

Tema 4: Aprendizaje Automático

Abdelmalik Moujahid, Iñaki Inza y Pedro Larrañaga

Departamento de Ciencias de la Computación e Inteligencia Artificial

Universidad del País Vasco

<http://www.sc.ehu.es/isg/>

Aprendizaje automático

- Machine learning
- Computational intelligence
- Data mining
- Supervised pattern recognition
- Unsupervised pattern recognition

Minería de datos

Fases:

1. Recopilación y almacenamiento de la información
2. Filtrado de la información
3. Extracción de datos para llevar a cabo el estudio
4. Modelado
5. Toma de decisiones

Reconocimiento de patrones (supervisado)

Clasificación supervisada

	X_1	$\dots$	X_n	C
$(\mathbf{x}^{(1)}, c^{(1)})$	$x_1^{(1)}$	$\dots$	$x_n^{(1)}$	$c^{(1)}$
$(\mathbf{x}^{(2)}, c^{(2)})$	$x_1^{(2)}$	$\dots$	$x_n^{(2)}$	$c^{(2)}$
$\dots$		$\dots$		$\dots$
$(\mathbf{x}^{(N)}, c^{(N)})$	$x_1^{(N)}$	$\dots$	$x_n^{(N)}$	$c^{(N)}$
$\mathbf{x}^{(N+1)}$	$x_1^{(N+1)}$	$\dots$	$x_n^{(N+1)}$	???

Reconocimiento de patrones (no supervisado)

Clasificación no supervisada

	Y_1	...	Y_i	...	Y_n	C
$\mathbf{x}_1$	y_1^1	...	y_i^1	...	y_n^1	$c^1?$
...	...	...	...	...	...	...
$\mathbf{x}_j$	y_1^j	...	y_i^j	...	y_n^j	$c^j?$
...	...	...	...	...	...	...
$\mathbf{x}_N$	y_1^N	...	y_i^N	...	y_n^N	$c^N?$

Aprendizaje automático

	Estadística	Aprendizaje automático
Premisas	si	no
Score	verosimilitud	porcentaje bien clasif.
Búsqueda	heu. deter. + test hipo.	heu. no deter.
Transparencia	poca	mucha
Validación	no	si
Selección de variables	indirecta	directa e indirecta

Verosimilitud

	X_1	X_2
1	0	1
2	0	1
3	0	1
4	0	1
5	0	0
6	1	0
7	1	0
8	1	0
9	1	0
10	1	1

M_1 : X_1 y X_2 son independientes

$$X_1 \rightarrow B(1; \frac{1}{2})$$

$$X_2 \rightarrow B(1; \frac{1}{2})$$

$$L_{M_1}((x_1^{(1)}, x_2^{(1)}), \dots, (x_1^{(10)}, x_2^{(10)})) \\ = (\frac{1}{2} \cdot \frac{1}{2}) \cdots (\frac{1}{2} \cdot \frac{1}{2}) = (\frac{1}{2})^{20}$$

M_2 : X_1 y X_2 no son independientes

$$X_1 \rightarrow B(1; \frac{1}{2})$$

$$X_2|X_1 = 0 \rightarrow B(1; \frac{4}{5})$$

$$X_2|X_1 = 1 \rightarrow B(1; \frac{1}{5})$$

$$L_{M_2}((x_1^{(1)}, x_2^{(1)}), \dots, (x_1^{(10)}, x_2^{(10)})) \\ = (\frac{1}{2} \cdot \frac{4}{5})^4 (\frac{1}{2} \cdot \frac{1}{5}) \cdot (\frac{1}{2} \cdot \frac{4}{5})^4 \cdot (\frac{1}{2} \cdot \frac{1}{5}) \\ = (\frac{1}{2})^{10} \cdot (\frac{1}{5})^2 \cdot (\frac{4}{5})^8$$