Homework 1. Due: 1/27/16.
Note: This h/w contains two parts.

Part A is for grading and you have to submit your best answer. If the question is ambiguous for you, then make suitable assumptions (and justify your assumptions).

Part B is for practice; you don't need to submit solution. We will provide solution and you can self-grade your answers.

Part A (50 points). All answers should be your own.

Q1. (10 points)

(a) Define data mining (2 points)
(b) Distinguish data mining from statistical analysis (3 points)
(c) List 5 unique data characteristics that makes traditional techniques less effective (5 points)

Q2. (10 points)

(a) Describe any 5 unique datasets (5 points) (e.g., text and image are two unique datasets; but new papers articles and scientific papers are not two unique datasets for this h/w).
(b) For each of this dataset, identify one best data mining technique, and explain why it is best? (5 points)

Q3. (10 points)

(a) Define various types of attributes and list properties that are applicable to each type (4 points)
(b) Give an example of transformation on each attribute type (4 points)
(c) Define asymmetric attribute and give an example (2 points)

Q4. (10 points)

(a) Data = {55, 23, 28, 32, 18, 68, 72, 89, 98, 100}. Transform this data to new range of [0..255] using the min-max normalization, show new values (5 points).
(b) Apply z-score normalization for the same data. Show output. (5 points)

Q5. (10 points)

(a) For each of the 5 unique datasets identified in Q2 (a), identify most suitable similarity measure (2.5 points)
(b) For the following 2-d data, (1) draw scatter plot (2.5 points), (2) compute (pair-wise) distance matrix (5 points)

Object	Height (ft)	Weight (lb)
1	5.6	150
2	5.8	155
3	5.4	145
4	6.1	180
5	5.9	175

Part B (Don't submit solution; for your practice; from end of book chapter exercise). Collaboration is encouraged to solve Part B.

(From Chapter 1; 1.7 Exercises)

1. Q1

(From Chapter 2; 2.6 Exercises)

2. Q2, Q8, Q13, Q18, Q19, Q25, Q28.