

# Tratamiento de variables categóricas

- La mayoría de algoritmos de aprendizaje automático necesitan que las variables categóricas se conviertan a cuantitativas, para poder aplicar operaciones aritméticas sobre las mismas.
- Tipo de animal  $\in \{\text{mamífero, reptil, ave}\}$
- Posible conversión (1 variable categórica discreta con tres valores):

Tipo	mamífero	reptil	ave
Valores	1	2	3

- **Inconvenientes:**
  - ▶ Se ha asumido que existe un orden entre las categorías (reptil > mamífero, reptil > ave, ...).
  - ▶ Se ha asumido una distancia entre cada una de las categorías.

# Tratamiento de variables categóricas

Tipo	Nº Extrem.	Peso	...
mamífero	4	20	...
mamífero	4	15	...
reptil	0	5	...
ave	2	0.5	...
reptil	4	2	...
...	...	...	...

Tipo	Nº Extrem.	Peso	...
1	4	20	...
1	4	15	...
2	0	5	...
3	2	0.5	...
2	4	2	...
...	...	...	...

# Tratamiento de variables categóricas

Tipo	Nº Extrem.	Peso	...
1 mamífero	4	20	...
2 mamífero	4	15	...
3 mamífero	0	5	...
4 ave	2	0.5	...
5 reptil	2	2	...
...	...	...	...

Tipo	Nº Extrem.	Peso	...
1 mamífero	4	20	...
2 mamífero	4	15	...
3 mamífero	0	5	...
4 ave	2	0.5	...
5 reptil	4	2	...
...	...	...	...

## Tratamiento de variables categóricas

- Binarización de variables categóricas a través de la **representación 1-de- $k$** , donde  $k$  es el número de categorías.
  - ▶ La variable categórica genera  $k$  variables binarias.
  - ▶ Por elemento de la base de datos (o patrón), la  $i$ -ésima variable binaria será igual a 1 si el patrón es de esa categoría y a 0 si no lo es.
  - ▶ Es decir, todas las variables binarias serán 0, salvo aquella que corresponda a la categoría del patrón (en la que habrá un 1).
- Binarización de la variable anterior:

Tipo	mamífero	reptil	ave
1-de- $k$	{1, 0, 0}	{0, 1, 0}	{0, 0, 1}

# Tratamiento de variables categóricas

Tipo	Nº Extrem.	Peso	...
mamífero	4	20	...
mamífero	4	15	...
reptil	0	5	...
ave	2	0.5	...
reptil	4	2	...
...	...	...	...

Tipo= mamífero	Tipo= reptil	Tipo= ave	Nº Extrem.	Peso	...
1	0	0	4	20	...
1	0	0	4	15	...
0	1	0	0	5	...
0	0	1	2	0.5	...
0	1	0	4	2	...
...	...	...	...	...	...