Data Mining: Assignment 2

Due date: February 7 (Thursday)

Problem 1 (5 points)
Suppose that you are trying to learn the concept of "fast food" based on the following examples:

	place	food type	price	good taste?	includes a plate?
positive	campus	fish	cheap	yes	no
positive	campus	chicken	cheap	no	no
positive	campus	beef	cheap	no	no
negative	campus	beef	medium	yes	yes
negative	mall	fish	costly	yes	no
negative	mall	chicken	cheap	no	yes

(a) Determine the most specific and most general hypotheses consistent with these examples.

(b) Give a new instance that cannot be classified as positive or negative.

(c) Suppose that you can choose additional training examples. Which examples will you request in order to identify the correct hypothesis among the remaining candidate hypotheses?

Problem 2 (5 points)
Implement the candidate-elimination algorithm; it should read a file with training and test examples, use the training examples to determine the most specific and most general hypotheses, and then classify each test example as `positive`, `negative`, or `unknown`. The only required output is the classification of the test examples; it does not have to include the hypotheses themselves. The input format is as follows:

```
<classification> <attribute> <attribute> ... <attribute>
   ...
<classification> <attribute> <attribute> ... <attribute>

<attribute> <attribute> ... <attribute>
   ...
<attribute> <attribute> ... <attribute>
```

The training examples are above the blank line, and the test examples are below. `<classification>` is either `positive` or `negative`, and each `<attribute>` is a string of lower-case letters. The length of an attribute is at most twenty characters; successive attributes are separated by one or more spaces. For instance, the following file includes three training examples and two test examples.

```
positive   campus   fish      cheap    good   noplate
positive   campus   chicken   cheap    bad    noplate
negative   campus   beef      medium   good   plate

campus   fish   cheap    bad    noplate
mall     beef   costly   good   plate
```