Department of Computer Science and Engineering
University of California at San Diego

CSE 255
Winter 1998

Midterm Examination

Thursday February 12, 2:20pm to 3:40pm

Your name:

Instructions: This exam has 5 pages and 65 points in total. Look through the whole exam and answer the questions you find easiest first.

Whenever you make an assumption, state it clearly. In questions where you are asked to discuss an issue, it is important that you use relevant concepts and terminology as covered in class and/or in textbooks and handouts.

You may use a calculator and your own personal lecture notes, the online class lecture notes, your own old assignments, and all handouts and class textbooks. Other books or notes may not be used.

(Question 1) [15 points. This question was originally written by Zachary Dodds of Yale.]

A disgruntled student, we'll call her Chelsea, is considering transferring to two other schools that had caught her eye when checking out colleges: Yale and the Citadel. She likes the schools equally, and thus will base her decision on financial concerns. In particular, she is concerned because her father (due to other financial pressures) is leaving to her the cost of appropriate protection (i.e. bodyguards) at each place.

Regarding the Citadel, Chelsea feels that \$60,000 worth of bodyguard-hours would be sufficient there. Regarding Yale, she feels that if the mayor's pending "gated community" initiative is adopted by the city council, \$20,000 of guard-hours will suffice. On the other hand, if the initiative does not pass, she thinks that \$200,000 will be required. Upon initial inquiry, David Morris (a family friend near Yale), estimates that there is a 60% chance that the initiative will pass.

(a) [5 points] Given this information, what should Chelsea do? What will the savings be over the alternative?

(b) [10 points] Mr. Morris, looking for work, offers to research the issue more thoroughly. He can provide a report that will be $x\%$ accurate as to whether or not the initiative will pass. That is, Pr(report=pass | initiative=pass) = $x\%$, and similarly for report=fail and initiative=fail.

Mr. Morris is thinking about charging \$100 times x for his services, since more work will need to be done in order to increase x. Explain whether Chelsea should take him up on this offer.

(Question 2) [19 points] [This question is based on the problems on page 110 of the book by Gonzalez and Dankel.]

Suppose that PERS (Printer Emergency Resolution System, pronounced "purrs") is a knowledge-based system that solves problems with the printer of a personal computer. For example, if you issue a "print" command and nothing happens, PERS can tell you what to do. Assume that PERS is implemented using CLIPS.

(a) [3 points] Sketch an overall design for PERS, using forward chaining.

(b) [3 points] Is forward chaining appropriate for PERS? Why or why not?

(c) [3 points] Does the order in which CLIPS rules are written have any significance for a forward-chaining system such as PERS?

(d) [5 points] Give an example of some knowledge needed by PERS that should be represented in an object-oriented way. Explain why objects and classes are right for this knowledge, and sketch how to write down the knowledge using CLIPS 6.0 syntax.

(e) [5 points] Modern programming languages such as C++ and Java have "if-then" constructs, and they are object-oriented. So why not use one of these languages to implement a system such as PERS? Why might CLIPS be preferable?

(Question 3) [16 points]

This question concerns forecasting the stock market. Consider the following scenario. Every day 50% of stocks go up and 50% go down. For each stock, there is no correlation from day to day. There is also no correlation between stocks. If a stock goes up then down, or down then up, then with equal probability it finishes net down or net up. Assume for simplicity that stocks are always up or down: their price is never exactly the same on two different days.

Suppose we claim to have a data mining method that predicts at the end of each day for each stock whether it will go up or go down in the future. Each day, for each stock, we either keep our forecast unchanged, or we reverse it.

For a given stock, suppose that the most recent day on which we reversed our forecast was day t, and that the current forecast (i.e. the result of the reversal) is "up" ["down"]. We define the forecast to be "correct" on day $t + k$ where $k \geq 1$ if the price at the end of day $t + k$ is higher [lower] than the price at the end of day t.

Our forecasting method is in fact very simple. At the end of each day, let the forecast remain unchanged for all stocks for which their current forecast is "correct". For all other stocks, reverse the forecast. Before day 1 we start with a random "up" or "down" forecast for each stock.

Note: Professor Howard Phillips of the University of North Carolina at Charlotte publishes stock forecasts at `http://www.coe.uncc.edu/~hphillip/`. The accuracy of these forecasts is measured in the way described above. Dr. Phillips systematically obtains accuracies around 75%.

(a) [8 points] Explain why this forecasting method will lead to measured forecasting accuracies of well over 50%.

(b) [8 points] Explain why this forecasting method can or cannot be used to buy and sell stocks profitably.

(Question 4) [15 points] Suppose that you are a factory manager in a high-tech manufacturing company. One of the engineers working for you will leave unless she is allowed to work on what she wants to, which is a data mining project.

Based on your own experience, whether in CSE255 or elsewhere, and in your own words, give three important, concrete guidelines for the data mining project, and explain why they are important. At least one guideline should be software-related, and at least one should be business-related.

(a) [5 points] Guideline 1.

(b) [5 points] Guideline 2.

(c) [5 points] Guideline 3.