



# APLICACIONES DE DATA SCIENCE



¿Cuál es la utilidad de las  
probabilidades en el  
modelamiento predictivo?



## Logro Unidad 2

El estudiante podrá utilizar las habilidades que necesita para aplicar la ciencia de datos a las diferentes tareas de la minería de textos.



# Contenido 4

## Modelo de lenguaje

- Modelamiento de Lenguaje
- Problema del Modelamiento del Lenguaje
- Repaso de Conceptos Probabilísticos
- Trigramas



## 1. Motivación

**Problema:** Speech Recognition (reconocimiento de voz)



Dado una señal de audio encontrar/asignar la oración en texto mas probable.

- Modelo Estadístico y Probabilístico



## 1. Motivación

**Problema:** Speech Recognition (reconocimiento de voz)



## 1. Motivación

**Problema:** Speech Recognition (reconocimiento de voz)

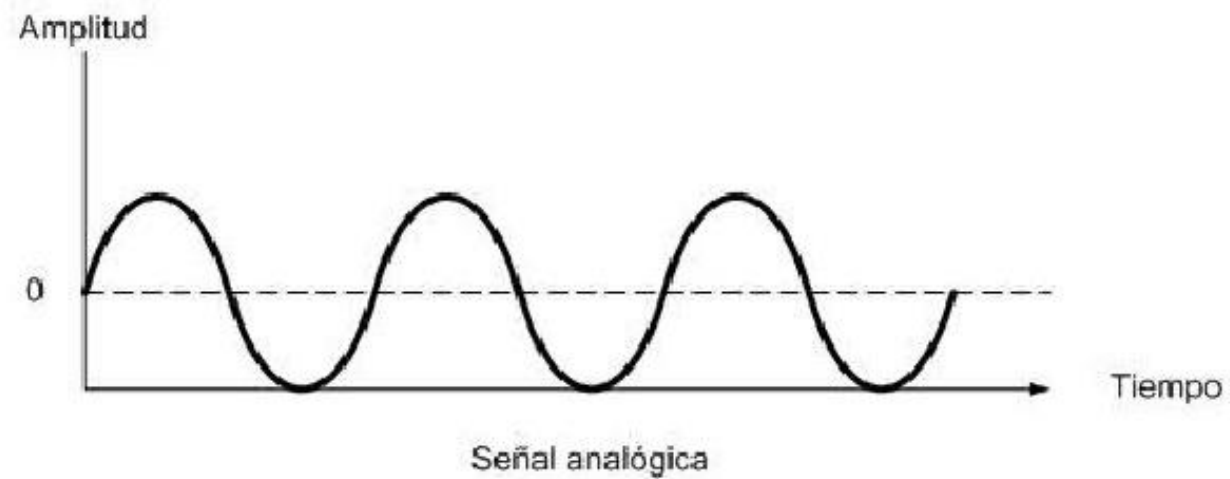
**INPUT**



## 1. Motivación

**Problema:** Speech Recognition (reconocimiento de voz)

**INPUT**





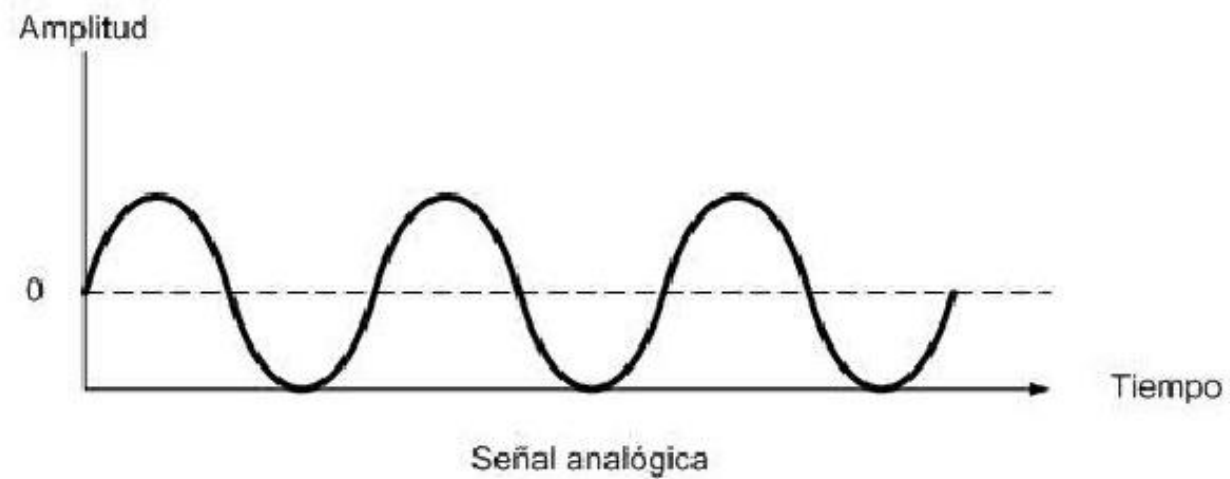


## 1. Motivación

**Problema:** Speech Recognition (reconocimiento de voz)

INPUT

OUTPUT

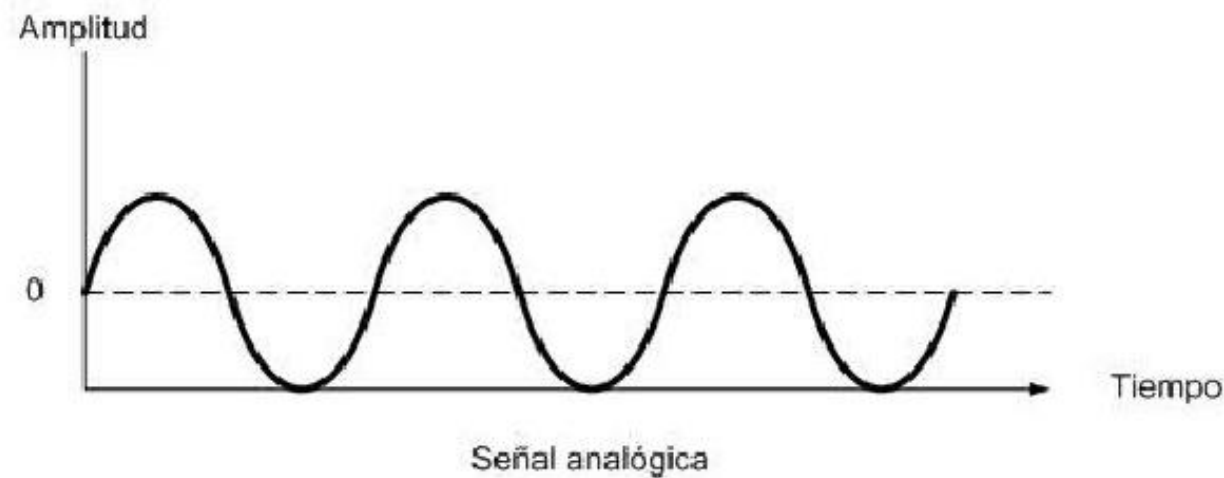




## 1. Motivación

**Problema:** Speech Recognition (reconocimiento de voz)

INPUT



OUTPUT

### Modelamiento de Lenguaje

Trata de encontrar la oración más probable .

- Modelo Estadístico y Probabilístico



## 2. Contenido

- Modelo Estadístico y Probabilístico



## 2. Contenido

- ❑ Modelar probabilísticamente el lenguaje.
- ❑ Modelos Trigrama.
- ❑ Modelos de evaluación de lenguaje: Perplexity.
- ❑ Técnicas de Estimación
  - Interpolación Lineal
  - Método de desconteo

- Modelo Estadístico y Probabilístico



### **3. Problema del Modelamiento del Lenguaje**



### 3. Problema del Modelamiento del Lenguaje

- Tenemos un conjunto finito de términos **V**.

$$\mathcal{V} = \{\text{the, a, man, telescope, Beckham, two, \dots}\}$$



### 3. Problema del Modelamiento del Lenguaje

- Tenemos un conjunto finito de términos  $\mathbf{V}$ .

$$\mathcal{V} = \{\text{the, a, man, telescope, Beckham, two, \dots}\}$$

- Tenemos un conjunto de oraciones (*strings*) que se pueden formar  $\mathbf{V}^+$ .



### 3. Problema del Modelamiento del Lenguaje

- Tenemos un conjunto finito de términos  $\mathbf{V}$ .

$$\mathcal{V} = \{\text{the, a, man, telescope, Beckham, two, ...}\}$$

- Tenemos un conjunto de oraciones (*strings*) que se pueden formar  $\mathbf{V}^+$ .

the STOP

a STOP

the fan STOP

the fan saw Beckham STOP

the fan saw saw STOP

the fan saw Beckham play for Real Madrid STOP





### 3. Problema del Modelamiento del Lenguaje

- Tenemos un conjunto finito de términos  $\mathbf{V}$ .

$$\mathcal{V} = \{\text{the, a, man, telescope, Beckham, two, ...}\}$$

- Tenemos un conjunto de oraciones (*strings*) que se pueden formar  $\mathbf{V}^+$ .

the **STOP**      Se usa un símbolo especial para identificar  
a STOP      el fin de la oración (token especial).  
the fan STOP  
the fan saw Beckham STOP  
the fan saw saw STOP  
the fan saw Beckham play for Real Madrid STOP

- Modelo Estadístico y Probabilístico



### **3. Problema del Modelamiento del Lenguaje**

- Si quiero obtener un Modelamiento de Lenguaje.



### **3. Problema del Modelamiento del Lenguaje**

- Si quiero obtener un Modelamiento de Lenguaje.
- Debemos tener un conjunto de entrenamiento de oraciones (CORPUS de entrenamiento).



### 3. Problema del Modelamiento del Lenguaje

- Si quiero obtener un Modelamiento de Lenguaje.
- Debemos tener un conjunto de entrenamiento de oraciones (CORPUS de entrenamiento).
- Necesitamos aprender una distribución de probabilidad  $P$ .

$$\sum_{x \in \mathcal{V}^{\dagger}} p(x) = 1, \quad p(x) \geq 0 \text{ for all } x \in \mathcal{V}^{\dagger}$$



### 3. Problema del Modelamiento del Lenguaje

- Si queremos obtener un Modelamiento de Lenguaje:
  - ❑ Debemos tener un conjunto de entrenamiento de oraciones (CORPUS de entrenamiento).
  - ❑ Necesitamos aprender una distribución de probabilidad P.

$$\sum_{x \in \mathcal{V}^{\dagger}} p(x) = 1, \quad p(x) \geq 0 \text{ for all } x \in \mathcal{V}^{\dagger}$$

