

# Introducción a la Clasificación

## KNN-Naive Bayes

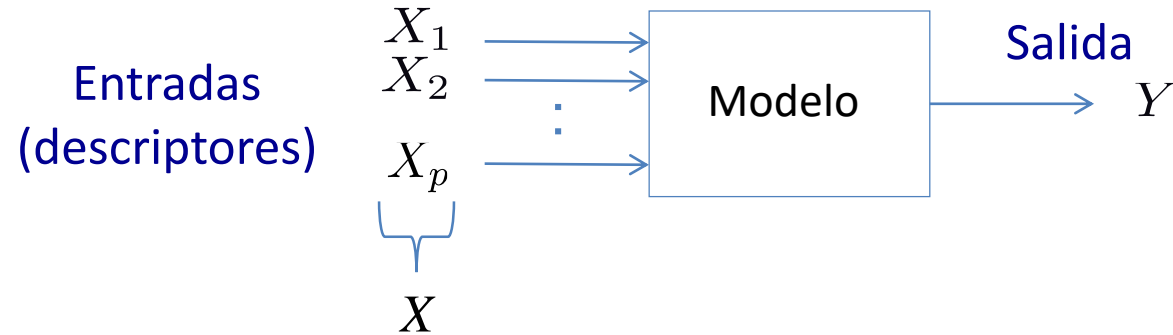
Aprendizaje Bioestadístico

**Santiago Alférez**

Departament de Matemàtiques (DMAT)  
Escola d'Enginyeria de Barcelona Est (EEBE)  
Universitat Politècnica de Catalunya (UPC)

Enero 8 de 2024

# Clasificación



- En regresión  $Y$  es cuantitativa
- **En clasificación  $Y$  es cualitativa (categórica)**

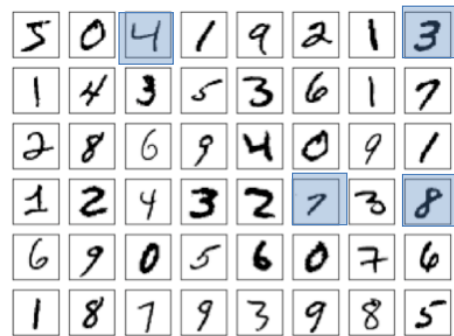
Muchas situaciones:

- Condición de paciente  $\in \{\text{sano}, \text{enfermo}\}$
- Célula  $\in \{\text{normal}, \text{drepanocito}, \text{esferocito}\}$
- .....

Dado un objeto caracterizado por un vector de descriptores, el **objetivo de un clasificador (modelo)** es predecir a qué clase pertenece el objeto dentro de un conjunto de clases predefinidas.

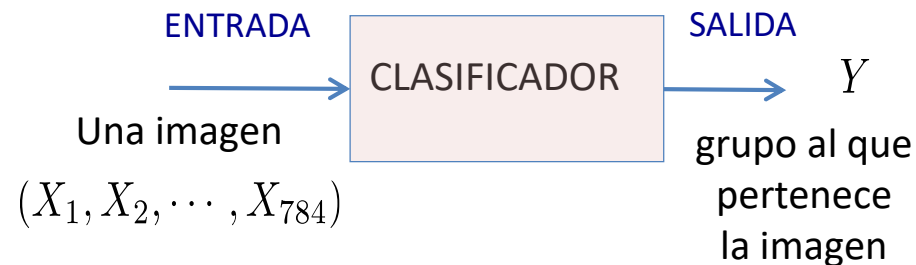
# Clasificación - Ejemplo

Base de datos MNIST: 70000 imágenes de los **dígitos (0,1,..., 9)** escritos a mano



Cada imagen contiene un único dígito en escala de grises y está **etiquetada**, es decir se conoce cuál es el valor del dígito.

