UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Introducción al entorno de laboratorio y razonamiento probabilístico

DSIC
Departamento de Sistemas Informáticos y Computación

Objetivos formativos

- Introducir el entorno de laboratorio

- Aplicar conceptos y técnicas de razonamiento probabilístico

Índice

1. Introducción al entorno de laboratorio: python

- Python es un lenguage interpretado
- Uso interactivo o con ficheros que guardan programas
- Introduciremos python con ejemplos sobre razonamiento probabilístico
- Usaremos la librería numpy de python para computo numérico
- Inicio de sesión python: `python -q`

2. Representación probabilística

El conocimiento probabilístico puede representarse con la distribución de probabilidad conjunta de las variables aleatorias de interés.

Ejemplo del dentista: conocimiento para diagnosticar caries

Variables aleatorias de interés:

$$Dolor: \quad D \in \{0, 1\}$$
$$Caries: \quad C \in \{0, 1\}$$
$$Hueco: \quad H \in \{0, 1\}$$

Representación:

$$P(D = d, C = c, H = h)$$

d	c	h	P
0	0	0	0,576
0	0	1	0,008
0	1	0	0,144
0	1	1	0,072
1	0	0	0,064
1	0	1	0,012
1	1	0	0,016
1	1	1	0,108
Suma:			1,000

La tabla del dentista en python

d	c	h	P
0	0	0	0,576
0	0	1	0,008
0	1	0	0,144
0	1	1	0,072
1	0	0	0,064
1	0	1	0,012
1	1	0	0,016
1	1	1	0,108

Introduce la tabla del dentista en python:

```
import numpy as np
T=np.array([[0,0,0,.576],[0,0,1,.008],[0,1,0,.144],
            [0,1,1,.072],[1,0,0,.064],[1,0,1,.012],
            [1,1,0,.016],[1,1,1,.108]])
```

Elemento en la fila 0, columna 3:

```
T[0,3]
```

```
0.57600
```

Elemento en la fila 0, última columna:

```
T[0,-1]
```

```
0.57600
```

Elementos de la fila 0 (inicial) a 3 de la última columna:

```
T[0:4,-1]
T[:4,-1]
```

```
array([0.576, 0.008, 0.144, 0.072])
array([0.576, 0.008, 0.144, 0.072])
```

Elementos de la fila 4 a 7 (final) de la última columna:

```
T[4:8,-1]
T[4:,-1]
```

```
array([0.064, 0.012, 0.016, 0.108])
array([0.064, 0.012, 0.016, 0.108])
```

Elementos (de todas las filas) de la última columna:

```
T[:,-1]
```

```
array([0.576, 0.008, 0.144, 0.072,
       0.064, 0.012, 0.016, 0.108])
```

$d\,c\,h$	P
0 0 0	0,576
0 0 1	0,008
0 1 0	0,144
0 1 1	0,072
1 0 0	0,064
1 0 1	0,012
1 1 0	0,016
1 1 1	0,108

Elementos en las filas 0, 1, 4 y 5 de la última columna:

```
T[[0,1,4,5],-1]
```

```
array([0.576, 0.008, 0.064, 0.012])
```

Suma de los elementos de la última columna:

```
np.sum(T[:,-1])
```

```
1.0
```

Filas con elementos de la columna 0 nulos:

```
np.where(T[:,0]==0)[0]
```

```
array([0, 1, 2, 3])
```

Filas con elementos no nulos en la columna 1:

```
np.where(T[:,1]!=0)[0]
```

```
array([2, 3, 6, 7])
```

Filas con elementos nulos en las columnas 1 y 2:

```
np.where((T[:,1]==0) & (T[:,2]==0))[0]
```

```
array([0, 4])
```

Elementos en última col. de filas con 0 en las cols. 1 y 2:

```
T[np.where((T[:,1]==0) & (T[:,2]==0))[0],-1]
```

```
array([0.576, 0.064])
```

Suma de elem. en última col. de filas con 1 en las cols. 1 y 2:

```
np.sum(T[np.where((T[:,1]==1) & (T[:,2]==1))[0],-1])
```

```
0.18
```

3. Inferencia probabilística

A partir de la distribución conjunta podemos calcular la probabilidad de cualquier *suceso* (*proposición*) mediante aplicación de:

La regla suma:

$$P(x) = \sum_y P(x, y)$$

La regla producto:

$$P(x, y) = P(x)\, P(y \mid x)$$

En general no es necesario conocer la tabla completa de probabilidades conjuntas para calcular la probabilidad de un suceso dado.

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

$d\,c\,h$	P
0 0 0	0,576
0 0 1	0,008
0 1 0	0,144
0 1 1	0,072
1 0 0	0,064
1 0 1	0,012
1 1 0	0,016
1 1 1	0,108

Probabilidad de caries y hueco (a la vez):

$$P(c=1,h=1)=\sum_{d=0,1} P(d,c=1,h=1)=0{,}180$$

```
Pc1h1=np.sum(T[np.where((T[:,1]==1) & (T[:,2]==1))[0],-1])
print('Pc1h1 = %.5f' % Pc1h1)
```

```
Pc1h1 = 0.18000
```

Probabilidad de hueco:

$$P(h=1)=\sum_{d=0,1}\sum_{c=0,1} P(d,c,h=1)=0{,}200$$

```
Ph1=np.sum(T[np.where(T[:,2]==1)[0],-1])
```

```
Ph1 = 0.20000
```

Probabilidad de caries tras observar (sabiendo que hay) hueco:

$$P(c=1\mid h=1)=\frac{P(c{=}1,h{=}1)}{P(h{=}1)}=\frac{0{,}180}{0{,}200}=0{,}900$$

```
Pc1Dh1=Pc1h1/Ph1
```

```
Pc1Dh1 = 0.90000
```

Probabilidad de dolor sabiendo que hay caries:

$$P(d=1\mid c=1)=\frac{P(d{=}1,c{=}1)}{P(c{=}1)}=\frac{0{,}124}{0{,}340}=0{,}365$$

```
Pd1c1=np.sum(T[np.where((T[:,0]==1)&(T[:,1]==1))[0],-1])
Pc1=np.sum(T[np.where(T[:,1]==1)[0],-1])
Pd1Dc1=Pd1c1/Pc1
```

```
Pd1c1  = 0.12400
Pc1    = 0.34000
Pd1Dc1 = 0.36471
```

4. Ejercicio: aplicación del teorema de Bayes

El ***teorema de Bayes*** permite actualizar nuestro conocimiento sobre una hipótesis y tras observar una nueva evidencia x:

$$P(y \mid x) = \frac{P(x, y)}{P(x)} = P(y)\,\frac{P(x \mid y)}{P(x)}$$

De otra forma: $P(y \mid x)$ es la probabilidad de que se produzca el efecto y tras observar que se ha producido la causa x.

Ejercicio: calcula la probabilidad de caries sabiendo que hay dolor

$$P(c = 1 \mid d = 1) = P(c = 1)\,\frac{P(d = 1 \mid c = 1)}{P(d = 1)}$$