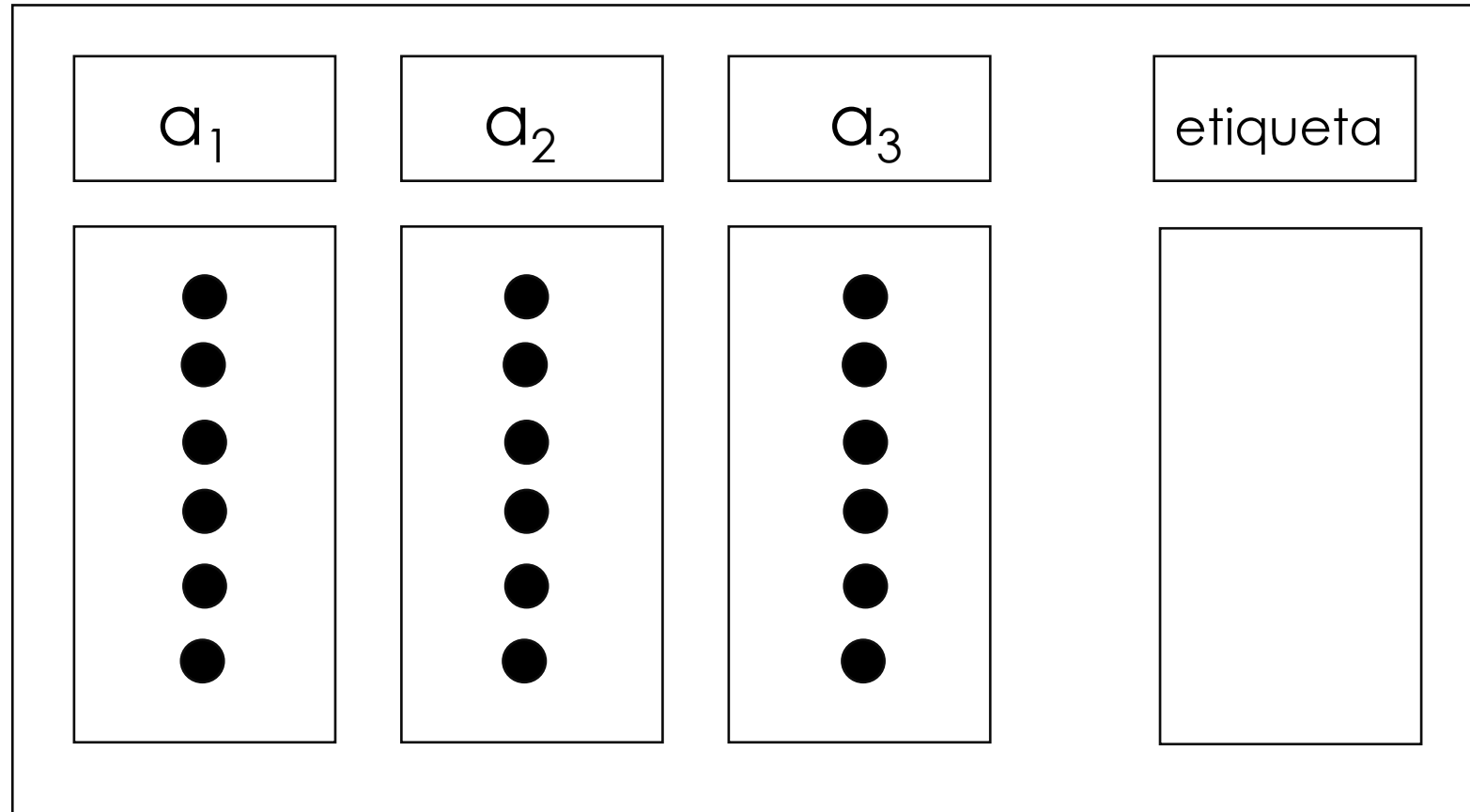
- Modelo Estadístico y Probabilístico



## ¿De donde partimos?



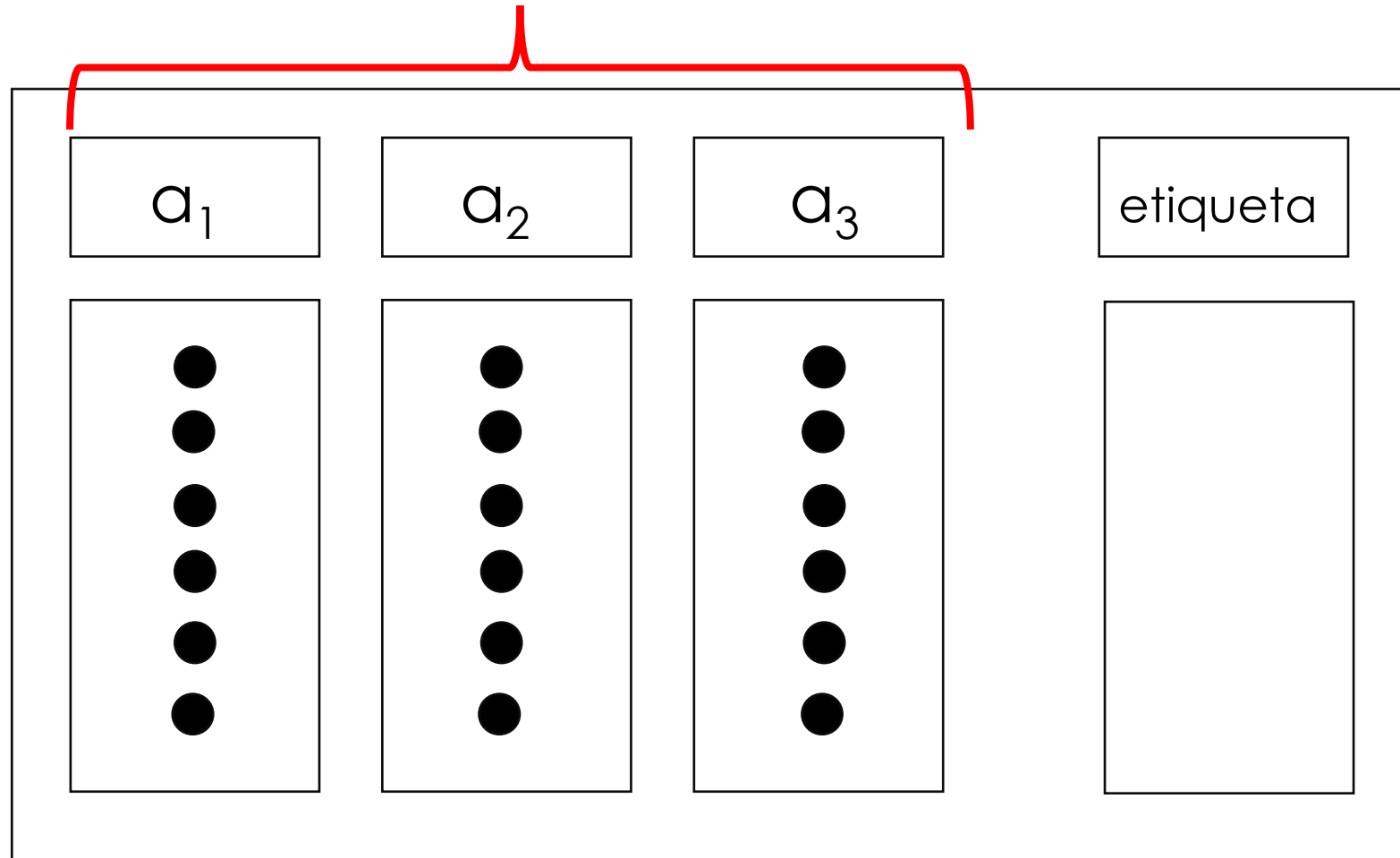
## ¿De donde partimos?





## ¿De donde partimos?

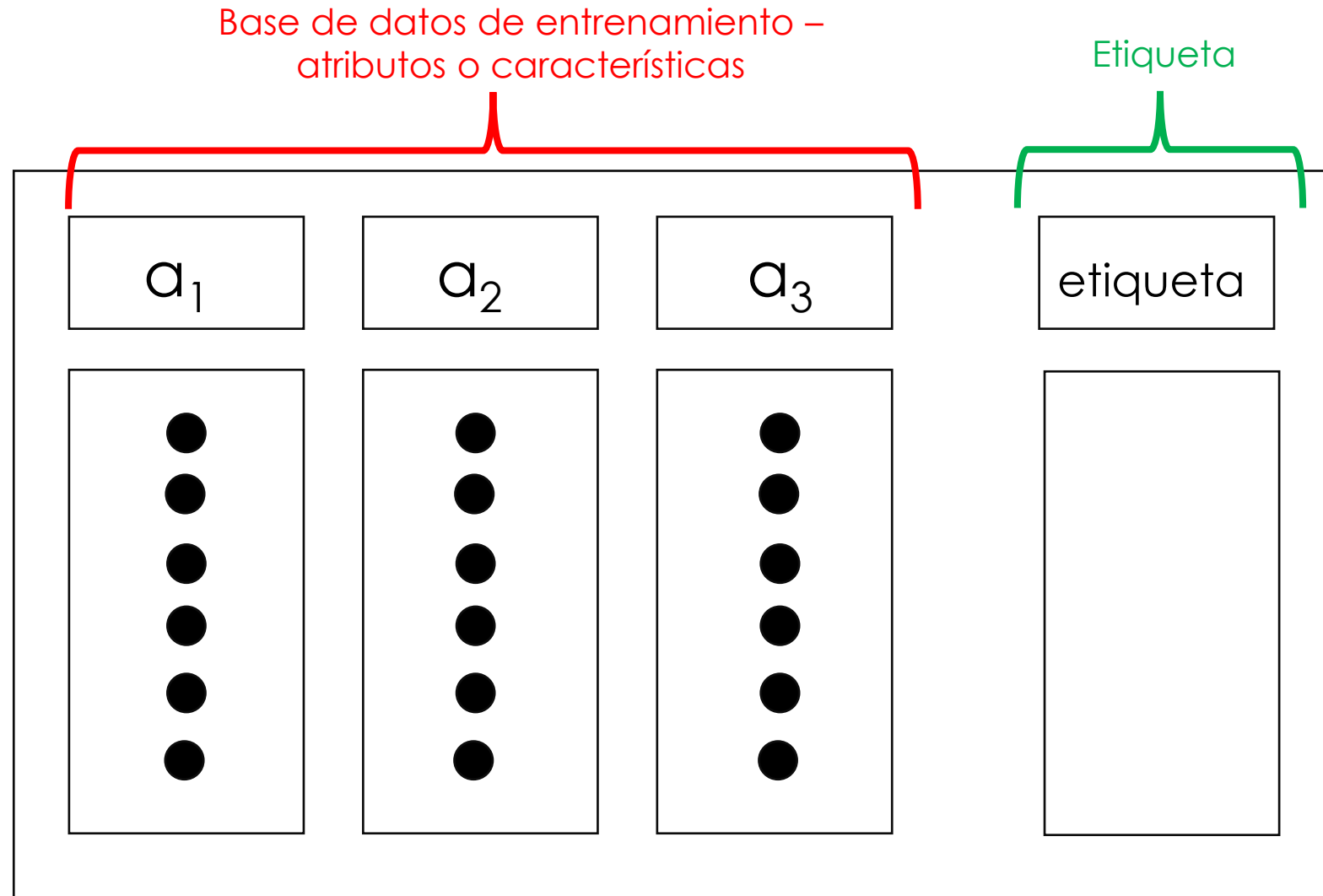
Base de datos de entrenamiento –  
atributos o características





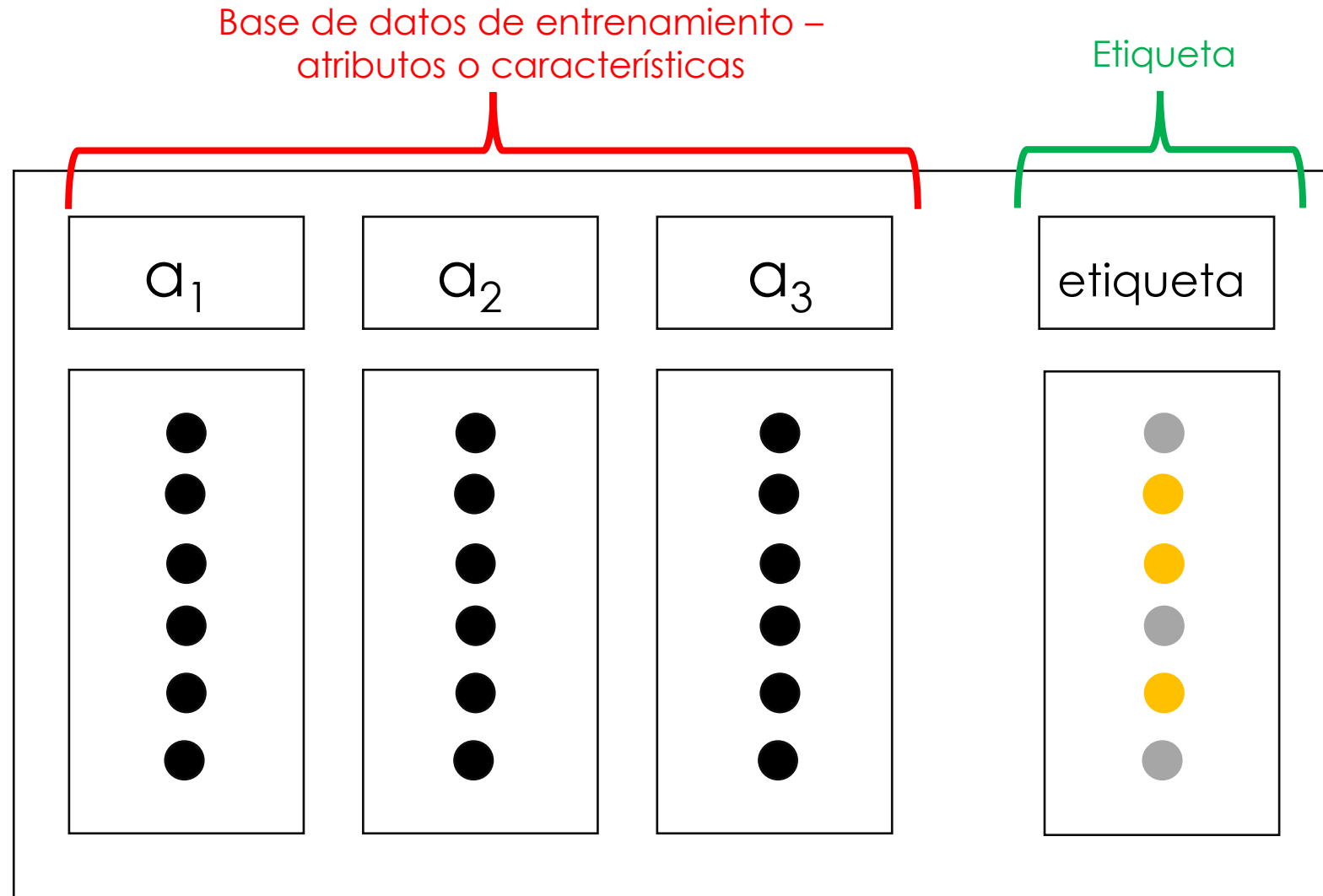


# ¿De donde partimos?





# ¿De donde partimos?



- Clase A
- Clase B

- Modelo Estadístico y Probabilístico

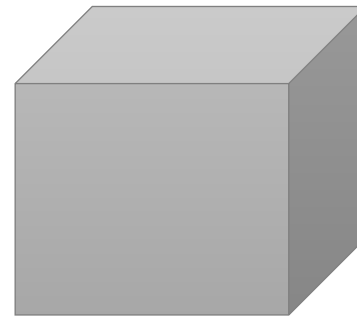


□ Tarea de **clasificación**.

- Modelo Estadístico y Probabilístico



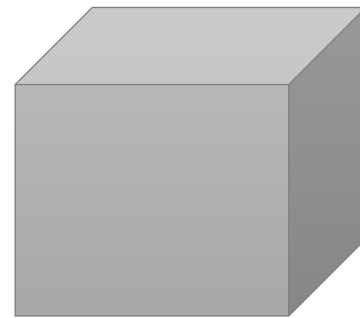
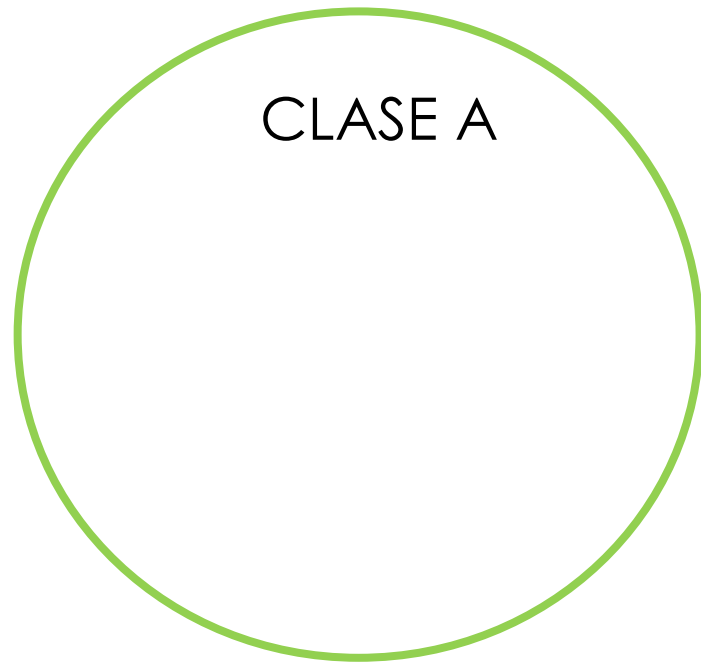
❑ Tarea de **clasificación**.