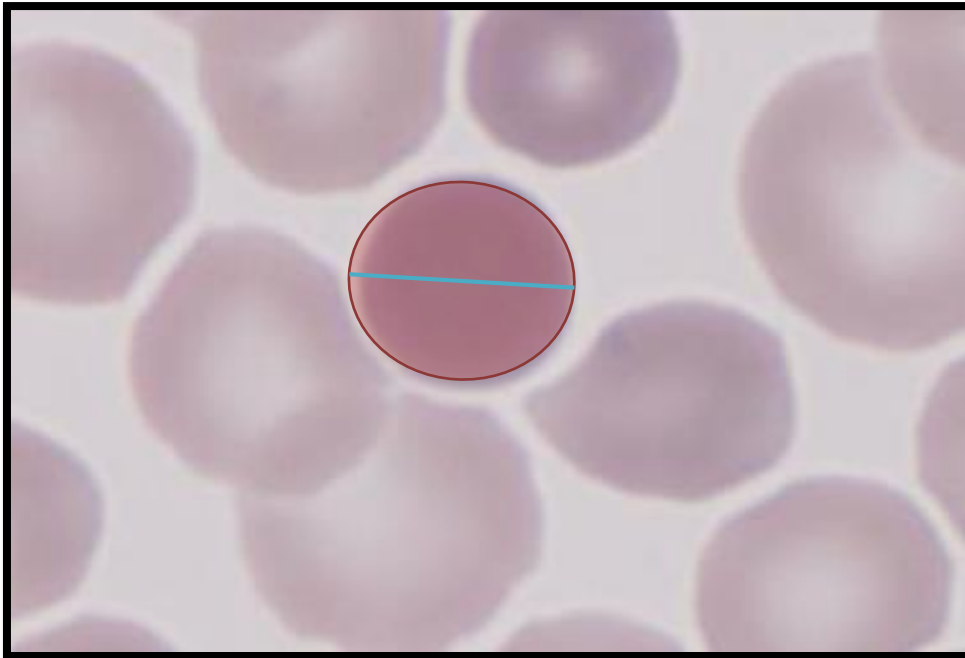
La imagen tiene 28 x 28 píxeles y está representada por **784 descriptores numéricos**, que son los valores de la intensidad de cada uno de los píxeles en un rango entre 0 y 255.



