Assignment 1
Inductive Learning

Instructions

- This assignment is based on class discussions and chapter 2 of Tom Mitchell's textbook. Read the course material and the chapter carefully before beginning this assignment.
- All work submitted must be your own. Do not copy from online sources. If you use any references, please list them.
- You should use a cover sheet, which can be downloaded from: http://www.utdallas.edu/˜axn112530/cs6375/CS6375_CoverPage.docx
- You are allowed to work in pairs i.e. a group of two students is allowed. Please write the names of the group members on the cover page.
- **You have a total of 4 free late days for the entire semester. You can use at most 2 days for any one assignment. After four days have been used up, there will be a penalty of 10% for each late day. The submission for this assignment will be closed 2 days after the due date.**
- Please ask all questions on Piazza, not via email.

1. Consider the problem of gradient descent that was discussed in class - you would like to predict the number of A grades that a student in the second year of the M.S. program receives (y) based on the number of A grades that the student received in the first year of the M.S. program (x). You propose a hypothesis of the form $h(x) = \theta_0 + \theta_1 x$, where θ_0 and θ_1 are parameters that you want to find. The data is presented below:

x	**y**
3	2
1	2
0	1
4	3

You start with an initial choice of parameters as: $\theta_0 = 0$ and $\theta_1 = 1$. You can assume that the error function is:

$$J = \frac{1}{2m} \sum_{i=1}^{i=m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

where m is the number of training examples. Run at most 5 rounds of the gradient descent algorithm discussed in class. Does your error go down after 5 rounds? Show all the steps of your calculation.

2. Suppose there is a testing machine for a disease that can identify the disease in 80% of the cases, and also in 90% of the cases it is able to correctly predict those who do not have the disease.
Identify the values of False Positive and False Negative in percent

3. What are the pros and cons of the following
a. Selecting the most specific hypothesis (S) based on a training data.
b. Selecting the most general hypothesis (G) based on a training data.

4. What is a consistent hypothesis and version space?

5. The most general hypothesis has __________________ value for each attribute.

6. Consider the ML task of finding an approximation to the job finding problem for UTD students i.e. the function $f : X \rightarrow Y$ where $X = (x_1, x_2, x_3, x_4)$ is a vector of attributes defined below and Y is a Boolean output. $X = (x_1, x_2, x_3, x_4)$ such that
x_1 is a Boolean indicating whether GPA ≥ 3.5
x_2 is a Boolean indicating whether student has taken CS 6375
x_3 is a Boolean indicating whether student has taken CS 6350
x_4 is a Boolean indicating whether student has taken Years of Work Experience > 2

a. How many instances i.e. $|X|$ are possible, assuming each attribute is included?

b. You decide to propose hypotheses in the form of conjunctions of Boolean literals. A Boolean literal can represent either an attribute (e.g. x_1) or negation of that attribute (e.g. $\neg x_1$). You are free to choose any number of literals from 0 to 4. How many such hypotheses are possible?

c. For each attribute, there can be three possible choices (labellings) - 1, 0, or ? (don't care). For example, some examples could be:
$c_1 = (1, 0, ?, ?)$, which can also be stated as: $c_1 = x_1 \wedge \neg x_2$
$c_2 = (0, 0, 1, ?)$, which can also be stated as: $c_2 = \neg x_1 \wedge \neg x_2 \wedge x_3$
For each such choice, you can propose a hypothesis that accepts it as a positive or negative choice. For example, you can create a hypothesis as: $h(X) = c_1$, or $h(X) = \neg c_2$.
This idea is called ***universal concept class***. More details can be found at link 1 or link 2 How many such hypotheses are possible?

d. How many fully-grown decision trees of depth 2 using only 2 attributes are possible?
Hint: You have to think about first selecting 2 attributes and then constructing a DT using them.

e. In how many ways can each tree constructed in part d be labeled? Hint: Labeling a tree implies giving labels to all the leaf nodes of the tree

7. Apply the **Find-S algorithm** on the following dataset for UTD students. There are 5 attributes:
x_{GPA} is a boolean indicating whether GPA > 3.5
x_{WorkEx} is a boolean indicating whether Years of Work Experience > 2
x_{CS6375} is a boolean indicating whether student has taken CS 6375
x_{CS6350} is a boolean indicating whether student has taken CS 6350
x_{Java} is a boolean indicating whether student has taken advanced Java skills