OBJETO



❑ Tarea de **clasificación**.

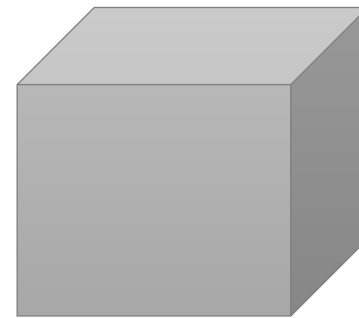
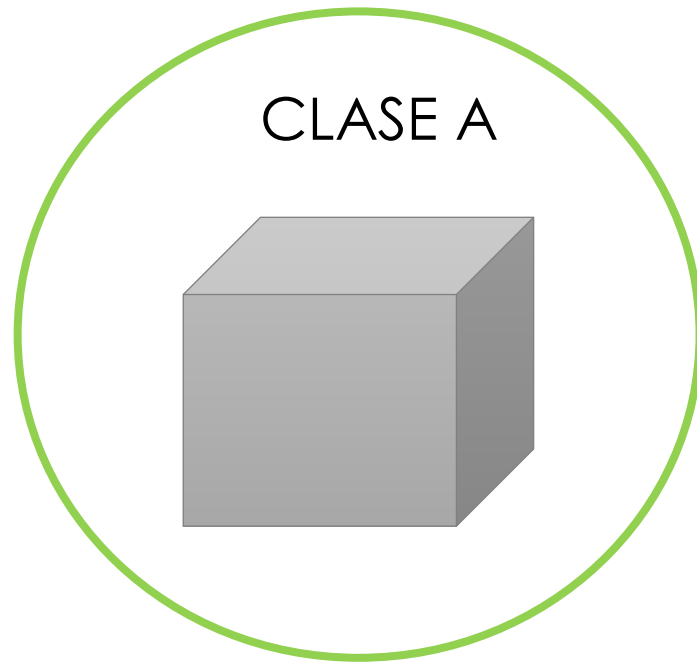


OBJETO





□ Tarea de **clasificación**.

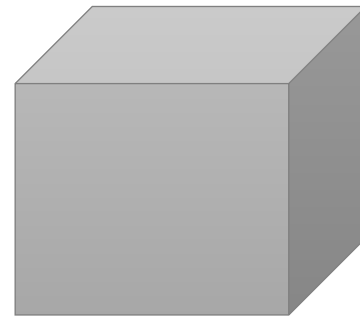
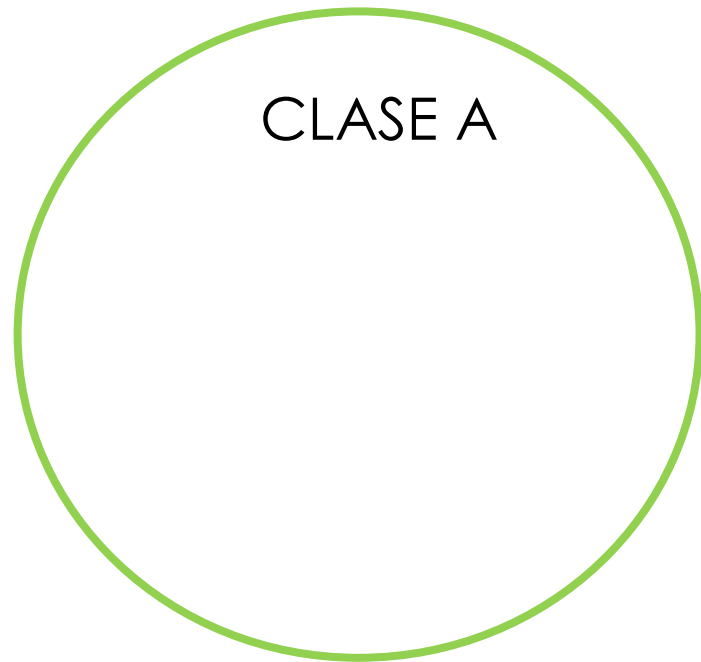


OBJETO

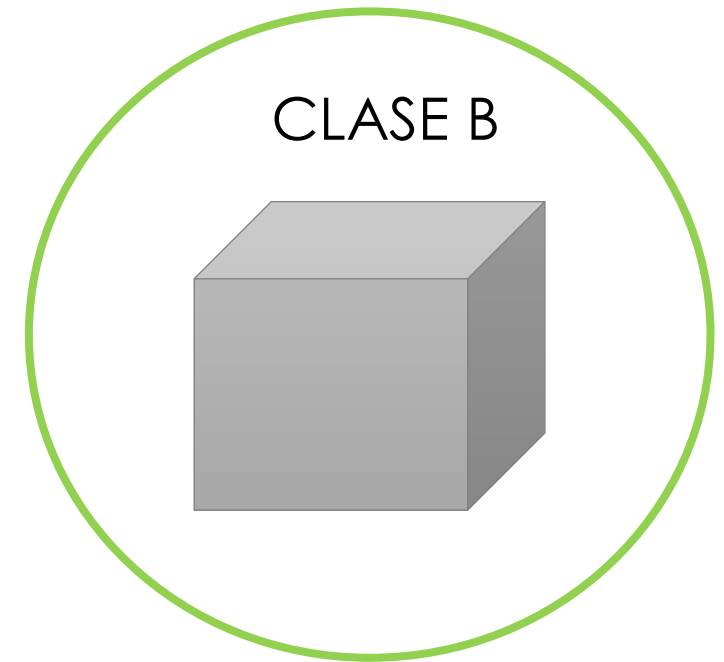




❑ Tarea de **clasificación**.

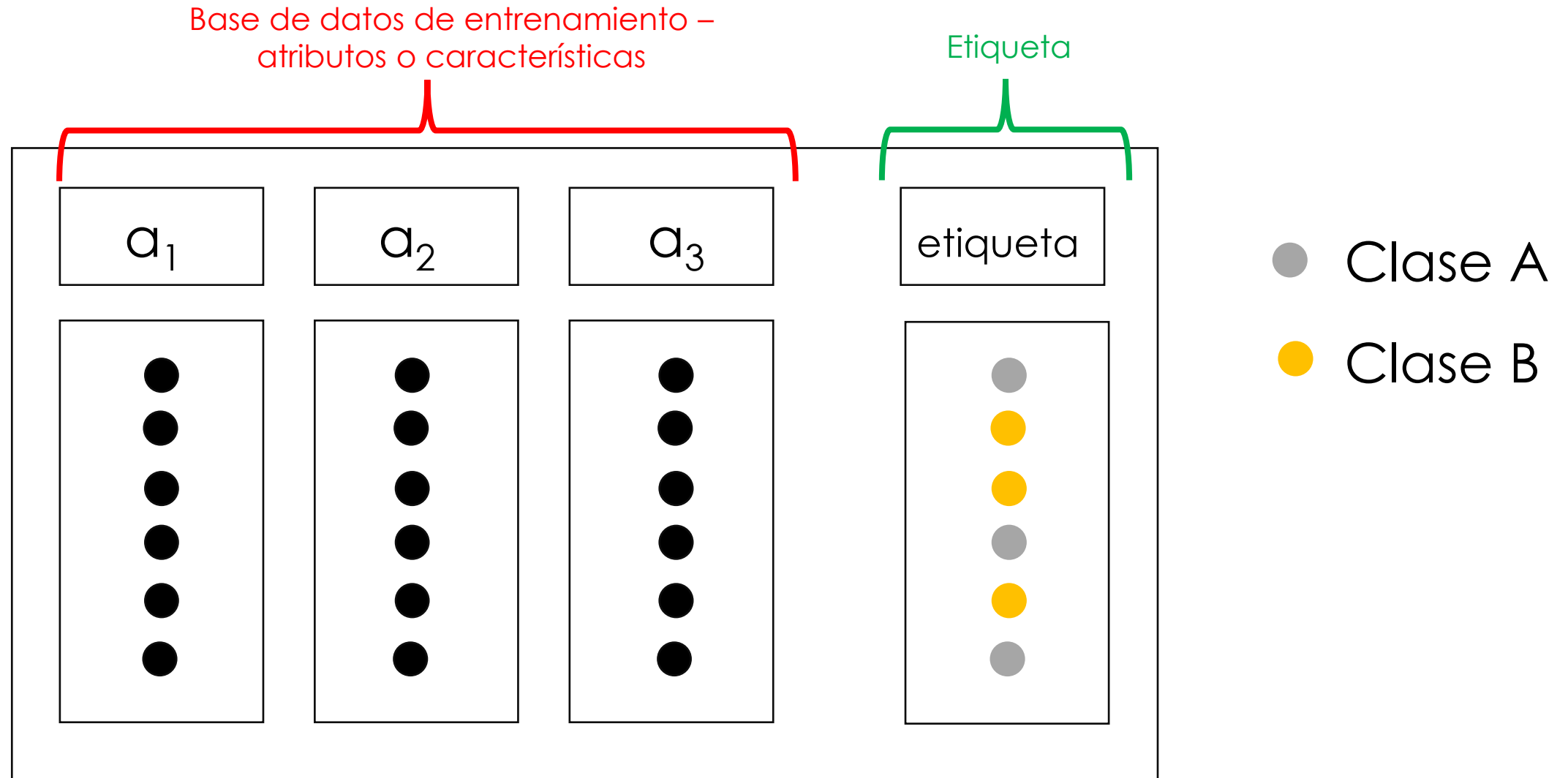


OBJETO





## ¿De donde partimos?







## **EJEMPLO**

- Modelo Estadístico y Probabilístico



Tabela adaptada de: Quinlan, J.R. Induction of decision trees.  
Machine Learning, 81-106 (1986).  
<https://doi.org/10.1007/BF00116251>

	outlook	temperature	humidity	windy	class
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rainy	mild	high	false	yes
5	rainy	cool	normal	false	yes
6	rainy	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rainy	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rainy	mild	high	true	no

# Modelo Estadístico y Probabilístico



Tabela adaptada de: Quinlan, J.R. Induction of decision trees.  
Machine Learning, 81-106 (1986).  
<https://doi.org/10.1007/BF00116251>

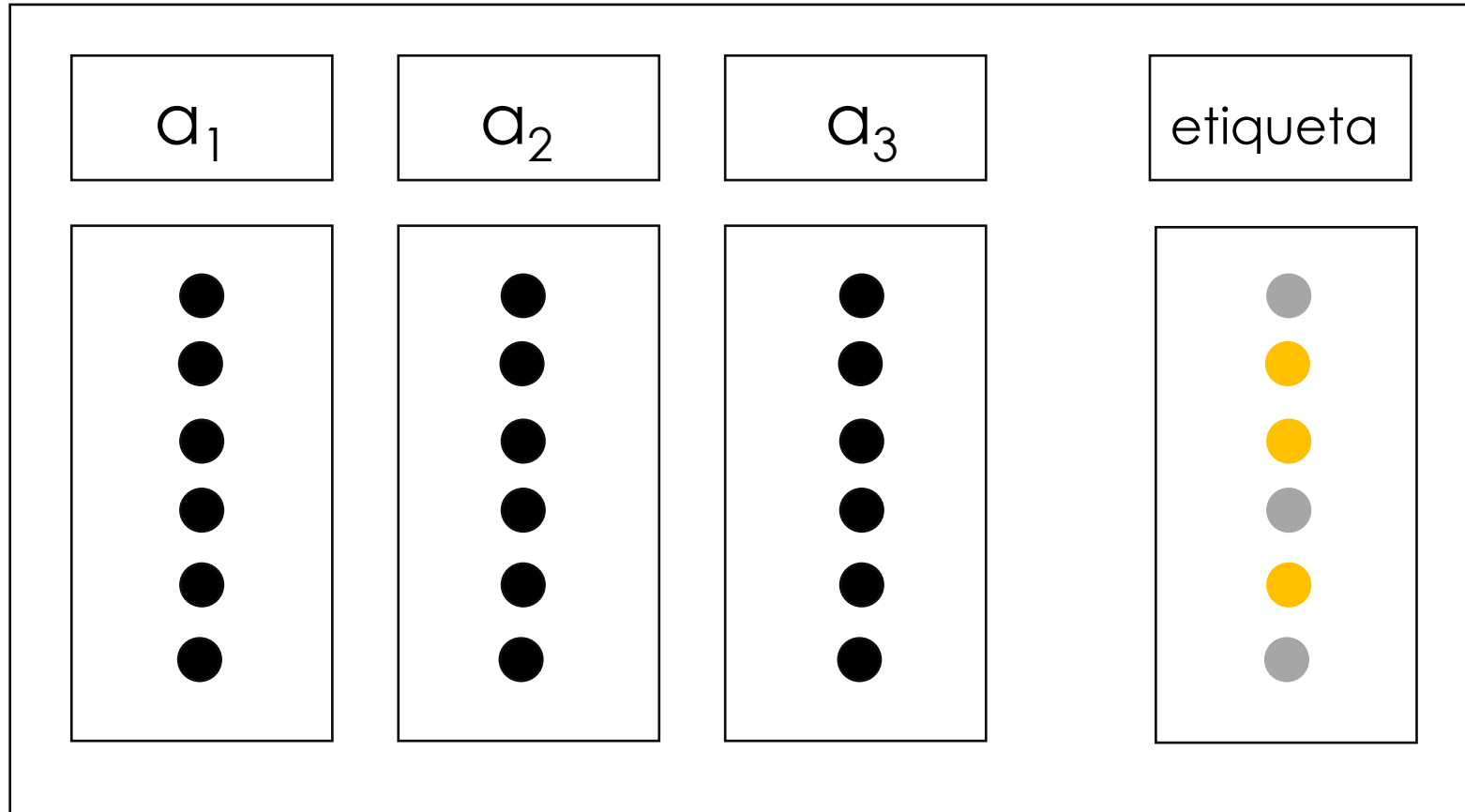
	outlook	temperature	humidity	windy	class
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rainy	mild	high	false	yes
5	rainy	cool	normal	false	yes
6	rainy	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rainy	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rainy	mild	high	true	no

