CSC 425/525 Homework #7 (Chapter 10)
Due Date: Wednesday, March 29

Word process all answers. Figures may be hand drawn. Undergraduates answer question 1 and any four others. Graduate students answer all 6 questions.

1. For each of the following types of learning by a human, classify it using the categories listed on slide 3 of the chapter 10 power point notes (e.g.,chunking, new case, altering parameters, etc). Explain your choice in comparison to other possibilities.
    a. Learning how to program in C having learned Java.
    b. Learning how to debug your Java programs better having been through CSC 260 and CSC 360.
    c. Learning the words of a new song.
    d. Learning how to drive (assuming you've never driven before).
    e. Learning recursion after having been programming for a couple of semesters.

2. Given the following positive and negative instances of C for loops (not Java for loops), provide a Winston-like semantic network that captures what a C for loop is. Assume {…} is 2 or more instructions and …; is a single instruction.
    + for(i=0;i<n;i++) …;
    - for(i=0;i<n) {…}
    + for(i=0,j=n;i<n;i++) {…}
    + for(; i<n;) {…}
    - for(int i=0;i<n;i++) {…}
    - for i=0; i<n; i++ {…}
    + for(i=0; i<n; i++, n--) {…}
    - for(i=0;i<n;i++) …; …; …;
    + for(i=n;i>0;i--) …;
    - for(; i>0;) …

3. From what you've seen of learning a version space through candidate elimination, provide a critique of this form of learning. Specifically, address the following.
    a. Is it a useful form of learning? Why or why not?
    b. What types of domains/problems seem adequate for learning using version spaces?
    c. To what extent, or in what part of its problem solving, did Lex use candidate elimination and was it a useful demonstration to indicate candidate elimination's utility?

4. The following table provides data that indicates the major chosen by various entering freshmen at NKU from 2012-2016. We want to build an expert system to advice new students at NKU and so we want to build a decision tree from this data. Draw this decision tree. There are many possible solutions to this problem based on the order that you tackle the attributes (recall that the ID3 algorithm uses the notion of information gain to determine which attribute to test next). You do not need to provide an optimal tree but try to match the data as much as possible.

Math	Art	Bus.	Prob. solving	Analysis	Prog.	Money	Theory	Comp.	Major
Y	N	Y	Y	Y	Y	Y	Y	Y	CSC
Y	Y	N	N	N	N	Y	N	Y	MIN
N	N	Y	N	N	N	Y	N	Y	BIS
N	N	N	Y	N	Y	Y	N	Y	CSC

Y	Y	N	Y	Y	Y	Y	N	Y	DSC
Y	N	N	Y	N	N	Y	Y	Y	CIT
Y	Y	Y	Y	N	N	Y	N	N	BIS
N	N	N	Y	N	N	N	N	Y	CIT
N	Y	N	Y	Y	Y	Y	N	N	MIN
Y	Y	Y	Y	Y	Y	Y	N	Y	DSC
Y	N	Y	Y	Y	N	N	Y	Y	MAT
N	Y	Y	N	N	N	Y	N	Y	BIS
Y	N	Y	Y	Y	Y	N	N	Y	MAT
N	Y	N	Y	Y	Y	N	N	Y	MIN
N	N	Y	Y	Y	N	Y	N	N	BIS
Y	Y	Y	Y	N	N	Y	N	Y	CSC
Y	Y	N	Y	Y	N	Y	N	Y	DSC
Y	N	N	Y	Y	Y	Y	N	Y	CSC
Y	N	N	Y	Y	Y	N	N	Y	CIT
N	Y	N	N	N	Y	Y	N	Y	MIN

5. In analyzing 10000 store receipts, we came up with the following four rules with the given frequency and utility values.

   If (a, b, c, d, e) → f (frequency = 50%, utility = 50%)
   If (h, i, j, k, l) → m (frequency = 25%, utility = 75%)
   If (n, o, p, q) → r (frequency 80%, utility = 10%)
   If (s, t, u, v) → w, x, y (frequency 25%, utility = 25%)

   If you were going to select a rule to adopt, which one would it be and why? You can select multiple rules but justify each one.

6. We have the following pieces of data to classify and have three proposed classes. These classes' centers are given below.

   Datum 1: (10, 53, 12, 6, 4, 14, 22)
   Datum 2: (25, 34, 4, 18, 18, 17, 12)
   Datum 3: (9, 17, 12, 16, 17, 5, 11)
   Datum 4: (9, 44, 18, 7, 9, 15, 19)
   Datum 5: (33, 28, 9, 11, 20, 17, 9)
   Datum 6: (33, 35, 7, 7, 10, 14, 12)
   Datum 7: (9, 45, 13, 7, 8, 9, 24)
   Datum 8: (20, 25, 12, 18, 10, 10, 10)

   Class 1 center: (28, 28, 6, 15, 17, 16, 9)
   Class 2 center: (12, 47, 16, 10, 9, 10, 19)
   Class 3 center: (15, 24, 10, 15, 13, 9, 13)

   a. Determine for the 8 data items which class it should be placed into.
   b. Having completed part a, recompute the centers for the three classes (average the 7 data points for all of the data found to be in that class). Provide the new centers. Are the new centers better than the originals? Explain.