

0.1 Gestion d'attributs continus

Soit l'ensemble de données suivant :

Température	Faiblesse	Toux	Malade?
10	0	oui	Angines
30	10	non	Non_malade
20	10	oui	Angines
30	10	non	Non_malade
40	0	non	Fievre
20	10	oui	Angines

L'algorithme C4.5 gère les attributs continus de la manière suivante :

— Étape 01 : Trier les valeurs de l'attribut

Ce qui donnera le résultat suivant :

10	20	30	40
----	----	----	----

— Étape 02 : Créer un seuil à partir de chaque valeur consécutives, ce qui donnera le résultat suivant :

$$\frac{10+20}{2} = 15, \frac{20+30}{2} = 25, \frac{30+40}{2} = 35$$

— Étape 03 : Calculer la quantité d'information apportée par chacune des valeurs précédentes, en transformant les valeurs initiales de l'attribut en valeurs ($\geq VAL$) ou ($< VAL$)

1. Valeur : 15

Température	Faiblesse	Toux	Malade?
< 15	0	oui	Angines
≥ 15	10	non	Non_malade
≥ 15	10	oui	Angines
≥ 15	10	non	Non_malade
≥ 15	0	non	Fievre
≥ 15	10	oui	Angines

$$\text{Entropie(Température)} = \frac{1}{6}(-\frac{1}{1}\text{Log}_2(\frac{1}{1})) + \frac{5}{6}(-(\frac{1}{5}\text{Log}_2(\frac{1}{5}) + \frac{2}{5}\text{Log}_2(\frac{2}{5}) + \frac{2}{5}\text{Log}_2(\frac{2}{5})))$$

$$\text{Entropie(Température)} = 1.27$$

2. Valeur : 25

Température	Faiblesse	Toux	Malade?
< 25	0	oui	Angines
≥ 25	10	non	Non_malade
< 25	10	oui	Angines
≥ 25	10	non	Non_malade
≥ 25	0	non	Fievre
< 25	10	oui	Angines

$$\text{Entropie(Température)} = \frac{3}{6}(-\frac{3}{3}\text{Log}_2(\frac{3}{3})) + \frac{3}{6}(-(\frac{1}{3}\text{Log}_2(\frac{1}{3}) + \frac{2}{3}\text{Log}_2(\frac{2}{3})))$$

$$\text{Entropie(Température)} = 0.48$$

3. Valeur : 35

Température	Faiblesse	Toux	Malade?
< 35	0	oui	Angines
< 35	10	non	Non_malade
< 35	10	oui	Angines
< 35	10	non	Non_malade
≥ 30	0	non	Fievre
< 35	10	oui	Angines

$$\text{Entropie(Température)} = \frac{1}{6}(-\frac{1}{1}\text{Log}_2(\frac{1}{1})) + \frac{5}{6}(-(\frac{2}{5}\text{Log}_2(\frac{2}{5}) + \frac{3}{5}\text{Log}_2(\frac{3}{5})))$$

$$\text{Entropie(Température)} = 0.81$$

Le seuil choisi sera donc ≥ 25 , < 25 , l'ensemble d'apprentissage qui sera utilisé sera donc le suivant :

Température	Faiblesse	Toux	Malade?
< 25	0	oui	Angines
≥ 25	10	non	Non_malade
< 25	10	oui	Angines
≥ 25	10	non	Non_malade
≥ 25	0	non	Fievre
< 25	10	oui	Angines