15	rainy	mild	normal	false	?
----	-------	------	--------	-------	---

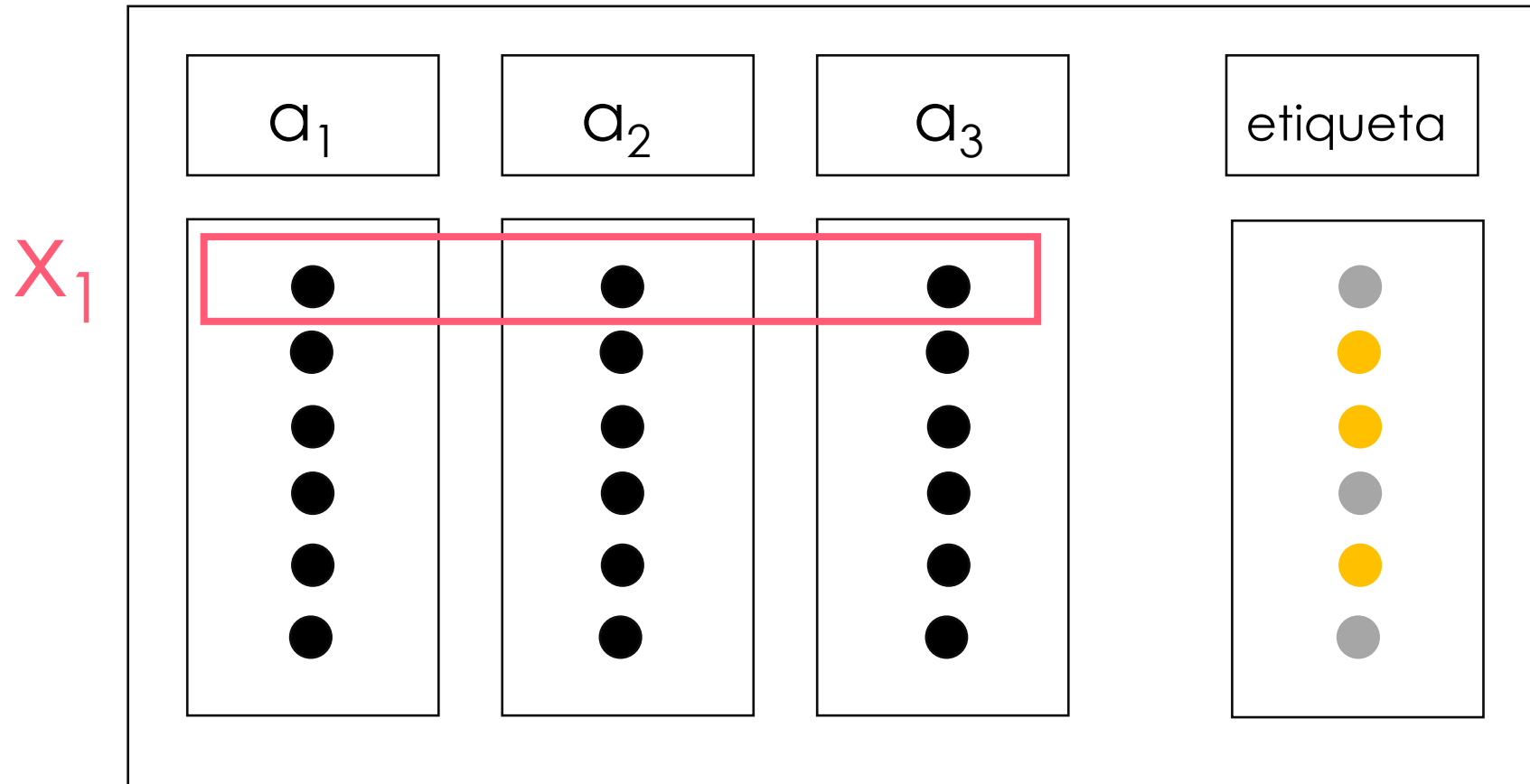


# **¿Cómo aprendemos probabilidades en tareas de NLP?**

- 



- Modelo Estadístico y Probabilístico





### 3. Problema del Modelamiento del Lenguaje

- Tenemos un conjunto finito de términos  $\mathbf{V}$ .

$$\mathcal{V} = \{\text{the, a, man, telescope, Beckham, two, ...}\}$$

- Tenemos un conjunto de oraciones (*strings*) que se pueden formar  $\mathbf{V}^+$ .