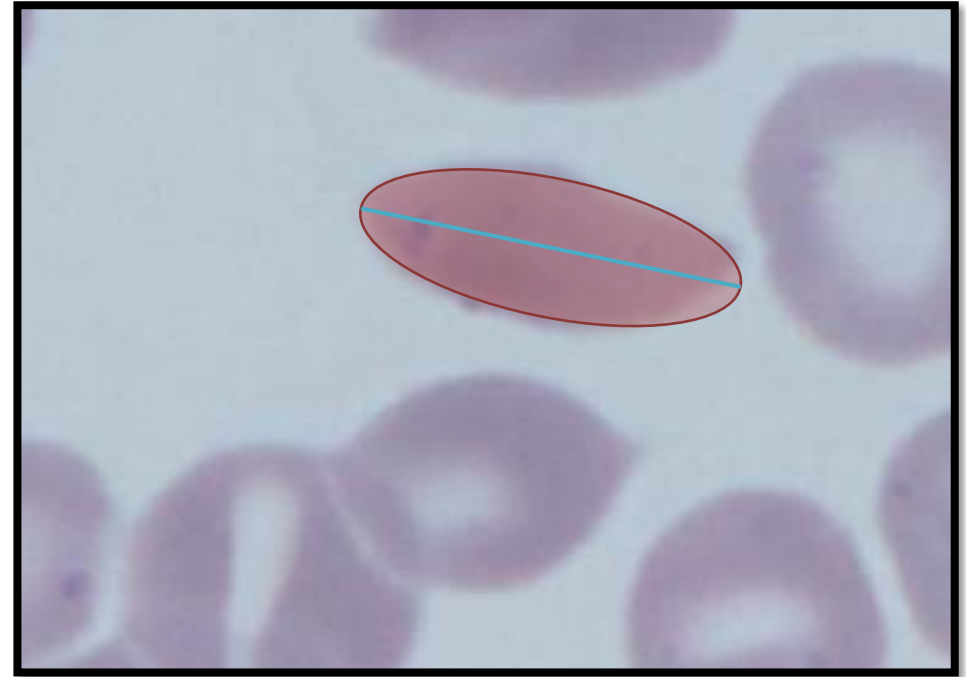
**(Cuál es el dígito ?)**

# Ejemplo: Anemias Hemolíticas

**Esferocitos**



**Drepanocitos**

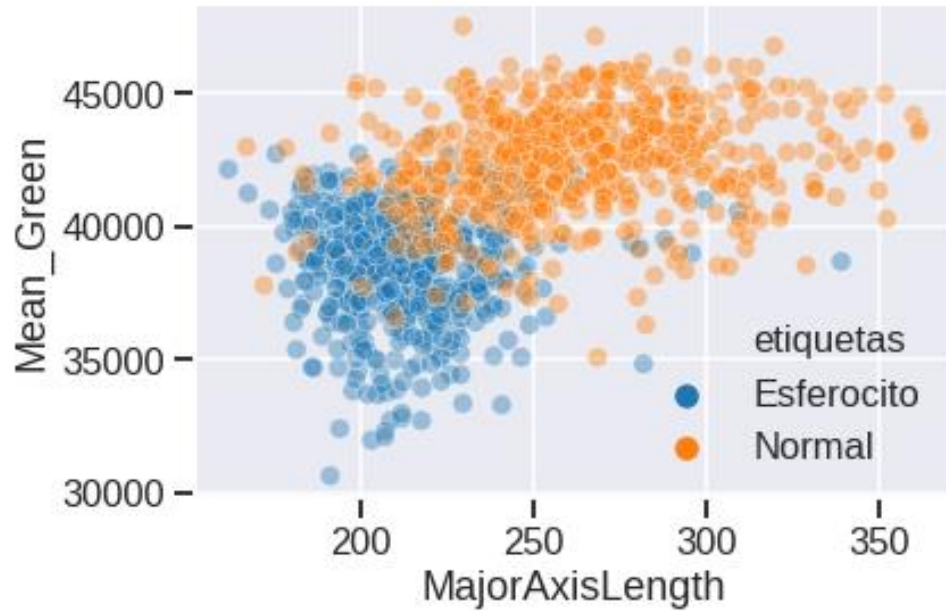


# Ejemplo: Tipo de Glóbulos Rojos

	etiquetas	MajorAxisLength	Area_cualitativa	Mean_Green
1565	Normal	305.976345	High	41625.075155
413	Drepanocits	410.030612	High	36756.232305
1428	Normal	285.253194	High	42508.682130
1095	Esferocito	205.203827	Low	38928.427773
856	Esferocito	193.184427	Low	41742.858976
9	Drepanocits	432.632437	High	35829.883802

```
1 df['etiquetas'].value_counts()
Esferocito      552
Normal          552
Drepanocits     552
Name: etiquetas, dtype: int64
```

# Ejemplo: Tipo de RBCs

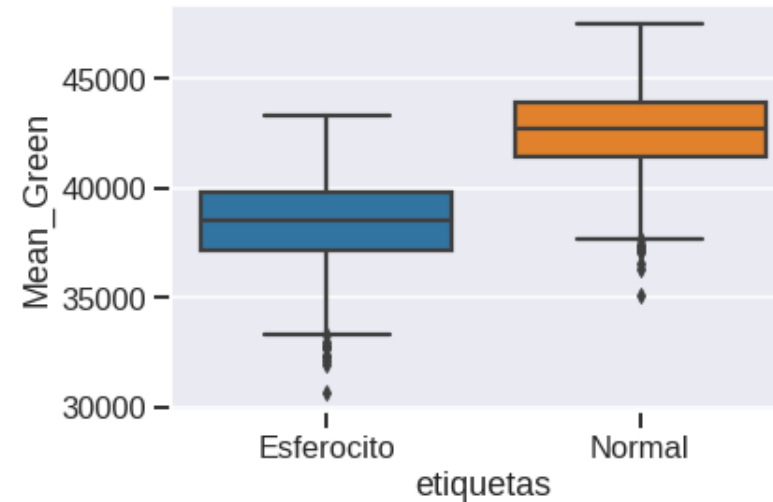
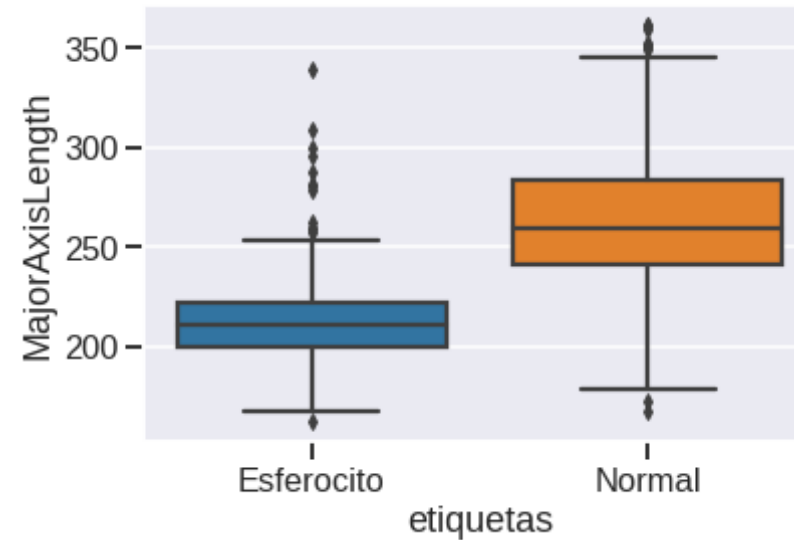


