

PRÁCTICA 1

CONCEPTOS BÁSICOS

En esta práctica vamos a trabajar algunos conceptos básicos relativos a la asignatura. Para ello vamos a suponer el siguiente conjunto de datos:

Instancia	Atr1	Atr2	Atr3	Atr4	Atr5	Atr6
1	SI	SI	5	14	21000	275
2	NO	NO	8	20	30000	313
3	NO	SI	24	52	78000	30
4	SI	SI	32	68	102000	342
5	NO	SI	3	10	15000	279
6	SI	NO	24	52	78000	348
7	SI	NO	7	18	27000	285
8	NO	SI	8	20	30000	32
9	NO	NO	12	28	42000	326
10	SI	NO	34	72	108000	301

1. Determinar la dependencia/independencia de los atributos nominales Atr1 y Atr2 de acuerdo al valor χ^2 (el valor de referencia para aceptar la hipótesis de independencia es $\chi^2 < 3.84$ con un grado de confianza del 95%). ¿Qué decisión de pre-procesamiento tomarías tras este análisis?
2. Determinar la dependencia/independencia de los atributos numéricos Atr3 y Atr4 de acuerdo al coeficiente de correlación. ¿Qué decisión de pre-procesamiento tomarías tras este análisis?
3. Indica la expresión que usarías para llevar a cabo la normalización lineal uniforme al intervalo [0,1] de los atributos atr4 y atr5. ¿Por qué es importante la normalización de los atributos si vamos a trabajar con distancias?
4. Discretizar el atributo Atr6 en 3 intervalos de acuerdo al método Binning:
 - a. De igual profundidad
 - b. De igual amplitud