

Data Processing Homework

CISC 4631 Data Mining

Points: 50 points

Suppose that the data for analysis includes the attribute **age**. The age values for the data tuples are 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40, 45, 46, 52, 70.

1. Use smoothing by bin means to smooth the data, using a bin depth of 3. Illustrate your steps.
2. Use min-max normalization to transform the age values onto the range [0.0, 1.0].
3. Use z-score normalization to transform the age values, where the standard deviation of age is 12.94 years.
4. Use normalization by decimal scaling to transform the values.
5. Comment on which method you would prefer to use for the given data and Why?