CS 6604 Assignment #3
Assigned: Nov 7, 2007
Date Due: Nov 16, 2007, 11:59:59pm, by email to the instructor

---

1. (10 points) Give five examples of data mining problems for which relational mining techniques such as ILP are necessary. Your examples must be those that have not been covered in class. Argue with 1-2 sentences (each) why non-relational techniques will not work or be effective.

2. (15 points) Consider a relational data mining problem such as the animals example studied in class. Explain how you can 'transform' the dataset for this problem so that you can mine it with a decision tree algorithm and obtain the same result as if you had applied a relational data mining algorithm.

3. (20 points) In class, we worked out the posterior distribution for the probability of heads of a given coin based on data. Assume a slightly modified version of this experiment as follows: We toss a coin (call it coin 1, or the first coin) that comes up with heads in some unknown probability. Now, based on the results of this coin, we choose to toss either coin 2 (if coin 1 falls heads) or coin 3 (if coin 1 falls tails). Both coin 2 and coin 3 have their own (unknown) probabilities of heads. Over a lot of trials, you are given both the results of the first coin and the results of the second coin (but you are not told which of coins 2 and 3 resulted in the second value). Explain how you can compute all the three unknown probabilities using this information. For full credit, identify a prior, define a likelihood function, and update the posterior systematically as data arrives. Use the technique described in class to identify the maximum of the posterior distribution and hence the unknown probabilities.

4. (25 points) The uniform prior we have used in class belongs to a class of distributions known as the Beta distribution. This distribution has the property that, when used for reasoning about a boolean variable (such as the coin toss), the posterior is also a Beta distribution! First, research the definition of the Beta distribution and prove that this is the case. Second, explain why this property is important and how you can design efficient algorithms for probabilistic reasoning that use this property.

5. (10 points) In class, we learnt that the Naive Bayes classifier makes a strong assumption, namely that attributes are independent given the class. Give an example of a dataset where the attributes are not independent given the class so that the Naive Bayes classifier makes wrong classifications.

6. (20 points) Same as above, except give an example of a dataset where the attributes are not independent given the class but yet the Naive Bayes classifier makes the right classification. How can this be?

Turnin a typed (not handwritten) paper copy giving answers to the questions, plots, including a brief description of how you solved each question. Write enough to convince us that you completed the assignment independently.