`seaborn.scatterplot`

**X1 = Mean\_Green.**

**X2 = MajorAxisLength**

**Y= RBC class {Normal, Spherocyte}**



`seaborn.boxplot`

# Cómo abordar la clasificación

Una forma natural para clasificar un objeto es mediante las **probabilidades** de que, dado un valor de los descriptores ( $X = x$ ), la salida pertenezca a una u otra de las categorías ( $Y = k$ ):

$$\Pr(Y = k | X = x)$$

Esto requiere estimar o modelizar la probabilidad condicional de la respuesta  $Y$  de cada clase dados los descriptores.

¿ Cómo se hace ?

# Clasificador ideal de Bayes

Supongamos que existen  $K$  clases  $\{label_1, label_2, \dots, label_K\}$ , y que se **codificadas** como  $\{1, 2, \dots, K\}$ . Sean

$$p_k(x) = \Pr(Y = k|X = x), \quad k = 1, 2, \dots, K.$$

las **probabilidades condicionadas** en  $x$  (particularmente, la probabilidad de que  $Y = k$ , dado el vector predictor  $x$ ). Entonces, **el clasificador óptimo Bayesiano** en  $x$ , es definido por:

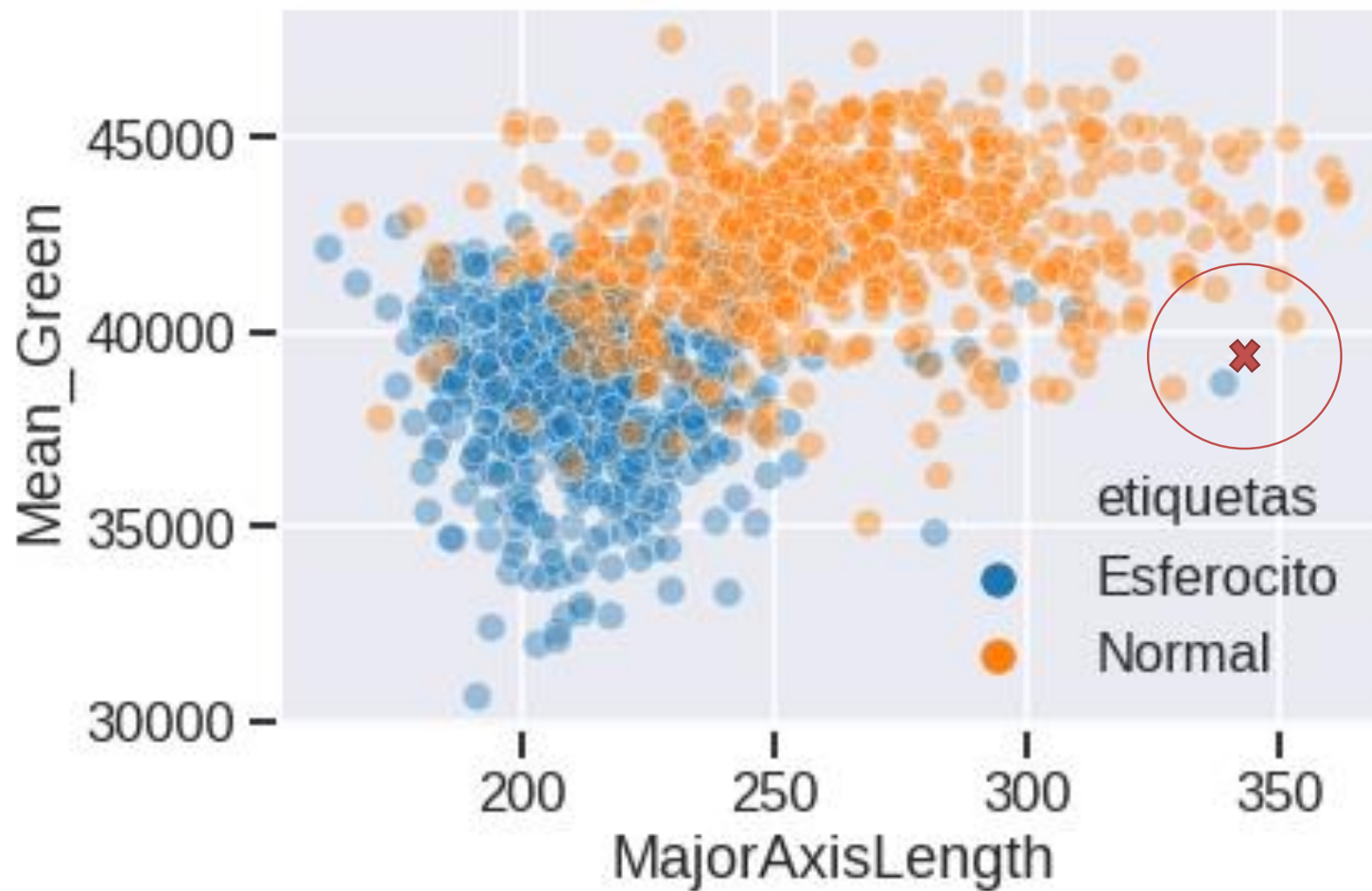
$$f(X = x) = j \text{ si } p_j(x) = \max\{p_1(x), p_2(x), \dots, p_K(x)\}$$

