

# Machine Learning

## Teoría Bayesiana de la Decisión

Christian Oliva Moya  
Pedro Ramón Ventura Gómez

# Repaso

- Aprendizaje Automático

- Supervisado

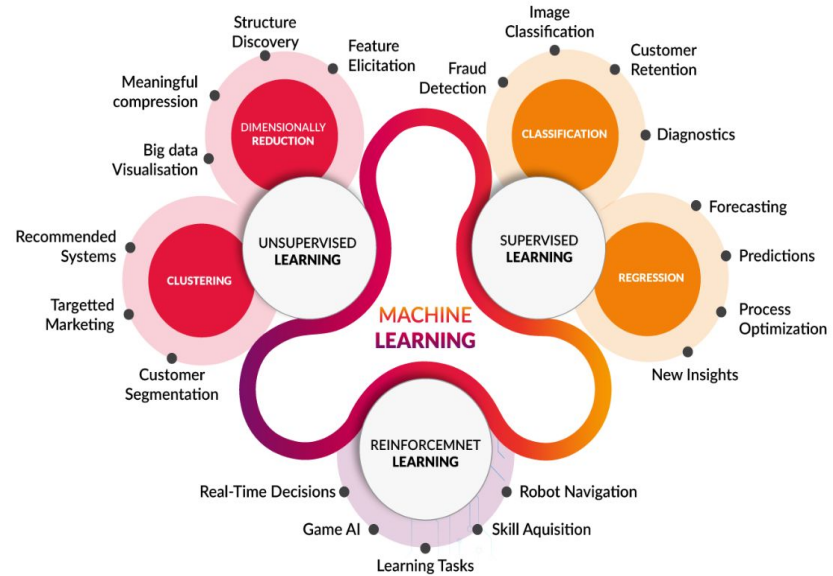
- Clasificación

- Regresión

- No supervisado

- Clustering

- Otros paradigmas



# Repaso

- Aprendizaje Supervisado

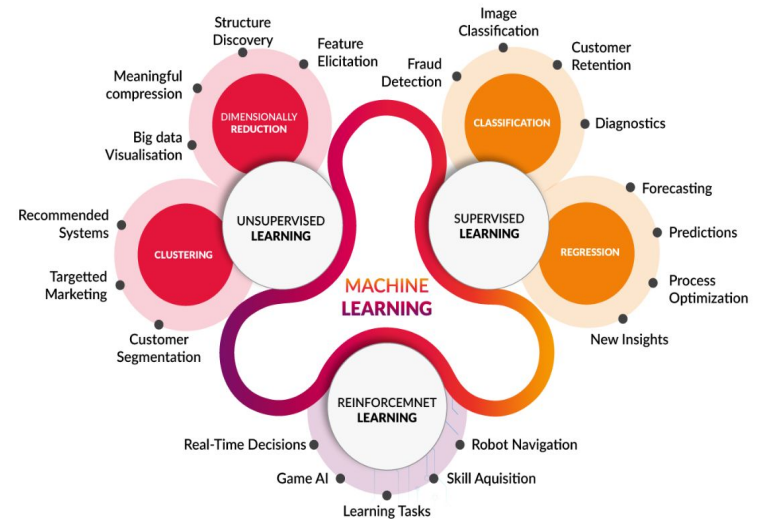
- Problema:  $\{(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), \dots, (\mathbf{x}_N, t_N)\}$

- $\mathbf{x}_i \equiv$  vector de atributos

- $t_i \equiv$  variable objetivo (target)

- $N \equiv$  número de datos/patrones

- Objetivo: predecir  $t_i$  a partir de  $\mathbf{x}_i \rightarrow y_i = f(\mathbf{x}_i) \approx t_i$



