

BÀI TẬP VỀ TIỀN XỬ LÝ DỮ LIỆU

Bài 1. Tiền xử lý tập dữ liệu sau sử dụng kỹ thuật binning:

[16 17 20 21 22 23 23 25 28 28 33 34 36 37]

Mô tả, tóm tắt dữ liệu trên

Bài 2. Cho mảng một chiều $X = \{-5.0, 23.0, 17.6, 7.23, 1.11\}$, hãy chuẩn hóa mảng sử dụng

- a. Decimal scaling: trong khoảng $[-1, 1]$.
- b. Min-max: trong khoảng $[0, 1]$.
- c. Min-max: trong khoảng $[-1, 1]$.
- d. z-score

Bài 3. Làm mịn dữ liệu sử dụng kỹ thuật làm tròn cho tập sau:

$Y = \{1.17, 2.59, 3.38, 4.23, 2.67, 1.73, 2.53, 3.28, 3.44\}$

Sau đó biểu diễn tập thu được với các độ chính xác:

- a. 0.1
- b. 1.0

Bài 4. Cho tập mẫu với các giá trị bị thiếu

- o $X1 = \{0, 1, 1, 2\}$
- o $X2 = \{2, 1, -, 1\}$
- o $X3 = \{1, -, -, 0\}$
- o $X4 = \{-, 2, 1, -\}$

Nếu miền xác định của tất cả các thuộc tính là $[0, 1, 2]$, hãy xác định các giá trị bị thiếu biết rằng các giá trị đó có thể là một trong số các xác trị của miền xác định?