Artificial Intelligence: Assignment 5

Due date: November 20 (Tuesday)

Problem 1 (2 points)
Read Chapter 18 and answer the following questions:
(a) What are the main advantages and drawbacks of using Occam's razor in learning?
(b) What are the main limitations of the decision-tree learning?

Problem 2 (3 points)
Suppose that we are trying to learn the concept of "fast food" based on the following examples:

	food type	price	good taste?	includes a plate?
positive	fish	low	yes	no
positive	chicken	medium	no	yes
positive	beef	medium	no	yes
positive	beef	medium	no	no
positive	chicken	low	no	no
negative	beef	high	no	yes
negative	fish	medium	yes	yes
negative	chicken	low	no	yes
negative	fish	high	yes	no
negative	chicken	medium	yes	no
negative	chicken	low	yes	yes
negative	fish	low	yes	yes
negative	beef	high	yes	yes
negative	chicken	high	no	no

Use these data to construct a decision tree; you should compute the information gains to decide which attributes are more important. For each node of the tree, indicate the corresponding information gain.

Problem 3 (5 points)
Implement a program for building decision trees. It should read a file with training and test examples, use the training examples to build a tree, and then classify the test examples. The only required output is the classification of the test examples; it does *not* have to include the tree itself. 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   fish   low      good   noplate
negative   beef   high     bad    plate
negative   fish   medium   good   plate

fish   high     bad    noplate
beef   medium   good   plate
```