# Repaso

---

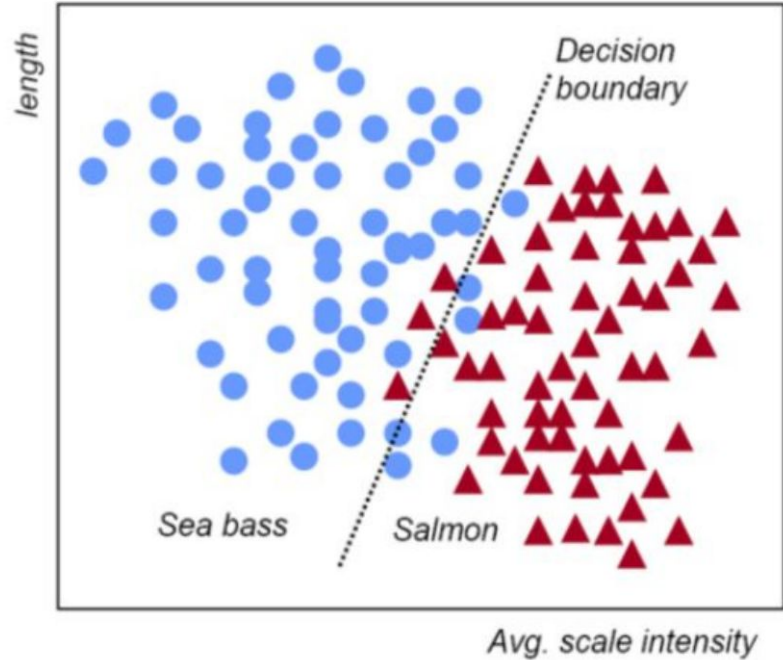
- Ejemplo: salmones vs lubinas



- Atributos:
  - Longitud
  - Brillo
  - Posición de las aletas
  - Color
  - ...

# Repaso

- Ejemplo: salmones vs lubinas

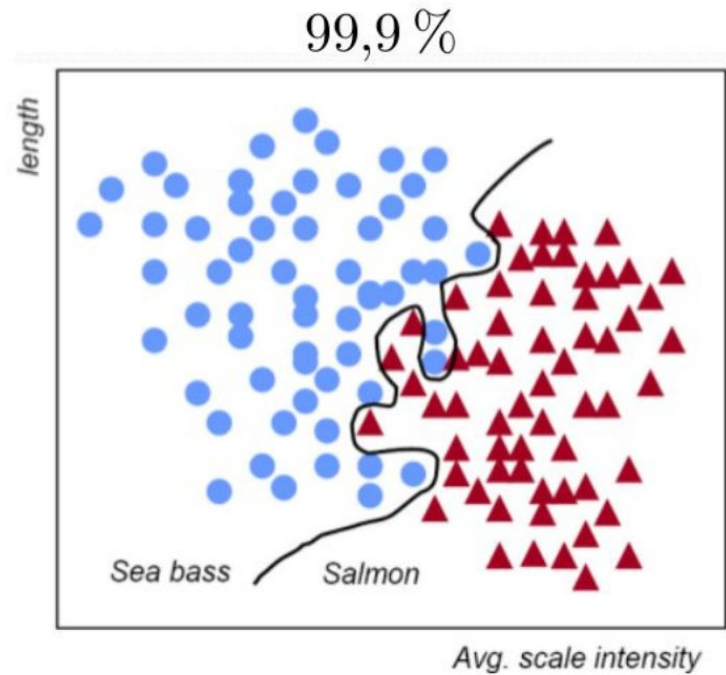
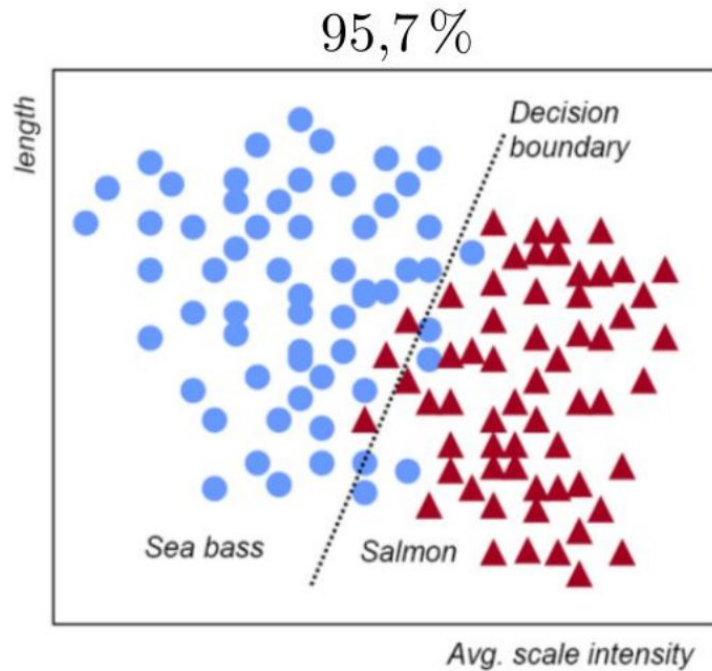


(From Duda, Hart and Stork, *Pattern Classification*, 2001)

