

1. text book page 141/3.3(a)

Suppose that the data for analysis include the attribute age. The age values for the data tuples are (in increasing order): 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. Use smoothing by bin means to smooth the above data, using a bin depth of 3.

2. text book page 141/3.5. Answer the following question using data in question 1:

- (a) Use min-max normalization to transform the value 35 for age onto the range [0.0-1.0]
- (b) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years.
- (c) Use normalization by decimal scaling to transform the value 35 for age.
- (d) Which method would you prefer to use? Why?

3. Assuming we are trying to identify the type of customer who are likely to buy a new product based on customers' past purchasing history. One of the attributes in the data is the number of purchases a customer has made for the same brand product. Apply the Entropy-based discretization method to discretize this attribute using data given below. (Show all calculation steps involved in the **first level** discretization only)

customer	number of purchase	class
1	3	not buy
2	5	buy
3	3	not buy
4	3	not buy
5	4	buy
6	6	buy
7	4	not buy
8	5	buy
9	6	buy
10	4	not buy

4. Apply *Natural partitioning* to discretize the attribute “profit” in a data. The histogram of the attribute is shown below. Start the first level partitioning by focusing on the data between 5%-tile and 95%-tile range. Show the complete partitioning (discretizing) of this data.