the STOP

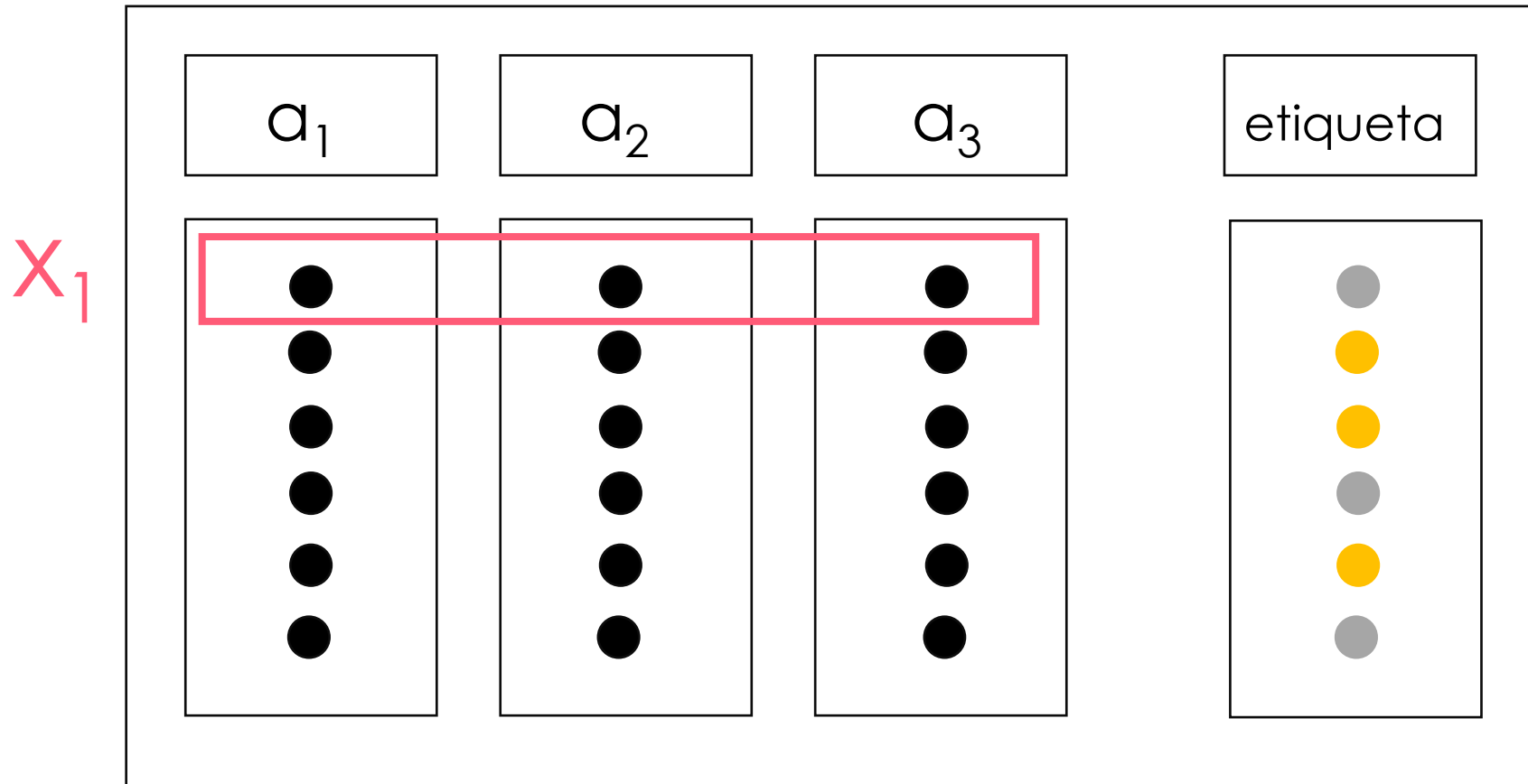
a STOP

the fan STOP

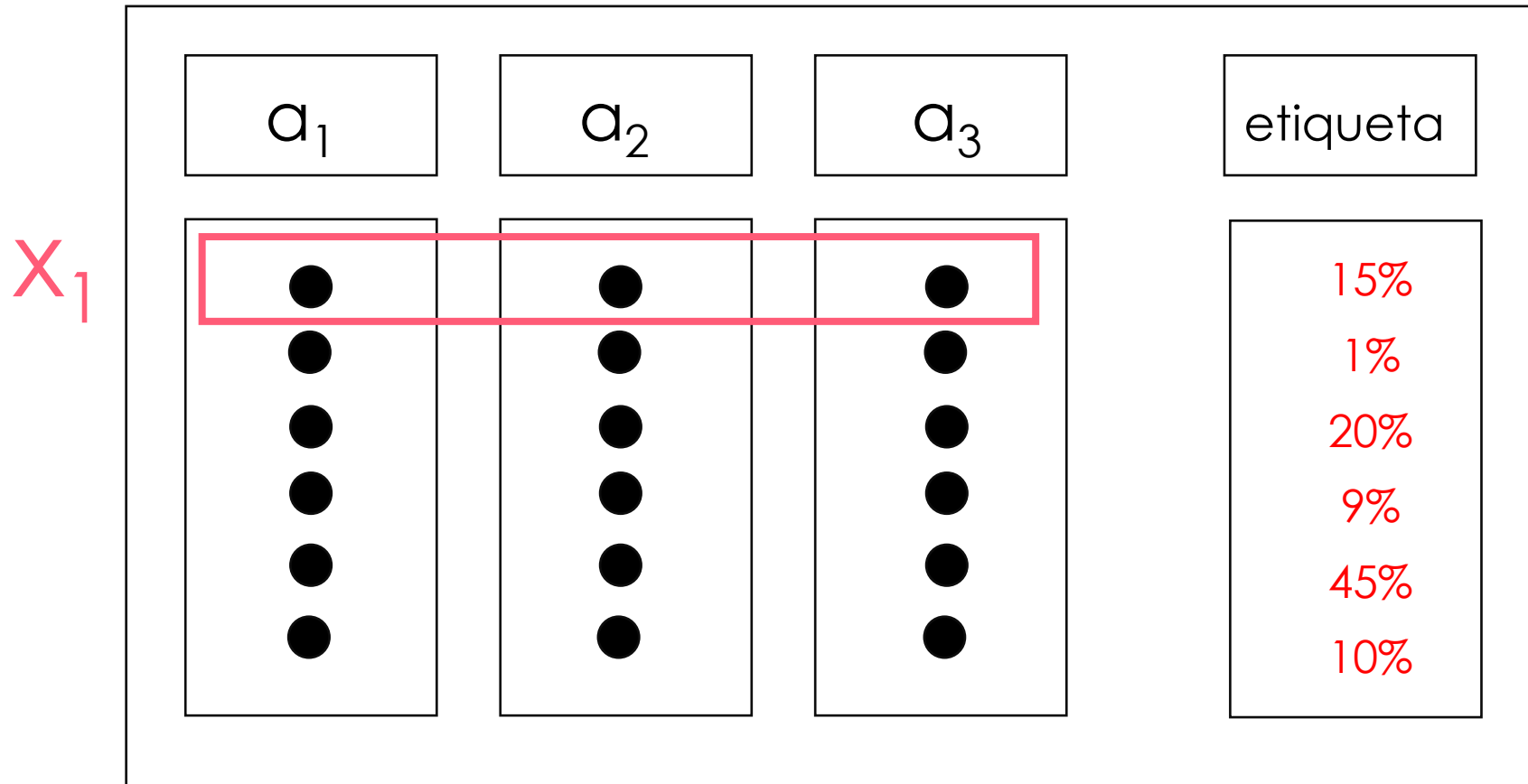
the fan saw Beckham STOP

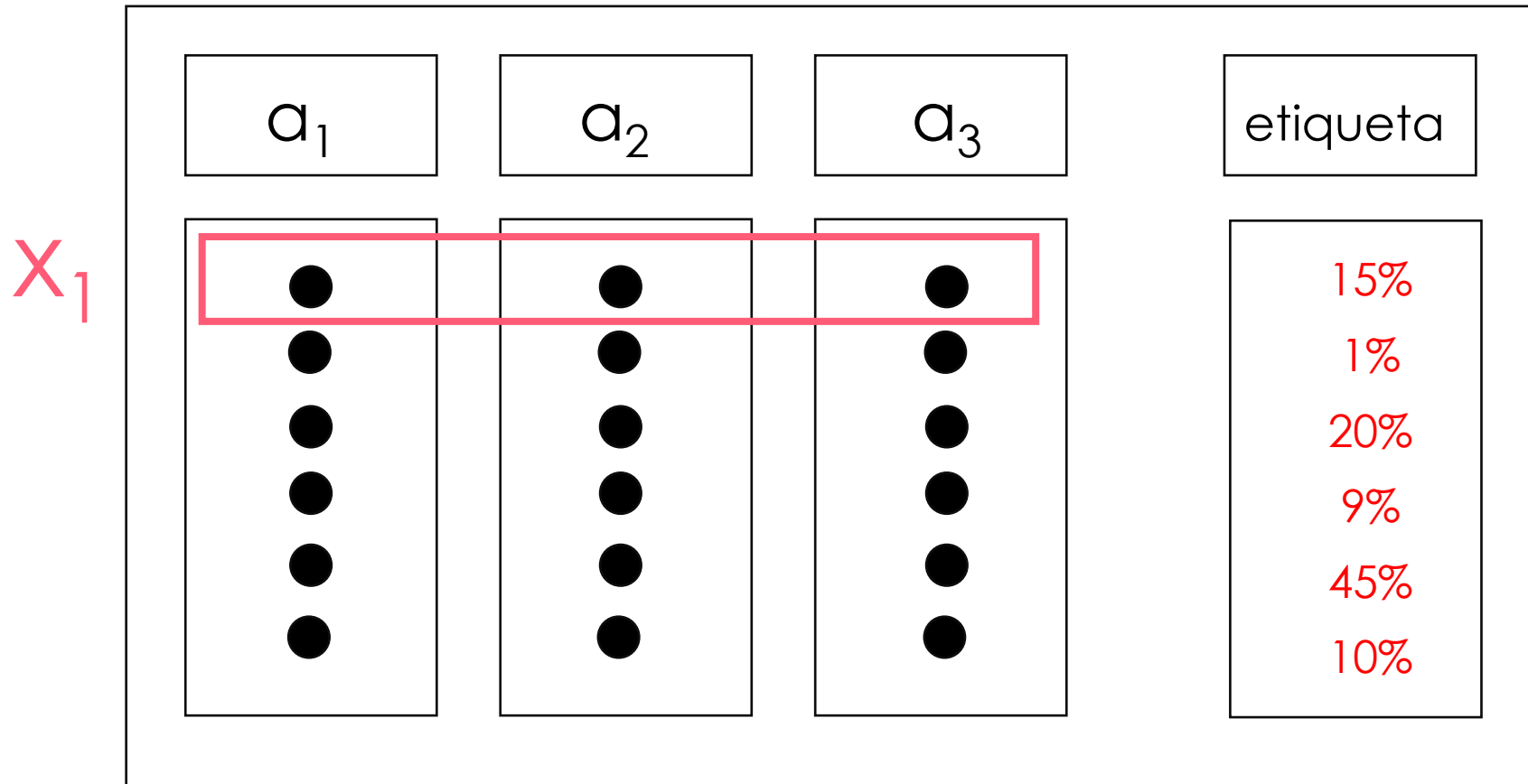
the fan saw saw STOP

the fan saw Beckham play for Real Madrid STOP

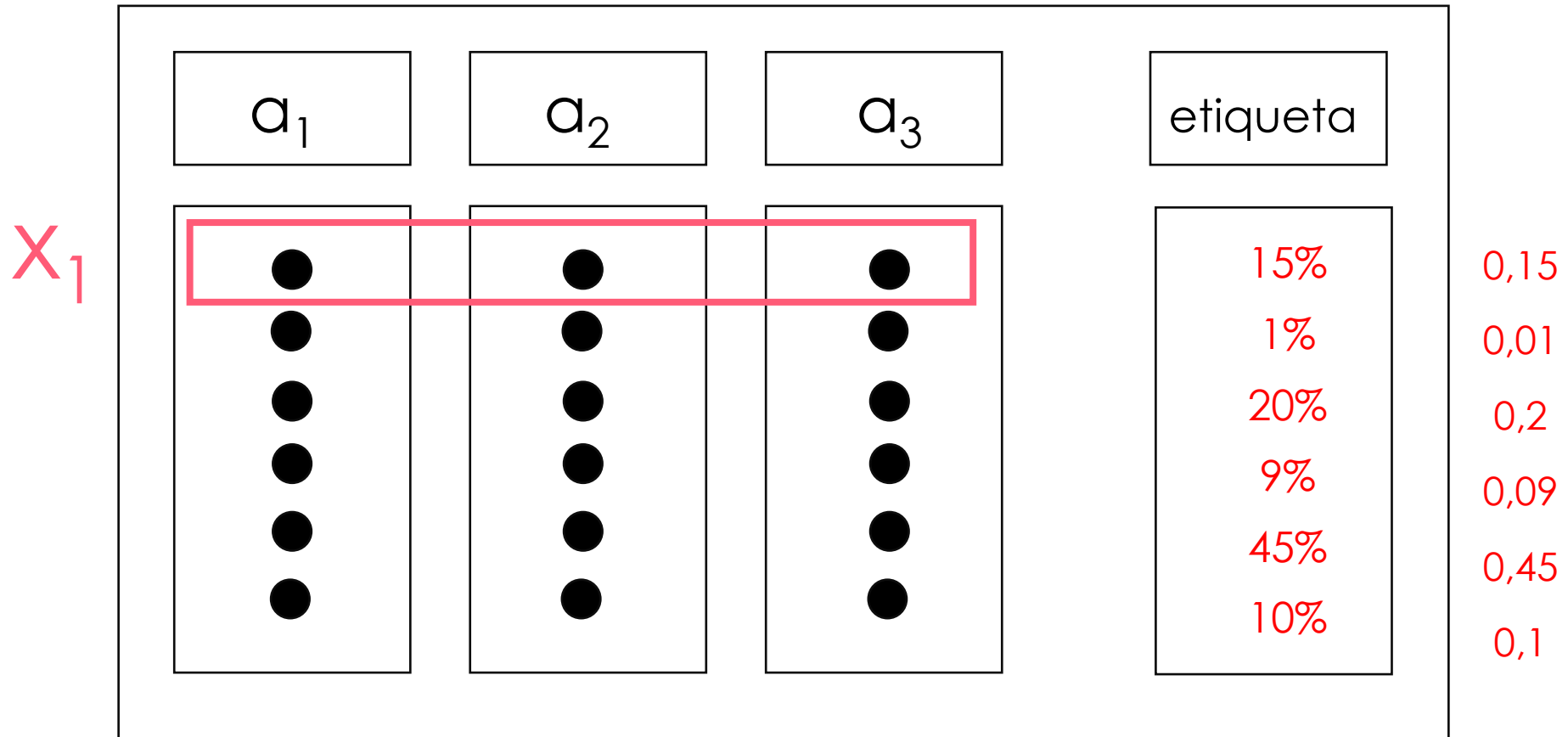




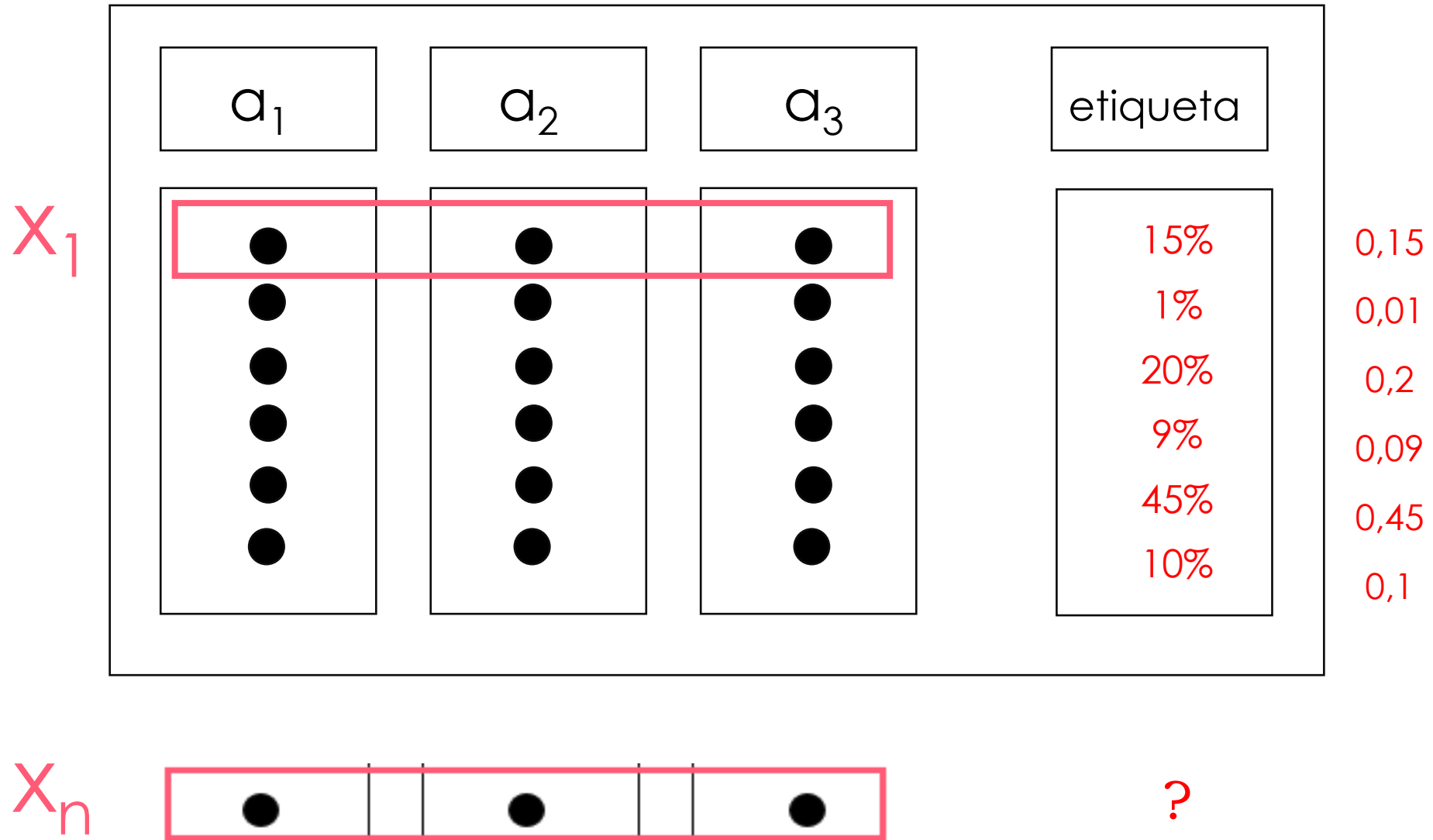




$$\sum_{x \in \mathcal{V}^\dagger} p(x) = 1, \quad p(x) \geq 0 \text{ for all } x \in \mathcal{V}^\dagger$$



$$\sum_{x \in \mathcal{V}^\dagger} p(x) = 1, \quad p(x) \geq 0 \text{ for all } x \in \mathcal{V}^\dagger$$



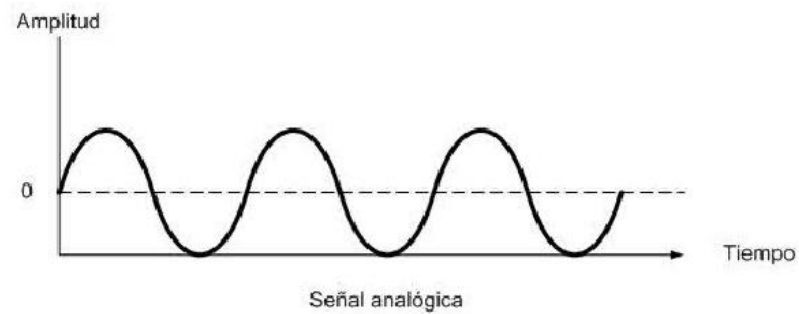


**¿En que nos ayuda aprender la  
probabilidad?**



## **Problema:** Speech Recognition (reconocimiento de voz)

**INPUT**

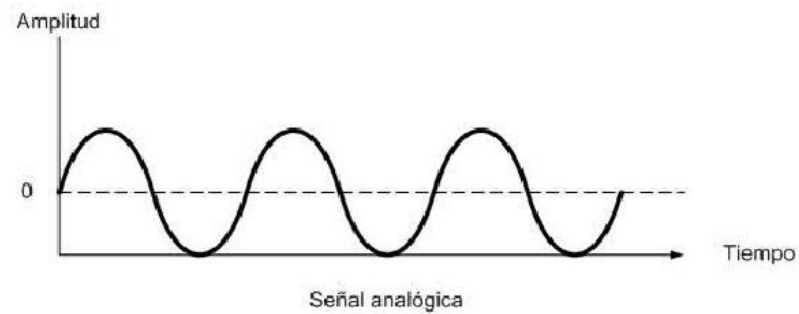


**1 entrada**



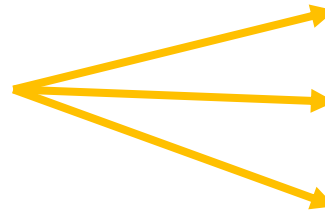
## Problema: Speech Recognition (reconocimiento de voz)

INPUT



1 entrada

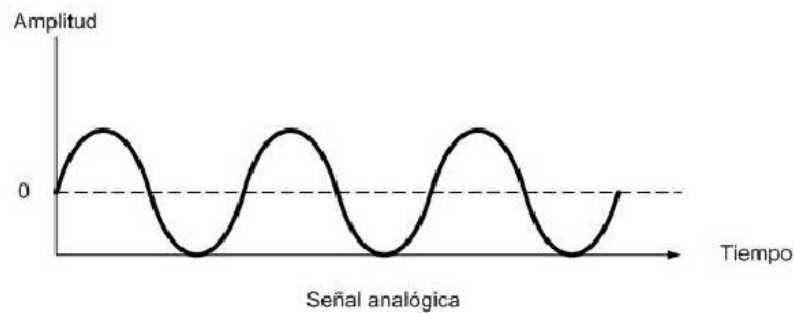
OUTPUT





**Problema:** Speech Recognition (reconocimiento de voz)

**INPUT**



**1 entrada**

**OUTPUT**

$O_1$ : "Las palabras parónimas"

