Inductive Learning

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 is the set of attributes defined below and Y is a boolean output.
   $X = \langle x_1, x_2, x_3, x_4 \rangle$ 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

   For each attribute, there can be three possible choices - 1, 0, or ? (don't care).
   a. How many instances i.e. $|X|$ are possible?
   b. How many labeling of these instances are possible? (Remember it's binary classification problem and each labeling represents a possible hypothesis)
   c. If you would like to limit the classifier to a decision tree of depth 2, how many hypotheses are possible?
   Hint: First choose 2 attributes out of 4, create decision trees out of them, and find possible ways of labeling

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
   $(\langle x_{GPA}, x_{WorkEx}, x_{CS6375}, x_{CS6350}, x_{Java}\rangle$, 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)

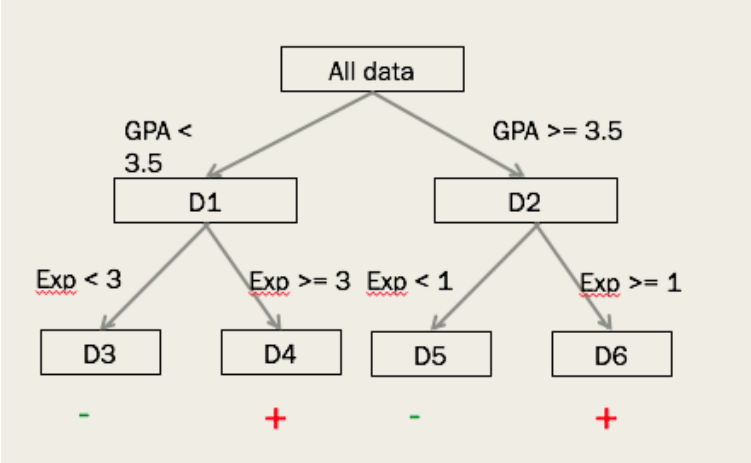

Figure 1: The labels near the leaf nodes represent class attribute i.e. outcome

9. Solve question 2.4 from Tom Mitchell's book

10. Solve question 2.5 from Tom Mitchell's book