(\text{feliz} \mid \text{rico}) = \frac{0.2 * 0.4}{0.1}$$

$$= 0.8 = 80\%$$

# ¿Y por qué se llama “Naïve” ?

- Naïve significa “**ingenuo**”
- Es “ingenuo” por que asume independencia de los eventos\*

- \* en realidad, asume independencia condicional

Manzana	Carne	Pastel	¿Alergia?
No	Sí	No	Sí
No	Sí	Sí	Sí
No	Sí	No	Sí
Sí	Sí	Sí	Sí
Sí	Sí	No	No
No	No	Sí	No
Sí	No	No	No
No	No	No	No

- ¿Cuál es la probabilidad de haber consumido el alimento Manzana, dado que hubo alergia, es decir  $P(\text{Manzana}=\text{Si}|\text{Alergia}=\text{Si})$ ? ¿Cuál es la probabilidad de haber consumido el alimento pastel, dado que no hubo alergia, es decir,  $P(\text{Pastel}=\text{Si}|\text{Alergia}=\text{No})$  ?

$$P(M|A) = (P(A|M) * P(M)) / P(A) = (1/3 * 3/8) / 1/2 = 1/4$$

$$P(P|\text{No } A) = (P(\text{No } A|P) * P(P)) / P(\text{No } A) = (1/3 * 3/8) / 1/2 = 1/4$$

# Volviendo: Ejemplo de Clasificación

- Consideremos un auto SUV, color rojo, doméstico. ¿La probabilidad de que la roben es mayor o menor de que no la roben?

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

$$P(\text{Robo}) = (L(\text{Robo})P(\text{Color} | \text{Robo}) * P(\text{Tipo} | \text{Robo}) * P(\text{Origen} | \text{Robo})) / N$$

$$N = L(\text{Robo})P(\text{Color} | \text{Robo}) * P(\text{Tipo} | \text{Robo}) * P(\text{Origen} | \text{Robo})$$

+

$$L(\text{No Robo})P(\text{Color} | \text{No Robo}) * P(\text{Tipo} | \text{No Robo}) * P(\text{Origen} | \text{No Robo})$$

$$P(\text{Robo}) = (5/10 * 3/5 * 1/5 * 2/5) / ((5/10 * 3/5 * 1/5 * 2/5) + (5/10 * 2/5 * 3/5 * 3/5))$$

$$P(\text{Robo}) = 0.25$$

Manzana	Carne	Pastel	¿Alergia?
No	Sí	No	Sí
No	Sí	Sí	Sí
No	Sí	No	Sí
Sí	Sí	Sí	Sí
Sí	Sí	No	No
No	No	Sí	No
Sí	No	No	No
No	No	No	No

- Basado en los datos de la Tabla 1, usando un clasificador Naive Bayes, clasifique los siguientes dos casos dados los datos de la Tabla 1.

Manzana	Carne	Pastel	¿Alergia?
Sí	No	Sí	??
Sí	Sí	No	??



# Ejemplo de Clasificación Numérico

- ¿Qué ocurrirá en el caso con datos numéricos y no categóricos?

# Ejemplo de Clasificación Numérico

- ¿Qué ocurrirá en el caso con datos numéricos y no categóricos?
- R: En general se asume que los datos distribuyen como una gaussiana (puede ser otra distribución) y se calculan sus parámetros acorde a los datos.
- Luego, se utilizan para utilizar la función de densidad para calcular la probabilidad del dato a predecir.

# Referencias

- Material de Tom M. Mitchell, CMU:

[http://www.cs.cmu.edu/~awm/15781/slides/NB  
ayes-9-27-05.pdf](http://www.cs.cmu.edu/~awm/15781/slides/NB<br/>ayes-9-27-05.pdf)

[http://www.cs.cmu.edu/~tom/mlbook/NBayesL  
ogReg.pdf](http://www.cs.cmu.edu/~tom/mlbook/NBayesL<br/>ogReg.pdf)