Q1.
Partition the given data into 4 bins using the equi-depth Binning method and perform smoothing according to the following methods:  

a. smoothing by bin means,

b. smoothing by bin median,

c. smoothing by bin boundaries.

11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 21, 21, 22, 23, 24, 30, 40, 45, 45, 45, 71, 72, 73, 75

Q2.
Generate concept hierarchy for the given data up to 3 levels using 3 - 4 - 5 rule. You should clearly explain how the partitions are obtained at each level. [If Interval consists of 2 or fewer elements it need not be divided further] 

104, 110, 120, 121, 122, 123, 125, 125, 125, 127, 130, 133, 133, 135, 135, 135, 135, 136, 137, 138, 138, 139, 140, 143, 146, 146, 148, 148, 151, 153, 155, 156, 158, 158, 161, 163, 165, 168, 170, 178, 183, 188, 198
Q3.
Reduce the given data using histogram with the following algorithms: [12 points]

a. equi-width histogram,

b. equi-depth histogram, but use V-Optimal criterion to choose between the two options,

c. Max-Diff histogram.

You should partition the data into 4 buckets for each method.

Data 3, 5, 6, 6, 9, 10, 10, 11, 12, 12, 15, 15, 15, 15, 20, 21, 23, 24, 24, 25, 25, 25, 30, 35, 36, 37, 38, 40

Q4.
A Sample data table Containing Variables of Mixed type

<table>
<thead>
<tr>
<th>Object Identifier</th>
<th>Categorical</th>
<th>Ordinal</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N</td>
<td>Excellent</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>R</td>
<td>Ordinary</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>P</td>
<td>Good</td>
<td>58</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>Ordinary</td>
<td>68</td>
</tr>
<tr>
<td>5</td>
<td>P</td>
<td>Good</td>
<td>588</td>
</tr>
</tbody>
</table>

For the given table

i) Normalize each attribute, [8 points]

ii) Compute dissimilarity between tuples according to individual attributes, and [12 points]

iii) Compute distance between tuples by combining results of ii). [12 points]

Note: The reference material is section 2.4 in the 3rd edition and section 7.2 in the 2nd edition. The same material is in section 8.2 in the 1st edition of the book.
Q5.
Given Sample survey data which tells us regarding what kind of food do people eat in a restaurant. [12 points]
Steak – 55.2%
Shrimp – 48.0%
Pizza – 50.0%
Fish – 43.1%
Bread – 42.7%
Salad – 53.6%
Chicken – 33.3%
Using Naïve Prediction
i). What is the prediction for each type of food?
ii). What is the percentage error associated with each of these predictions.

Q6.
Classify following attributes as binary, discrete or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (Interval or ratio); your answers should be provided with explanations of why you reached that conclusion. [10 points]

a. Hair color
b. Phone number
c. Volume in gallons
d. Military rank
e. Temperature in degree