*Accuracy: 95%*

# Repaso

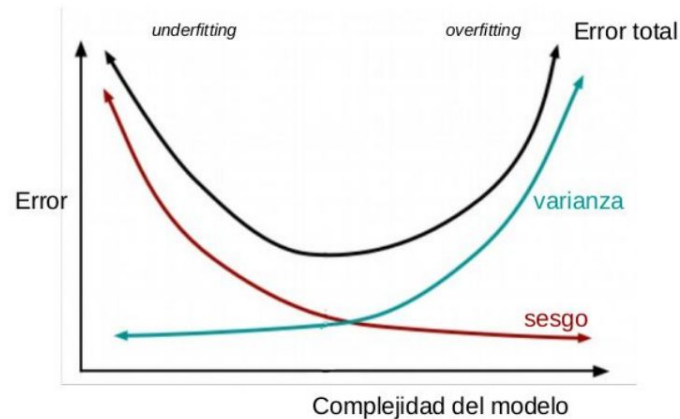
*¿Cuál es el mejor modelo?*



(From Duda, Hart and Stork, *Pattern Classification*, 2001)

# Repaso

- Complejidad, generalización y sobreajuste
  - Modelos más complejos se adaptan mejor a los datos
  - Modelos más complejos generalizan peor (sobreajuste / overfitting)



# Teoría Bayesiana de la Decisión

---

- Planteamiento estadístico del problema, asumiendo que:
  - El problema se puede plantear en términos de probabilidades y costes
  - Todas las probabilidades relevantes son conocidas
- Problema de ejemplo:

	<b>L &lt; 50 cm</b>	<b>L &gt; 50 cm</b>	<b>Total</b>
<b>Salmón</b>	20	40	60
<b>Lubina</b>	30	10	40



# Teoría Bayesiana de la Decisión

- Probabilidad a priori:

- Probabilidad de observar un salmón/lubina en la cinta

- $P(\text{salmón}) = 60 / 100 = 60\% = 0.6$

- $P(\text{lubina}) = 40 / 100 = 40\% = 0.4$

- $P(\text{salmón}) + P(\text{lubina}) = 0.6 + 0.4 = 1.0$

- En general:  $P(C1) + P(C2) + \dots + P(Cn) = 1.0$

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40

# Teoría Bayesiana de la Decisión

- Regla de decisión:

- Define qué acción a tomar basándose en la observación

- Supongamos que:

- Solo conocemos el prior, es decir, no hay datos de entrada

- Lo mejor que podemos hacer es:

$$\text{decisión} = \begin{cases} \text{lubina si } p(\text{lubina}) > p(\text{salmón}) \\ \text{salmón si } p(\text{lubina}) < p(\text{salmón}) \end{cases}$$

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40

# Teoría Bayesiana de la Decisión

- Regla de decisión en base al prior:
  - Elegimos siempre la clase más probable
  - Es la regla óptima en ausencia de más información
  - Siempre se asigna la misma clase a todos los patrones

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40

*En el ejemplo,  $P(\text{salmón}) > P(\text{lubina}) \rightarrow$  todos los datos son salmones*

# Teoría Bayesiana de la Decisión

- Regla de decisión:

- Define qué acción a tomar basándose en la observación

- Supongamos que:

- Conocemos un conjunto de variables que describen a cada patrón

- Probabilidad a posteriori:

- $P(\text{salmón} \mid L < 50 \text{ cm}) = 20 / 50 = 0.4$

$$P(\text{lubina} \mid L < 50 \text{ cm}) = 30 / 50 = 0.6$$

- $P(\text{salmón} \mid L > 50 \text{ cm}) = 40 / 50 = 0.8$

$$P(\text{lubina} \mid L > 50 \text{ cm}) = 10 / 50 = 0.2$$

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40

# Teoría Bayesiana de la Decisión

- Regla de decisión:

- Define qué acción a tomar basándose en la observación

- Supongamos que:

- Conocemos un conjunto de variables que describen a cada patrón

- Probabilidad a posteriori:

- $P(\text{salmón} \mid L < 50 \text{ cm}) = 20 / 50 = 0.4$

$$P(\text{lubina} \mid L < 50 \text{ cm}) = 30 / 50 = 0.6$$

- $P(\text{salmón} \mid L > 50 \text{ cm}) = 40 / 50 = 0.8$

$$P(\text{lubina} \mid L > 50 \text{ cm}) = 10 / 50 = 0.2$$

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40

# Teoría Bayesiana de la Decisión

- Regla de decisión MAP:

