

Data Mining : Prétraitement des données

TD N : 1

Exercice 1

Supposons que les données à analyser incluent l'attribut Age. Les valeurs d'âge des tuples de données sont (par ordre croissant):

13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

Utilisez le lissage par la moyenne pour lisser les données ci-dessus, en utilisant une profondeur de partition de 3. Illustrez vos étapes.

Exercice 2

Utiliser les méthodes de normalisation pour normaliser l'ensemble de données suivant:

200, 300, 400, 600, 1000

- a) min-max en réglant $\min = 0$ et $\max = 1$
- b) z-score
- c) mise à l'échelle décimale

Exercice 3

En utilisant les données de l'attribut Age données de l'exercice 1, répondez aux questions suivantes

- a) Utilisez la normalisation min-max pour transformer la valeur 35 pour l'âge dans la plage $[0, 1]$.
- b) Utilisez la normalisation z-score pour transformer la valeur 35 pour l'âge, où l'écart-type de l'âge est de 12,94 ans.
- c) Utilisez la normalisation par mise à l'échelle décimale pour transformer la valeur 35 pour l'âge.

Exercice 4

Supposons qu'un hôpital ait testé les données d'âge et de graisse corporelle de 18 adultes sélectionnés au hasard avec le résultat suivant

<i>age</i>	23	23	27	27	39	41	47	49	50
<i>%fat</i>	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
<i>age</i>	52	54	54	56	57	58	58	60	61
<i>%fat</i>	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

a) Normaliser les deux attributs en fonction de la normalisation z-score.