**Se elige la clase que tiene mayor probabilidad para el valor del descriptor calculado.**

**Puede demostrarse que este clasificador minimiza el error medio en los aciertos.**



# K-Nearest Neighbors (KNN)



# K-Nearest Neighbors (KNN)

- El clasificador de Bayes es ideal (imposible de calcular) y sirve como un estándar de **oro** para comparar otros métodos.
- Un método para **estimar** la probabilidad condicional de  $Y$  dado  $X$ , es KNN.
- Dados un posible entero  $K$ , y una observación de *prueba*  $x_0$ :

$$\Pr(Y = j | X = x_0) = \frac{1}{K} \sum_{i \in \mathcal{N}_0} I(y_i = j)$$

Estima la probabilidad condicional para la clase  $j$  como la fracción de puntos en la vecindad  $\mathcal{N}_0$  que pertenecen a la clase  $j$ .

# Probabilidad Condicional

La probabilidad de que  $A$  ocurra cuando el evento  $B$  ya ha ocurrido es dada por:

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

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

**Regla de la multiplicación**

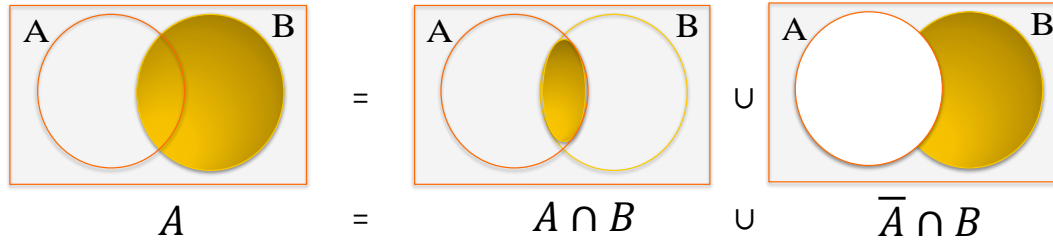
$$P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A)$$

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

**Teorema de Bayes**



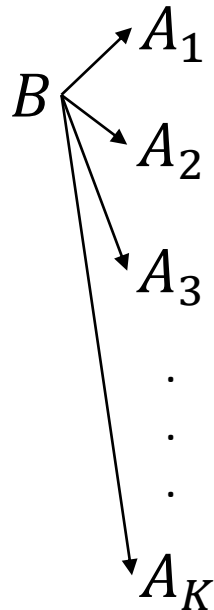
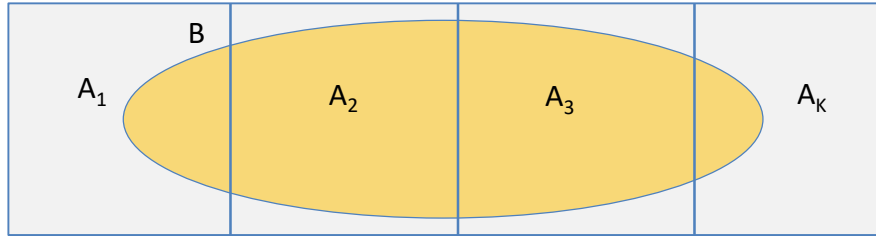
# Probabilidad Total