$O_2$ : "Mas descalabras anónimas"

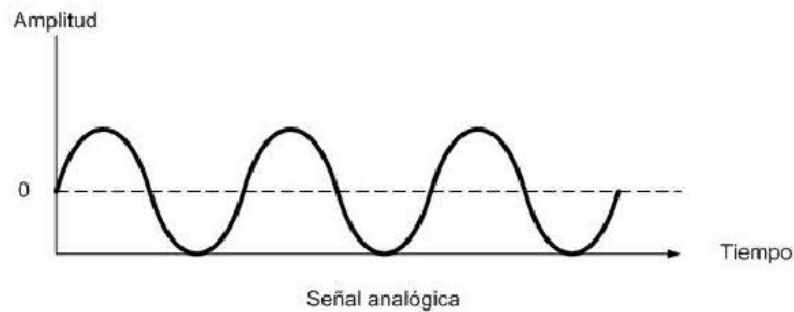
$O_3$ : "Las reabras homónimas"





**Problema:** Speech Recognition (reconocimiento de voz)

**INPUT**



**1 entrada**

**OUTPUT**

$O_1$ : "Las palabras parónimas"

**60%**

$O_2$ : "Mas descalabras anónimas"

**15%**

$O_3$ : "Las reabras homónimas"

**25%**

- Modelo Estadístico y Probabilístico



## **4. Repaso de Conceptos Probabilísticos**

- Modelo Estadístico y Probabilístico



## 4. Repaso de Conceptos Probabilísticos

**$P(A)$**

- Modelo Estadístico y Probabilístico



## 4. Repaso de Conceptos Probabilísticos

**P(A)**

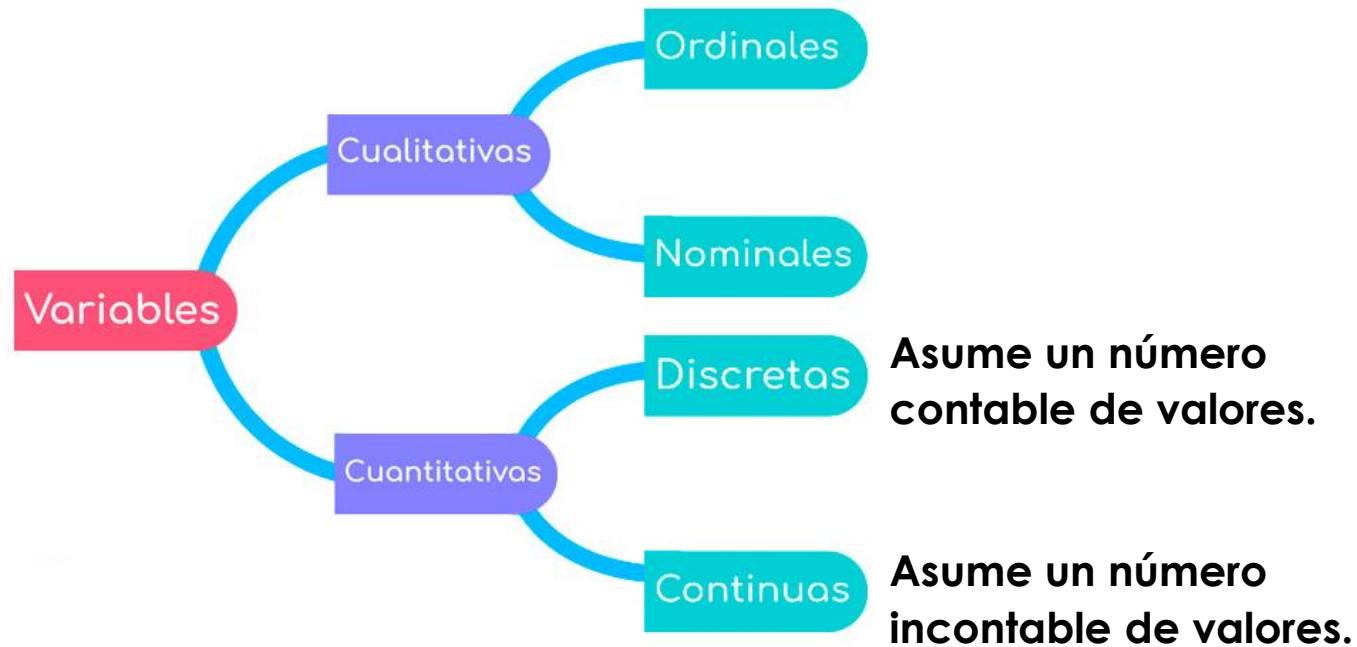
Variable aleatoria **discreta**



## 4. Repaso de Conceptos Probabilísticos

**$P(A)$**

Variable aleatoria **discreta**

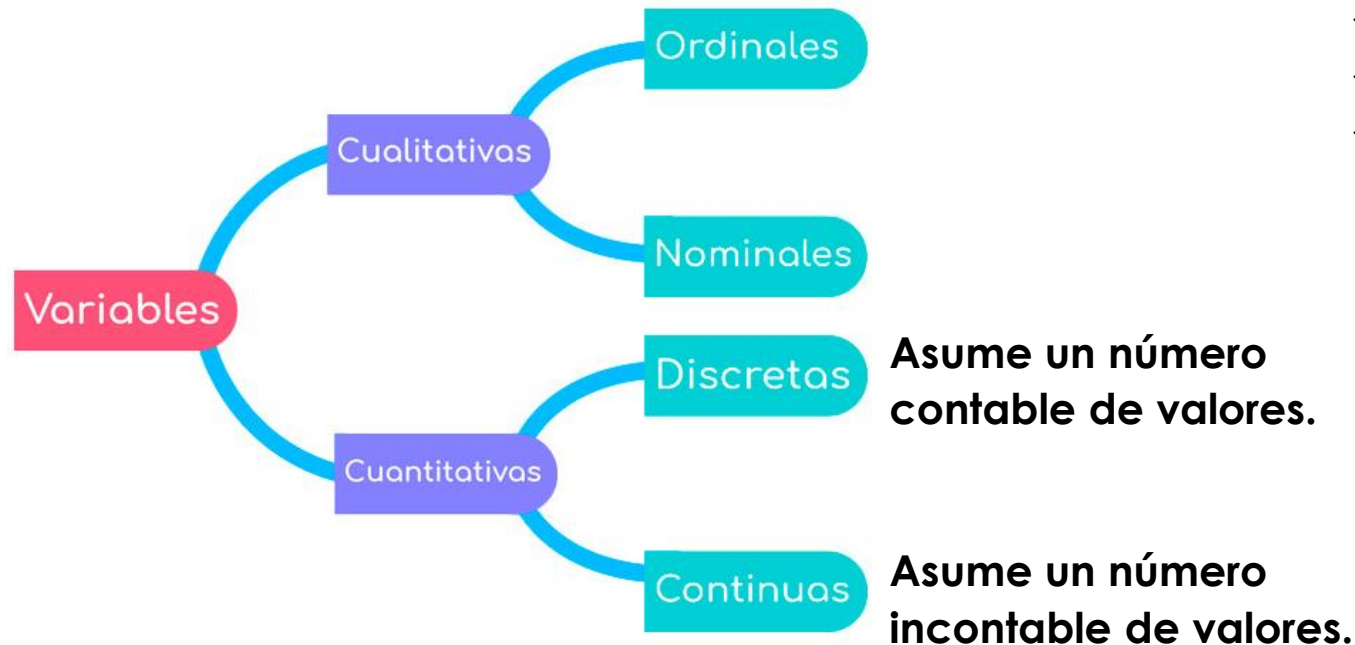




## 4. Repaso de Conceptos Probabilísticos

# $P(A)$

Variable aleatoria **discreta**



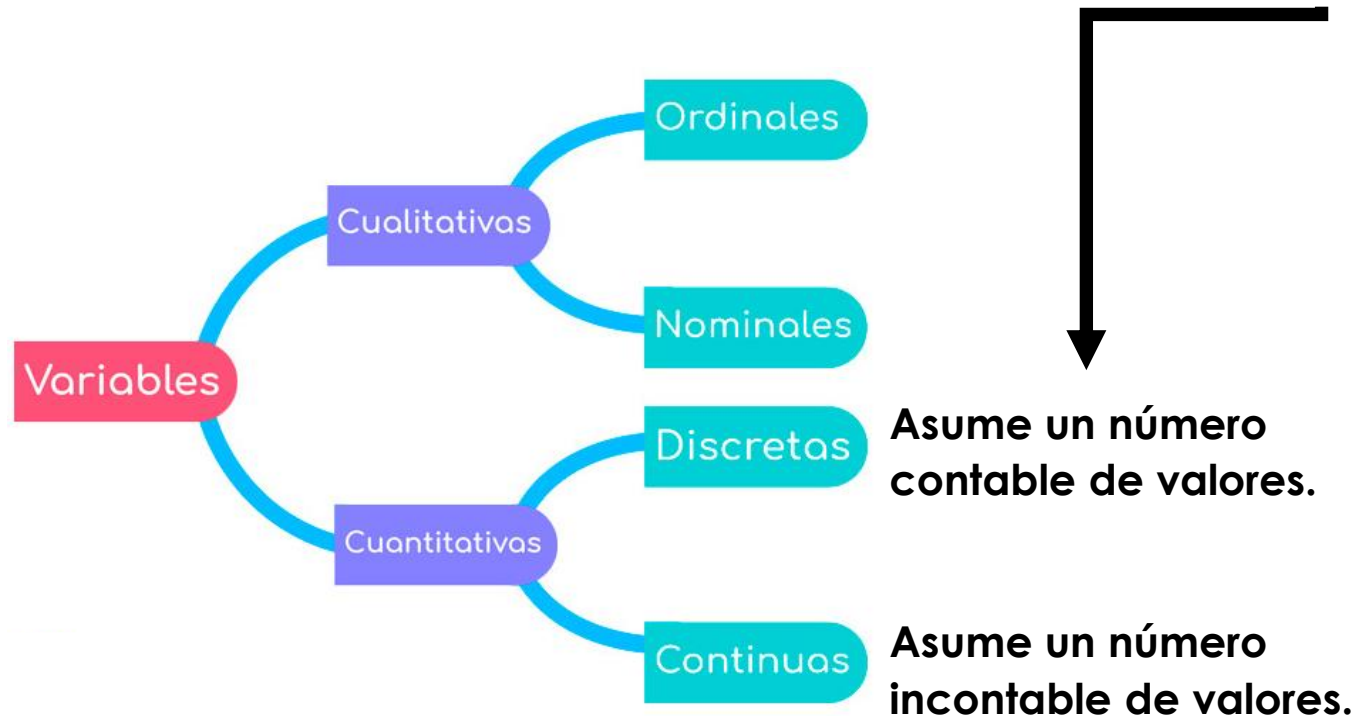
- ❖ El número de hijos de una familia.
- ❖ La cantidad de dedos que tiene una mano.
- ❖ El número de animales de una granja.
- ❖ Cantidad de empleados de una tienda.



## 4. Repaso de Conceptos Probabilísticos

**P(A)**

Variable aleatoria **discreta**



- ❖ El número de hijos de una familia.
- ❖ La cantidad de dedos que tiene una mano.
- ❖ El número de animales de una granja.
- ❖ Cantidad de empleados de una tienda.

- ❖ La estatura de un amigo.
- ❖ El ancho de una pelota.
- ❖ El volumen de agua en una piscina.
- ❖ El peso de una persona. (velocidad, tiempo, longitud)

- Modelo Estadístico y Probabilístico



## 4. Repaso de Conceptos Probabilísticos

**$P(A)$**

Variable **aleatoria** discreta





## 4. Repaso de Conceptos Probabilísticos

**$P(A)$**

Variable **aleatoria** discreta

Depende del azar o la suerte.





#### 4. Repaso de Conceptos Probabilísticos

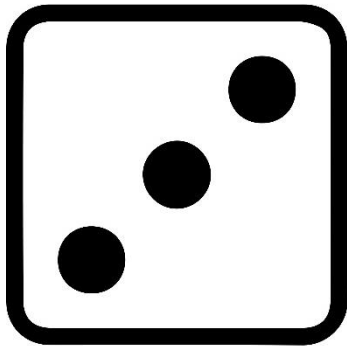
$$P(A) = \frac{\text{Casos Favorables}}{\text{Casos Totales}}$$



#### 4. Repaso de Conceptos Probabilísticos

$$P(A) = \frac{\text{Casos Favorables}}{\text{Casos Totales}}$$

$$P(\text{dado} = 3) =$$

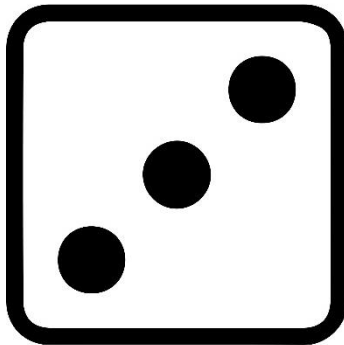




#### 4. Repaso de Conceptos Probabilísticos

$$P(A) = \frac{\text{Casos Favorables}}{\text{Casos Totales}}$$

$$P(\text{dado} = 3) = \frac{1}{6}$$





#### 4. Repaso de Conceptos Probabilísticos

Probabilidad Condicional

$$P(A | B)$$

$$P(\text{dado} = 3 | \text{impar}) =$$



#### 4. Repaso de Conceptos Probabilísticos

Probabilidad Condicional

$$P(A | B)$$

$$P(\text{dado} = 3 | \text{impar}) = \frac{1}{3}$$

- Modelo Estadístico y Probabilístico



## **4. Método Básico**



## 4. Método Básico

- Dado mi CORPUS de entrenamiento **N (cantidad total de oraciones)**.