- Elegimos siempre la clase más probable dados los atributos observados
- Estimación MAP (máximo a posteriori)

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40

*En el ejemplo:*

$$\text{Si } L < 50 \text{ cm} \rightarrow P(\text{salmón} \mid L < 50 \text{ cm}) = 20 / 50 = 0.4 \quad P(\text{lubina} \mid L < 50 \text{ cm}) = 30 / 50 = 0.6$$

$$\text{Si } L > 50 \text{ cm} \rightarrow P(\text{salmón} \mid L > 50 \text{ cm}) = 40 / 50 = 0.8 \quad P(\text{lubina} \mid L > 50 \text{ cm}) = 10 / 50 = 0.2$$

# Teoría Bayesiana de la Decisión

---

- Teorema de Bayes:

$$P(C_j|x_i) = \frac{P(x_i|C_j)P(C_j)}{P(x_i)}$$

# Teoría Bayesiana de la Decisión

- Teorema de Bayes:

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40
Total	50	50	100

$$p(\text{salmón} \mid L < 50) = \frac{p(L < 50 \mid \text{salmón}) p(\text{salmón})}{p(L < 50)}$$



# Teoría Bayesiana de la Decisión

- Teorema de Bayes:

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40
Total	50	50	100

$$p(\text{salmón} \mid L < 50) = \frac{p(L < 50 \mid \text{salmón}) p(\text{salmón})}{p(L < 50)}$$

$20/50 = 0.4$

# Teoría Bayesiana de la Decisión

- Teorema de Bayes:

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40
Total	50	50	100

$$p(\text{salmón} | L < 50) = \frac{p(L < 50 | \text{salmón}) p(\text{salmón})}{p(L < 50)}$$

$20/60 = 0.33$

$20/50 = 0.4$

# Teoría Bayesiana de la Decisión

- Teorema de Bayes:

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40
Total	50	50	100

$$p(\text{salmón} \mid L < 50) = \frac{\overset{20/60 = 0.33}{p(L < 50 \mid \text{salmón})} \overset{60/100 = 0.6}{p(\text{salmón})}}{20/50 = 0.4 \quad p(L < 50)}$$

# Teoría Bayesiana de la Decisión

- Teorema de Bayes:

	L < 50 cm	L > 50 cm	Total
Salmón	20	40	60
Lubina	30	10	40
Total	50	50	100

$$p(\text{salmón} \mid L < 50) = \frac{\overset{20/60 = 0.33}{p(L < 50 \mid \text{salmón})} \overset{60/100 = 0.6}{p(\text{salmón})}}{\underset{20/50 = 0.4}{p(L < 50)}} \quad \underset{50/100 = 0.5}{p(L < 50)}$$

# Teoría Bayesiana de la Decisión

- Regla de decisión con Teorema de Bayes:

$$P(C_j|x_i) = \frac{P(x_i|C_j)P(C_j)}{P(x_i)}$$

- La evidencia  $P(x_i)$  no depende de  $C$
- Por tanto:
  - Decisión  $C^* = \operatorname{argmax} P(x | C_i) P(C_i)$
- La decisión depende solo de los prioris y las verosimilitudes.

*¿Cómo podemos estimarlos a partir de una muestra finita?*

# Teoría Bayesiana de la Decisión

---

- Estimación paramétrica:
  - Suponemos una forma funcional conocida  $f(x, w)$
  - Ajustamos los parámetros  $w$  para maximizar la probabilidad de las observaciones
  - Algoritmo EM para clasificación suponiendo distribuciones Gaussianas

# Teoría Bayesiana de la Decisión

- Clasificador Naive Bayes:
  - Combina el teorema de Bayes suponiendo que los atributos son independientes dada la clase

$$P(y \mid x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n \mid y)}{P(x_1, \dots, x_n)}$$



$$P(y \mid x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i \mid y)}{P(x_1, \dots, x_n)}$$

$$\hat{y} = \arg \max_y P(y) \prod_{i=1}^n P(x_i \mid y)$$

# Teoría Bayesiana de la Decisión

## 1.9.1. Gaussian Naive Bayes

**GaussianNB** implements the Gaussian Naive Bayes algorithm for classification. The likelihood of the features is assumed to be Gaussian:

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

The parameters  $\sigma_y$  and  $\mu_y$  are estimated using maximum likelihood.

Presupone  
independencia entre  
los atributos