Probabilidad total (2 eventos)

$A \cap B$  y  $\bar{A} \cap B$   
 $A$  y  $\bar{A}$  son mutuamente exclusivos

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$$



Donde  $P(A_1) + P(A_2) + P(A_3) + \dots + P(A_K) = 1$

$A_1, A_2, A_3, \dots, A_K$  son eventos mutuamente exclusivos

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_K)P(A_K)$$

$$P(B) = \sum_{j=1}^K P(B|A_j)P(A_j)$$

# ¿Cómo podemos aplicar el Teorema de Bayes para Clasificar?

$$P(B) = \sum_{j=1}^K P(B|A_j)P(A_j)$$

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

$$P(Y = k|X = x) = \frac{P(X = x|Y = k)P(Y = k)}{P(X = x)}$$

$$\hat{y} = \operatorname{argmax}_k P(X = x|Y = k)$$

# Ejemplo Práctico

Dado un historial clínico, deseamos clasificar si una persona está *Sana* o *Enferma* dada su congestión nasal:

- *Nada*
- *Poco*
- *Mucha*

Congestión nasal	Clasificación
Mucha	Enferma
Mucha	Enferma
Nada	Sana
Poco	Sana
Poco	Sana
Poco	Enferma
Nada	Sana
Mucha	Enferma
Mucha	Sana
Poco	Sana
Mucha	Sana
Nada	Sana
Nada	Sana
Poca	Enferma

# Ejemplo Práctico

Llega un nuevo paciente que tiene un *Poco* de congestión nasal. ¿Podemos clasificar el diagnóstico de la persona como *Sana* o *Enferma*?

El objetivo es calcular:

1. La probabilidad de que esté *Sana* si el paciente tiene un *Poco* de congestión nasal:  $P(\textit{Sana}|\textit{Poco})$
2. La probabilidad de estar Enferma si el paciente tiene un *Poco* de congestión nasal:  $P(\textit{Enferma}|\textit{Poco})$
3. Comparar las dos probabilidades y decidir

$$P(\textit{Sana}|\textit{Poco}) = \frac{P(\textit{Poco}|\textit{Sana})P(\textit{Sana})}{P(\textit{Poco})}$$

$$P(\textit{Enferma}|\textit{Poco}) = \frac{P(\textit{Poco}|\textit{Enferma})P(\textit{Enferma})}{P(\textit{Poco})}$$

# Ejemplo Práctico

	Congestión nasal	Clasificación	
	Mucha	Enferma	
	Mucha	Enferma	
	Nada	Sana	S
P P	Poco	Sana	S
P P	Poco	Sana	S
P	Poco	Enferma	
	Nada	Sana	S
	Mucha	Enferma	
	Mucha	Sana	S
P P	Poco	Sana	S
	Mucha	Sana	S
	Nada	Sana	S
	Nada	Sana	S
P	Poco	Enferma	

$N = 14$

$$P(\text{Sana}|\text{Poco}) = \frac{P(\text{Poco}|\text{Sana})P(\text{Sana})}{P(\text{Poco})}$$

$$P(\text{Sana})=9/14$$

$$P(\text{Poco})=5/14$$

$$P(\text{Poco}|\text{Sana})=3/9$$



$$P(\text{Sana}|\text{Poco}) = \frac{3/9 \times 9/14}{5/14} = \frac{3}{5} = 0,6$$

$$P(\text{Enferma}|\text{Poco}) = \frac{P(\text{Poco}|\text{Enferma})P(\text{Enferma})}{P(\text{Poco})}$$

$$P(\text{Enferma})=5/14$$

$$P(\text{Poco}|\text{Enferma})=2/5$$

$$P(\text{Enferma}|\text{Poco}) = \frac{2/5 \times 5/14}{5/14} = \frac{2}{5} = 0,4$$



$$P(\text{Sana}|\text{Poco}) = 0,6$$

$$P(\text{Enferma}|\text{Poco}) = 0,4$$



En el anterior procedimiento, hay algo redundante, ¿qué es?