## 4. Método Básico

- Dado mi CORPUS de entrenamiento **N** (cantidad total de oraciones).
- Para cualquier oración  $\mathbf{x}_1$  tengo una función  $\mathbf{c}(\mathbf{x}_1)$  que cuenta el número de veces que esta oración fue vista en nuestros datos de entrenamiento.



## 4. Método Básico

- Dado mi CORPUS de entrenamiento **N** (cantidad total de oraciones).
- Para cualquier oración  $\mathbf{x}_1$  tengo una función  $\mathbf{c}(\mathbf{x}_1)$  que cuenta el número de veces que esta oración fue vista en nuestros datos de entrenamiento.

$$P(\mathbf{x}_1) =$$



## 4. Método Básico

- Dado mi CORPUS de entrenamiento **N** (cantidad total de oraciones).
- Para cualquier oración  $\mathbf{x}_1$  tengo una función  $\mathbf{c}(\mathbf{x}_1)$  que cuenta el número de veces que esta oración fue vista en nuestros datos de entrenamiento.

$$P(\mathbf{x}_1) = \mathbf{c}(\mathbf{x}_1)$$



## 4. Método Básico

- Dado mi CORPUS de entrenamiento **N** (cantidad total de oraciones).
- Para cualquier oración  $\mathbf{x}_1$  tengo una función  $\mathbf{c}(\mathbf{x}_1)$  que cuenta el número de veces que esta oración fue vista en nuestros datos de entrenamiento.

$$P(\mathbf{x}_1) = \frac{\mathbf{c}(\mathbf{x}_1)}{N}$$

- Modelo Estadístico y Probabilístico



## 5. Método Trigramas

,



## 5. Método Trigramas

PROCESOS DE MARKOV



## 5. Método Trigramas

### PROCESOS DE MARKOV

Tipo de proceso ESTOCASTICO, el cual tiene una secuencia de variables aleatorias  $X_1, X_2, \dots, X_n$ .



## 5. Método Trigramas

### PROCESOS DE MARKOV

Tipo de proceso ESTOCASTICO, el cual tiene una secuencia de variables aleatorias  $X_1, X_2, \dots, X_n$ .

$X_n$  cada variable aleatoria es el proceso de sacar una palabra del vocabulario



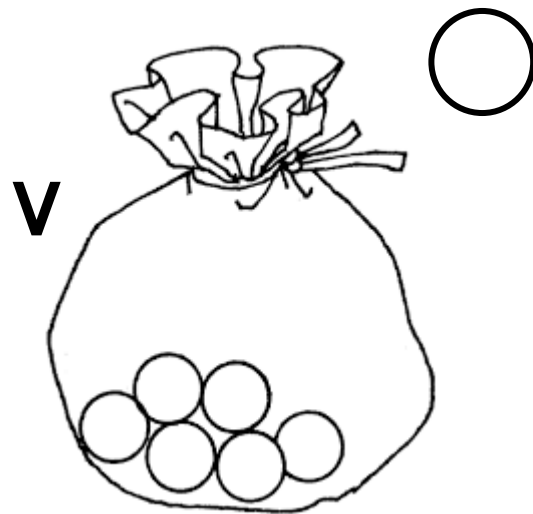


## 5. Método Trigramas

### PROCESOS DE MARKOV

Tipo de proceso ESTOCASTICO, el cual tiene una secuencia de variables aleatorias  $X_1, X_2, \dots, X_n$ .

$X_n$  cada variable aleatoria es el proceso de sacar una palabra del vocabulario



- Modelo Estadístico y Probabilístico



## 5. Método Trigramas

### PROCESOS DE MARKOV

Entonces una **oración** se puede ver como un proceso aleatorio.



## 5. Método Trigramas

### PROCESOS DE MARKOV

Entonces una **oración** se puede ver como un proceso aleatorio.

**“EL PERRO LADRA FUERTE”**

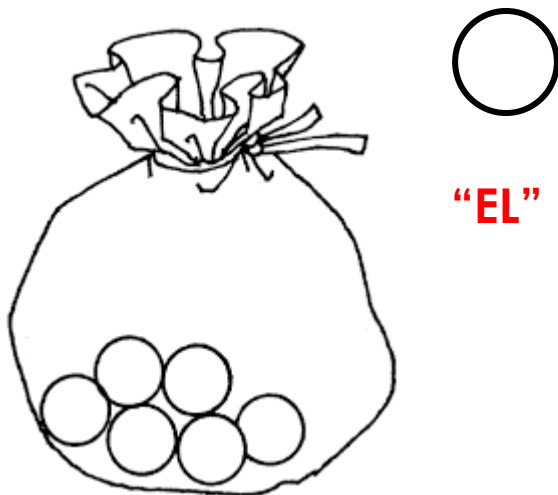


## 5. Método Trigramas

### PROCESOS DE MARKOV

Entonces una oración se puede ver como un proceso aleatorio.

**“EL PERRO LADRA FUERTE”**



En donde saco una palabra aleatoria, saco la siguiente palabra aleatoria, luego la siguiente y que cada proceso de sacar una palabra aleatoria pueda estar condicionada a la palabras anteriores para que así ojala las oraciones con mas sentido tengan mayor probabilidad.

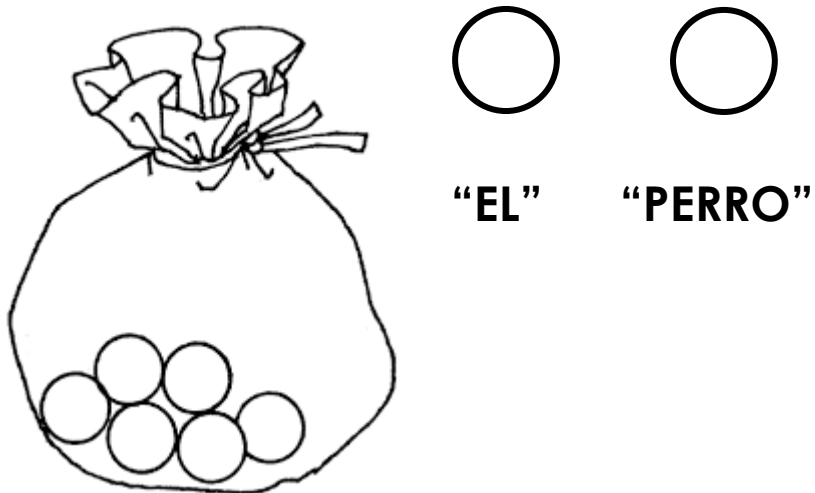


## 5. Método Trigramas

### PROCESOS DE MARKOV

Entonces una oración se puede ver como un proceso aleatorio.

**“EL PERRO LADRA FUERTE”**



En donde saco una palabra aleatoria, saco la siguiente palabra aleatoria, luego la siguiente y que cada proceso de sacar una palabra aleatoria pueda estar condicionada a la palabras anteriores para que así ojala las oraciones con mas sentido tengan mayor probabilidad.

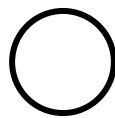
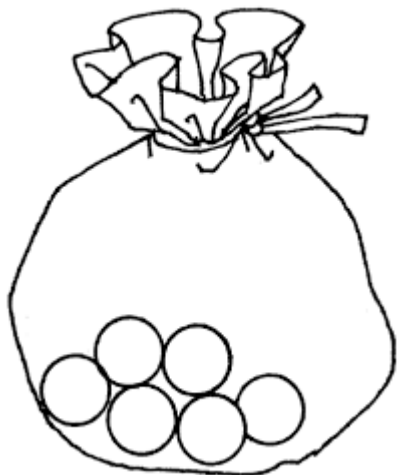


## 5. Método Trigramas

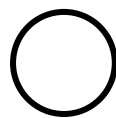
### PROCESOS DE MARKOV

Entonces una oración se puede ver como un proceso aleatorio.

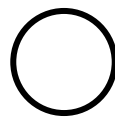
**“EL PERRO LADRA FUERTE”**



**“EL”**



**“PERRO”**



**“LADRA”**

En donde saco una palabra aleatoria, saco la siguiente palabra aleatoria, luego la siguiente y que cada proceso de sacar una palabra aleatoria pueda estar condicionada a la palabras anteriores para que así ojala las oraciones con mas sentido tengan mayor probabilidad.

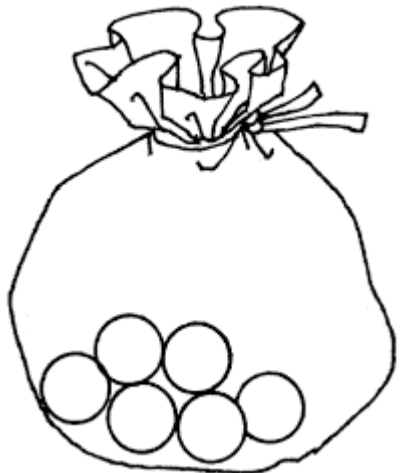


## 5. Método Trigramas

### PROCESOS DE MARKOV

Entonces una oración se puede ver como un proceso aleatorio.

**“EL PERRO LADRA FUERTE”**



○ ○ ○ ○  
**“EL” “PERRO” “LADRA” “FUERTE”**

En donde saco una palabra aleatoria, saco la siguiente palabra aleatoria, luego la siguiente y que cada proceso de sacar una palabra aleatoria pueda estar condicionada a la palabras anteriores para que así ojala las oraciones con mas sentido tengan mayor probabilidad.



## 5. Método Trigramas

### PROCESOS DE MARKOV

Nuestro objetivo es tener un modelo así:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$





## 5. Método Trigramas

### PROCESOS DE MARKOV

Nuestro objetivo es tener un modelo así:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Normalmente, se asume que la longitud  $n$  es fijo (con la idea de que todas las oraciones tengan el mismo numero de palabras).



## 5. Método Trigramas

### PROCESOS DE MARKOV

Nuestro objetivo es tener un modelo así:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Normalmente, se asume que la longitud  $n$  es fijo (con la idea de que todas las oraciones tengan el mismo numero de palabras).

Ejemplo: **“EL PERRO LADRA FUERTE”**



## 5. Método Trigramas

### PROCESOS DE MARKOV

Nuestro objetivo es tener un modelo así:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Normalmente, se asume que la longitud  $n$  es fijo (con la idea de que todas las oraciones tengan el mismo numero de palabras).

Ejemplo: **“EL PERRO LADRA FUERTE”**

$$P(X_1 = \text{“El”}, X_2 = \text{“perro”}, X_3 = \text{“ladra”}, X_4 = \text{“fuerte”})$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

Nuestro objetivo es tener un modelo así:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Normalmente, se asume que la longitud **n** es fijo (con la idea de que todas las oraciones tengan el mismo numero de palabras).

Ejemplo: **“EL PERRO LADRA FUERTE”**

$$P(X_1 = \text{“El”}, X_2 = \text{“perro”}, X_3 = \text{“ladra”}, X_4 = \text{“fuerte”})$$





## 5. Método Trigramas

### PROCESOS DE MARKOV

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Probabilidad Conjunta



## 5. Método Trigramas

### PROCESOS DE MARKOV

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Probabilidad Conjunta

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



## 5. Método Trigramas

### PROCESOS DE MARKOV

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Probabilidad Conjunta

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

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





## 5. Método Trigramas

### PROCESOS DE MARKOV

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Probabilidad Conjunta

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

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

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



## 5. Método Trigramas

### PROCESOS DE MARKOV

$$\underbrace{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)}_{\text{Probabilidad Conjunta}}$$

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

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

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





## 5. Método Trigramas

### PROCESOS DE MARKOV

$$\underbrace{P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)}_{\text{Probabilidad Conjunta}}$$

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

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

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



$$P(X_1 = \text{"El"}, X_2 = \text{"perro"}, X_3 = \text{"ladra"}, X_4 = \text{"fuerte"})$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Probabilidad Conjunta

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

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

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

**¿Hay algún problema aquí?**

$$P(X_1 = \text{"El"}, X_2 = \text{"perro"}, X_3 = \text{"ladra"}, X_4 = \text{"fuerte"})$$



- Modelo Estadístico y Probabilístico



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.

$$\begin{aligned} &P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \end{aligned}$$





## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.

$$\begin{aligned} &P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \end{aligned}$$

¿Qué dice Markov?



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.

$$\begin{aligned} &P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \end{aligned}$$

¿Qué dice Markov?

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



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.

$$\begin{aligned} &P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \end{aligned}$$

¿Qué dice Markov?

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

Es decir, solamente me importa donde estuve antes no me importa donde dos veces antes.



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.

$$\begin{aligned} &P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \end{aligned}$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_{i-1} = x_{i-1}) \end{aligned}$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Trata de simplificar estas probabilidades condicionales haciendo algunos supuestos de independencia.

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \\ &= P(X_1 = x_1) \prod_{i=2}^n P(X_i = x_i | X_{i-1} = x_{i-1}) \end{aligned}$$

Suposición de independencia = Solamente me importa un estado hacia atrás.



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Ejemplo :

$$P(C | A,B) = P(C | B) \rightarrow \text{Markov de Primer Orden}$$

Es decir, solamente me importa donde estuve antes no me importa dos veces antes (olvido eso).



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 1. Procesos de Markov de Primer Orden

Ejemplo :

$$P(C | A,B) = P(C | B) \rightarrow \text{Markov de Primer Orden}$$

Es decir, solamente me importa donde estuve antes no me importa dos veces antes (olvido eso).

$$P(X_1 = \text{"El"}, X_2 = \text{"perro"}, X_3 = \text{"ladra"}, X_4 = \text{"fuerte"})$$



- Modelo Estadístico y Probabilístico



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden

Proceso menos agresivo y solamente condiciono **dos palabras** hacia atrás.