You are given the following dataset along with the class variable (i.e. outcome variable) where 1 indicates student got internship and 0 means student didn't get it. Each data point is in the form $(\langle x_{\text{GPA}}, x_{\text{WorkEx}}, x_{\text{CS6375}}, x_{\text{CS6350}}, x_{\text{Java}}\rangle$, outcome$)$

$(\langle 1, 1, 0, 1, 1\rangle, 1)$
$(\langle 0, 1, 0, 1, 1\rangle, 0)$
$(\langle 1, 1, 1, 1, 0\rangle, 1)$
$(\langle 0, 0, 0, 1, 1\rangle, 0)$
$(\langle 1, 1, 1, 1, 1\rangle, 1)$

8. Consider the decision tree shown below. There are two splitting attributes GPA and years of work experience. The class labels are shown below the leaf nodes.
Write the final hypothesis shown by this decision tree in the form of Disjunctive Normal Form (DNF)

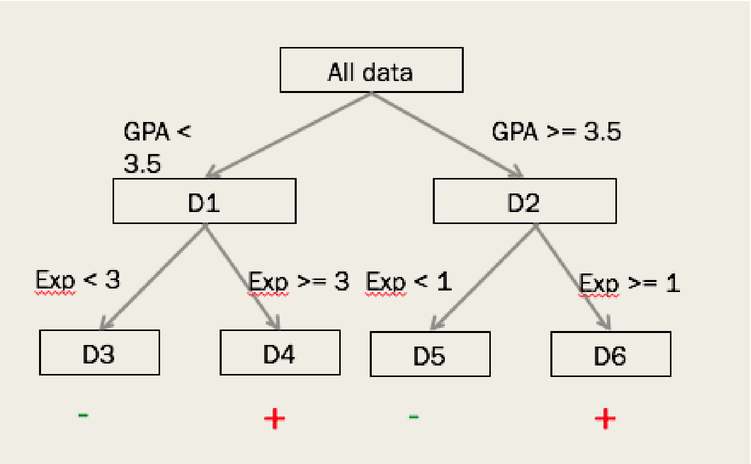

Figure 1: The labels near the leaf nodes represent class attribute i.e. outcome

9. Solve question 2.4 from Tom Mitchell's book

10. In this course, students work in pairs to solve programming assignments. We have kept track of previous groups based on the following attributes for each pair:
Level (ug, gr), where ug: undergraduate, gr: graduate
Major (cs, se, math), where cs: computer science, se: software engineering, math: mathematics
Programming-Background (l, m, h), where l: low, m: medium, h: high
Last-Math-Course (hs, fr, so, ju, se, ma, do), where hs: high school, fr: freshman, so: sophomore, ju: junior, se: senior, ma: master's, do: doctoral

For each attribute, there can be three acceptable constraints in the hypotheses - a specific value, "?" (don't care), or "∅" (null). For example, below is a possible pair of students:
$\langle(\text{ug}, ?, \text{low}, ?), (\text{gr}, ?, ?, \text{ma})\rangle$
It represents cases where the first student is an undergraduate with low programming background, and the second is a graduate student who took his last math course at the master's level.

Corresponding to each pair, we can add a class label as "+" (they teamed up) or "-" (they didn't team up). Below is the training data:

Student Pair Attributes	**Class Label**
$\langle(\text{ug}, \text{se}, \text{l}, \text{hs}), (\text{gr}, \text{cs}, \text{h}, \text{hs})\rangle$	+
$\langle(\text{ug}, \text{se}, \text{h}, \text{fr}), (\text{gr}, \text{cs}, \text{h}, \text{hs})\rangle$	+
$\langle(\text{gr}, \text{se}, \text{l}, \text{so}), (\text{gr}, \text{cs}, \text{h}, \text{se})\rangle$	-
$\langle(\text{ug}, \text{se}, \text{l}, \text{ju}), (\text{gr}, \text{se}, \text{h}, \text{ju})\rangle$	+

a. Run the *CANDIDATE-ELIMINATION* algorithm on the above training data and specify the S and G boundaries after each example.

b. How many total consistent hypotheses are returned after running the *CANDIDATE-ELIMINATION* algorithm on the 4 training examples above? How many of them are consistent with the following data point:

$\langle(\text{ug}, \text{cs}, \text{h}, \text{do}), (\text{gr}, \text{ma}, \text{l}, \text{se})\rangle$, with class label "+"