# Clasificación desde un punto de vista de probabilidad

La clase estimada ( $\hat{y}$ ) es aquella con mayor probabilidad

$$\hat{y} = \underset{k}{\operatorname{argmax}} [P(X = \mathbf{x} | Y = k) P(Y = k)]$$

$$\hat{y} = \operatorname{argmax} [f_k(\mathbf{x}) \pi_k]$$

Suponiendo que todos los descriptores son independientes,

$$f_k(\mathbf{x}) = P(X = \mathbf{x} | Y = k) \text{ y } \pi_k = P(Y = k)$$

se puede obtener por historia, utilizando las frecuencias relativas, asumiendo una distribución normal, etc. (de acuerdo con la naturaleza de los datos)

## Naive-Bayes Classifier



### 1.9. Naive Bayes

1.9.1. Gaussian Naive Bayes

1.9.2. Multinomial Naive Bayes

1.9.3. Complement Naive Bayes

1.9.4. Bernoulli Naive Bayes

1.9.5. Categorical Naive Bayes

1.9.6. Out-of-core naive Bayes

# Suponiendo que todos los descriptores son independientes ...

$$\hat{y} = \operatorname{argmax}_k \left[ \overbrace{P(X = \mathbf{x} | Y = k)}^{f_k(\mathbf{x})} \overbrace{P(Y = k)}^{\pi_k} \right]$$

$$P(X = \mathbf{x} | Y = k) \rightarrow P(\mathbf{x} | Y = k) = P(x_1, x_2, \dots, x_m | Y = k)$$

Asumiendo que los descriptores son independientes:

$$P(x_1, x_2, \dots, x_m | Y = k) = P(x_1 | Y = k) P(x_2 | Y = k) \dots P(x_m | Y = k) = \prod_{i=1}^m P(x_i | Y = k)$$

$$\hat{y} = \operatorname{argmax}_k \left[ \prod_{i=1}^m P(x_i | Y = k) P(Y = k) \right] = \operatorname{argmax}_k \left[ \prod_{i=1}^m f_k(x_i) \pi_k \right]$$

$$f_k(\mathbf{x}) = \prod_{i=1}^m f_k(x_i)$$

# Hemos visto

- Clasificación
- Ejemplos de clasificación
- ¿Cómo abordar la clasificación?
- Clasificador ideal de Bayes
- KNN
- Repaso de probabilidad condicional y teorema de Bayes
- Clasificador Naive Bayes

# Próxima clase

- Práctica sobre clasificación, KNN y NB
- Evaluar el desempeño en clasificación
- Validación cruzada



$$p(C_k | \mathbf{x}) = \frac{p(C_k) p(\mathbf{x} | C_k)}{p(\mathbf{x})}$$

In plain English, using [Bayesian probability](#) terminology, the above equation can be written as

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

In practice, there is interest only in the numerator of that fraction, because the denominator does not depend on  $C$  and the values of the features  $x_i$  are given, so that the denominator is effectively constant. The numerator is equivalent to the [joint probability](#) model

$$p(C_k, x_1, \dots, x_n)$$

which can be rewritten as follows, using the [chain rule](#) for repeated applications of the definition of [conditional probability](#):

$$\begin{aligned} p(C_k, x_1, \dots, x_n) &= p(x_1, \dots, x_n, C_k) \\ &= p(x_1 | x_2, \dots, x_n, C_k) p(x_2, \dots, x_n, C_k) \\ &= p(x_1 | x_2, \dots, x_n, C_k) p(x_2 | x_3, \dots, x_n, C_k) p(x_3, \dots, x_n, C_k) \\ &= \dots \\ &= p(x_1 | x_2, \dots, x_n, C_k) p(x_2 | x_3, \dots, x_n, C_k) \cdots p(x_{n-1} | x_n, C_k) p(x_n | C_k) p(C_k) \end{aligned}$$

Now the "naive" [conditional independence](#) assumptions come into play: assume that all features in  $\mathbf{x}$  are [mutually independent](#), conditional on the category  $C_k$ . Under this assumption,

$$p(x_i | x_{i+1}, \dots, x_n, C_k) = p(x_i | C_k).$$

Thus, the joint model can be expressed as

$$\begin{aligned} p(C_k | x_1, \dots, x_n) &\propto p(C_k, x_1, \dots, x_n) \\ &= p(C_k) p(x_1 | C_k) p(x_2 | C_k) p(x_3 | C_k) \cdots \\ &= p(C_k) \prod_{i=1}^n p(x_i | C_k), \end{aligned}$$