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden

Proceso menos agresivo y solamente condiciono **dos palabras** hacia atrás.

$$P(D | A,B,C) = P(D | C,B) \rightarrow \text{Markov de Segundo Orden}$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden

Proceso menos agresivo y solamente condiciono **dos palabras** hacia atrás.

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden

Proceso menos agresivo y solamente condiciono **dos palabras** hacia atrás.

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ = & P(X_1 = x_1) \times P(X_2 = x_2 | X_1 = x_1) \\ & \times \prod_{i=3}^n P(X_i = x_i | X_{i-2} = x_{i-2}, X_{i-1} = x_{i-1}) \end{aligned}$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden

Proceso menos agresivo y solamente condiciono **dos palabras** hacia atrás.

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ = & P(X_1 = x_1) \times P(X_2 = x_2 | X_1 = x_1) \\ & \times \prod_{i=3}^n P(X_i = x_i | X_{i-2} = x_{i-2}, X_{i-1} = x_{i-1}) \\ = & \prod_{i=1}^n P(X_i = x_i | X_{i-2} = x_{i-2}, X_{i-1} = x_{i-1}) \end{aligned}$$



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden

Proceso menos agresivo y solamente condiciono **dos palabras** hacia atrás.

$$\begin{aligned} & P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ &= P(X_1 = x_1) \times P(X_2 = x_2 | X_1 = x_1) \\ & \quad \times \prod_{i=3}^n P(X_i = x_i | X_{i-2} = x_{i-2}, X_{i-1} = x_{i-1}) \\ &= \prod_{i=1}^n P(X_i = x_i | X_{i-2} = x_{i-2}, X_{i-1} = x_{i-1}) \end{aligned}$$

(For convenience we assume  $x_0 = x_{-1} = *$ , where  $*$  is a special “start” symbol.)



## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden

Ejemplo:

$$P(X_1 = \text{"El"}, X_2 = \text{"perro"}, X_3 = \text{"ladra"}, X_4 = \text{"fuerte"})$$





## 5. Método Trigramas

### PROCESOS DE MARKOV

#### 2. Procesos de Markov de Segundo Orden

Ejemplo:

$$P(X_1 = \text{"El"}, X_2 = \text{"perro"}, X_3 = \text{"ladra"}, X_4 = \text{"fuerte"})$$

- Tenemos un conjunto de oraciones (*strings*) que se pueden formar  $V^+$ .

the STOP      Se usa un símbolo especial para identificar  
a STOP      el fin de la oración (token especial).  
the fan STOP  
the fan saw Beckham STOP  
the fan saw saw STOP  
the fan saw Beckham play for Real Madrid STOP

- Modelo Estadístico y Probabilístico



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

- Modelo Estadístico y Probabilístico



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Ejemplo: El perro ladra fuerte



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Ejemplo: El perro ladra fuerte

$x_{-1} = *$

$x_0 = *$

$x_1 = \text{el}$

$x_2 = \text{perro}$

$x_3 = \text{ladra}$

$x_4 = \text{fuerte}$

$x_5 = \text{STOP}$



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Ejemplo: El perro ladra fuerte

$x_{-1} = *$

$x_0 = *$

$x_1 = \text{el}$

$x_2 = \text{perro}$

$x_3 = \text{ladra}$

$x_4 = \text{fuerte}$

$x_5 = \text{STOP}$



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Ejemplo: El perro ladra fuerte

$x_{-1} = *$

$x_0 = *$

$P(\text{El, perro, ladra, fuerte}) =$

$x_1 = \text{el}$

$P(\text{El} \mid *, *)$

$x_2 = \text{perro}$

$x_3 = \text{ladra}$

$x_4 = \text{fuerte}$

$x_5 = \text{STOP}$



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Ejemplo: El perro ladra fuerte

$x_{-1} = *$

$x_0 = *$

$P(\text{El, perro, ladra, fuerte}) =$

$x_1 = \text{el}$

$P(\text{El} \mid *, *) * P(\text{perro} \mid *, \text{El})$

$x_2 = \text{perro}$

$x_3 = \text{ladra}$

$x_4 = \text{fuerte}$

$x_5 = \text{STOP}$



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Ejemplo: El perro ladra fuerte

$x_{-1} = *$

$x_0 = *$

$x_1 = \text{el}$

$x_2 = \text{perro}$

$x_3 = \text{ladra}$

$x_4 = \text{fuerte}$

$x_5 = \text{STOP}$

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} \mid *, *) * P(\text{perro} \mid *, \text{El}) * P(\text{ladra} \mid \text{El, perro})$





## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Ejemplo: El perro ladra fuerte

$x_{-1} = *$

$x_0 = *$

$x_1 = \text{el}$

$x_2 = \text{perro}$

$x_3 = \text{ladra}$

$x_4 = \text{fuerte}$

$x_5 = \text{STOP}$

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} \mid *, *) * P(\text{perro} \mid *, \text{El}) * P(\text{ladra} \mid \text{El, perro}) * P(\text{fuerte} \mid \text{perro, ladra})$



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Ejemplo: El perro ladra fuerte

$x_{-1} = *$

$x_0 = *$

$x_1 = \text{el}$

$x_2 = \text{perro}$

$x_3 = \text{ladra}$

$x_4 = \text{fuerte}$

$x_5 = \text{STOP}$

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} \mid *, *) * P(\text{perro} \mid *, \text{El}) * P(\text{ladra} \mid \text{El, perro}) * P(\text{fuerte} \mid \text{perro, ladra})$   
 $* P(\text{STOP} \mid \text{ladra, fuerte})$



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Consiste en:

- Tener un conjunto finito  $V$
- Un parámetro  $q(w | u, v)$  para cada **trigrama**  $u, v, w$  donde cada  $w \in V \cup \{STOP\}$  y donde  $u, v \in V \cup \{*\}$ .



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

Consiste en:

- Tener un conjunto finito  $V$
- Un parámetro  $q(w | u, v)$  para cada **trigrama**  $u, v, w$  donde cada  $w \in V \cup \{\text{STOP}\}$  y donde  $u, v \in V \cup \{*\}$ .

Entonces la idea es **APRENDER** esos  $q$  (esas probabilidades) dado un CORPUS de entrenamiento = **ESTIMACIÓN DE PARAMETROS**.

- Modelo Estadístico y Probabilístico



## 5. Método Trigramas

MODELO DE TRIGRAMAS

Estimación de Parámetros



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

#### Estimación de Parámetros

$$q(w_i \mid w_{i-2}, w_{i-1})$$



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

#### Estimación de Parámetros

$$q(w_i \mid w_{i-2}, w_{i-1})$$

#### Estimación de Máxima Similitud



## 5. Método Trigramas

### MODELO DE TRIGRAMAS

#### Estimación de Parámetros

$$q(w_i \mid w_{i-2}, w_{i-1})$$

#### Estimación de Máxima Similitud

$$q(w_i \mid w_{i-2}, w_{i-1}) = \frac{\text{Count}(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})}$$



- Modelo Estadístico y Probabilístico



# EJEMPLO



# EJEMPLO

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} \mid *, *) * P(\text{perro} \mid *, \text{El}) * P(\text{ladra} \mid \text{El, perro}) * P(\text{fuerte} \mid \text{perro, ladra}) * P(\text{STOP} \mid \text{ladra, fuerte})$



# EJEMPLO

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} \mid *, *) * P(\text{perro} \mid *, \text{El}) * P(\text{ladra} \mid \text{El, perro}) * P(\text{fuerte} \mid \text{perro, ladra}) * P(\text{STOP} \mid \text{ladra, fuerte})$



# EJEMPLO

$$q(w_i | w_{i-2}, w_{i-1}) = \frac{\text{Count}(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})}$$

P(El, perro, ladra, fuerte) =

$P(\text{El} | *, *) * P(\text{perro} | *, \text{El}) * P(\text{ladra} | \text{El}, \text{perro}) * P(\text{fuerte} | \text{perro}, \text{ladra}) * P(\text{STOP} | \text{ladra}, \text{fuerte})$



# EJEMPLO

$$q(w_i | w_{i-2}, w_{i-1}) = \frac{\text{Count}(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})}$$

P(El, perro, ladra, fuerte) =

$P(\text{El} | *, *) * P(\text{perro} | *, \text{El}) * P(\text{ladra} | \text{El}, \text{perro}) * P(\text{fuerte} | \text{perro}, \text{ladra}) * P(\text{STOP} | \text{ladra}, \text{fuerte})$

$$P(\text{El} | *, *) = \frac{\text{count}(*, *, \text{El})}{\text{count}(*, *)}$$



# EJEMPLO

$$q(w_i | w_{i-2}, w_{i-1}) = \frac{\text{Count}(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})}$$

$P(\text{El, perro, ladra, fuerte}) =$

$$P(\text{El} | *, *) * P(\text{perro} | *, \text{El}) * P(\text{ladra} | \text{El, perro}) * P(\text{fuerte} | \text{perro, ladra}) * P(\text{STOP} | \text{ladra, fuerte})$$

$$P(\text{El} | *, *) = \frac{\text{count}(*, *, \text{El})}{\text{count}(*, *)}$$

$$P(\text{perro} | *, \text{El}) = \frac{\text{count}(*, \text{El}, \text{perro})}{\text{count}(*, \text{El})}$$



# EJEMPLO

$$q(w_i | w_{i-2}, w_{i-1}) = \frac{\text{Count}(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})}$$

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} | *, *) * P(\text{perro} | *, \text{El}) * P(\text{ladra} | \text{El, perro}) * P(\text{fuerte} | \text{perro, ladra}) * P(\text{STOP} | \text{ladra, fuerte})$

$$P(\text{El} | *, *) = \frac{\text{count}(*, *, \text{El})}{\text{count}(*, *)}$$

$$P(\text{perro} | *, \text{El}) = \frac{\text{count}(*, \text{El}, \text{perro})}{\text{count}(*, \text{El})}$$

$$P(\text{ladra} | \text{El, perro}) = \frac{\text{count}(\text{El, perro, ladra})}{\text{count}(\text{El, perro})}$$



# EJEMPLO

$$q(w_i | w_{i-2}, w_{i-1}) = \frac{\text{Count}(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})}$$

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} | *, *) * P(\text{perro} | *, \text{El}) * P(\text{ladra} | \text{El, perro}) * P(\text{fuerte} | \text{perro, ladra}) * P(\text{STOP} | \text{ladra, fuerte})$

$$P(\text{El} | *, *) = \frac{\text{count}(*, *, \text{El})}{\text{count}(*, *)}$$

$$P(\text{fuerte} | \text{perro, ladra}) = \frac{\text{count}(\text{perro, ladra, fuerte})}{\text{count}(\text{perro, ladra})}$$

$$P(\text{perro} | *, \text{El}) = \frac{\text{count}(*, \text{El, perro})}{\text{count}(*, \text{El})}$$

$$P(\text{ladra} | \text{El, perro}) = \frac{\text{count}(\text{El, perro, ladra})}{\text{count}(\text{El, perro})}$$





# EJEMPLO

$$q(w_i | w_{i-2}, w_{i-1}) = \frac{\text{Count}(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})}$$

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} | *, *) * P(\text{perro} | *, \text{El}) * P(\text{ladra} | \text{El, perro}) * P(\text{fuerte} | \text{perro, ladra}) * P(\text{STOP} | \text{ladra, fuerte})$

$$P(\text{El} | *, *) = \frac{\text{count}(*, *, \text{El})}{\text{count}(*, *)}$$

$$P(\text{fuerte} | \text{perro, ladra}) = \frac{\text{count}(\text{perro, ladra, fuerte})}{\text{count}(\text{perro, ladra})}$$

$$P(\text{perro} | *, \text{El}) = \frac{\text{count}(*, \text{El, perro})}{\text{count}(*, \text{El})}$$

$$P(\text{STOP} | \text{ladra, fuerte}) = \frac{\text{count}(\text{ladra, fuerte, STOP})}{\text{count}(\text{ladra, fuerte})}$$

$$P(\text{ladra} | \text{El, perro}) = \frac{\text{count}(\text{El, perro, ladra})}{\text{count}(\text{El, perro})}$$



# EJEMPLO

$$q(w_i | w_{i-2}, w_{i-1}) = \frac{\text{Count}(w_{i-2}, w_{i-1}, w_i)}{\text{Count}(w_{i-2}, w_{i-1})}$$

$P(\text{El, perro, ladra, fuerte}) =$

$P(\text{El} | *, *) * P(\text{perro} | *, \text{El}) * P(\text{ladra} | \text{El, perro}) * P(\text{fuerte} | \text{perro, ladra}) * P(\text{STOP} | \text{ladra, fuerte})$

$$\frac{3}{3} * \frac{3}{3} * \frac{2}{3} * \frac{1}{2} * \frac{1}{1} = \frac{1}{3}$$



# TAREA N° 1

**❑ Explicar gráficamente cual seria el principal problema del método básico que acabamos de ver.**

Este modelo NO va a generalizar a oraciones que NO vimos en el CORPUS de entrenamiento. Cualquier oración que NO esta en el CORPUS de entrenamiento tiene probabilidad 0%.



# TAREA N° 2

1. Dado el siguiente CORPUS de entrenamiento.

“Mi amigo Miguel Diaz es feliz”

“Mi amigo Miguel Diaz es triste”

“Miguel Diaz es mi mejor amigo”

“ Tengo varios amigos”

2. Calcular la probabilidad de la siguiente oración:

“Miguel Diaz es mi mejor amigo”



CIERRE



¿Qué es la  
probabilidad?

¿Qué es un  
proceso de  
Markov?

¿Cuál es el rol de  
una base de  
datos de  
entrenamiento?



# CONSULTAS

[pcsirife@upc.edu.pe](mailto:pcsirife@upc.edu.pe)