CSE P546 Data Mining Homework 2

Due Date: May 9. We would prefer that you turn in a hard copy of your solutions at the start of class. However, you can also email them to bhushan@cs. Your file must contain your name at the top, and can be in any of pdf, Word or plaintext formats.

1. Mitchell 10.1

2. Suppose FOIL is considering adding a literal to a clause using a binary predicate P.

   (a) If the clause currently contains n different variables, how many distinct literals can be generated? Two literals are considered identical if they differ only in the names of the new variables that they contain.

   (b) Why do you think FOIL doesn't allow literals that contain no previously used variables?

3. Planets around distant stars can sometimes be detected by the tiny wobble they cause in the star's trajectory. Suppose that 2% of all stars have planets orbiting them, 75% of stars with planets show a wobble in their trajectory, and 10% of stars without planets show a wobble due to other causes. If a star shows a wobble in its trajectory, what is the probability that it has planets orbiting it?

4. Mitchell 6.6

5. Mitchell 4.2

6. What is the result of applying crossover to the strings 00110 and 10111, with crossover mask 11100?

7. Decision stumps are decision trees with only one node (the root). Which ensemble method would you use to improve the accuracy of decision stumps - bagging or boosting? Why?