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

CSE 255
Winter 1997

Final Examination

Your name: **Tim Leek**

Instructions: This exam has 3 pages and 100 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) [25 points] Suppose the task is to choose a policy for granting or rejecting loan applications. As in the fourth project assignment, the payoff matrix is

	repays	defaults
grant	x	$-4x$
reject	0	0

where x is the size of the loan. In the training and testing populations approximately 70% of the applicants repay, and 30% do not, i.e. they default.

(a) [5 points] Explain why a policy of the form "grant the loan if and only if $Pr(\text{repays}) \geq t$" is optimal, and calculate the optimal value of the probability threshold t.

(b) [5 points] Your manager has noticed that customers with large loans tend to default more often. His calculations give the following table:

decile	loan size	Pr(repays)
1	929	0.7
2	1258	0.7125
3	1474	0.75
4	1908	0.75
5	2320	0.725
6	2848	0.7875
7	3609	0.775
8	4844	0.6875
9	7393	0.6875
10	18424	0.475

Here deciles are according to loan size. The middle column gives the size of the largest loan in each decile, and the third column gives the observed frequency of repayment for loan sizes in the indicated decile. For example, reading from the first and second rows, the second smallest 10% of loans had size 930 DM to 1258 DM, and 71.25% of these loans were repaid.

Explain why there is no policy of the form "grant all loans of size less than z" that can be expected to be profitable.

(c) [5 points] Assuming that a policy of the form "grant all loans of size less than z" could be profitable, suggest two important practical reasons why it would still be undesirable compared to a policy based on a naive Bayes classifier.

(d) [5 points] Suppose that you want to use a naive Bayes classifier and MLC++. Given a training set and a test example, MLC++ tells you whether $Pr(\text{repays}) > 0.5$ but does not tell you the actual value of $Pr(\text{repays})$. As discussed in class, it is possible to replicate "default" examples in the training set to simulate decisions based on the optimal probability threshold. Write down the equations defining a naive Bayes classifier and use them to explain informally how to do this.

(e) [5 points] Calculate exactly how many times each "default" example should be replicated in order to simulate decisions based on the optimal probability threshold t of part (a).

(Question 2) [25 points] Consider the problem of designing an automatic air conditioning controller for a car. (Ignore the ethical question of whether such a controller is too sybaritic.) Since people feel warmer psychologically when the sun is shining, the target temperature ranges from 22 degrees Celsius at night to 20 degrees Celsius in bright sunshine. Two sensors are available. One measures the current temperature inside the car, in degrees Celsius. while the other measures the brightness of the light reaching the dashboard, in a range from 0% to 100% where 0% represents a cloudy, moonless night and 100% represents tropical sun at noon. Two actuators are available. One controls the speed of the fan over a continuous range from 0% to 100% and the other switches the air conditioning compressor on or off.

(a) [5 points] The air conditioning compressor can only be switched on or off–it cannot be partially on. How is this contrary to the spirit of fuzzy control? How might you cope with this difficulty?

(b) [5 points] Define trapezoidal fuzzy sets representing qualitative values for each sensor and actuator as appropriate. Briefly explain your design decisions.

(c) [5 points] Using the fuzzy sets you defined above, give a plausible small set of fuzzy rules for this control problem. Again, briefly explain your design decisions.

(d) [5 points] Assume that the temperature inside the car is 23 degrees Celsius and the strength of the sun is 60%. Trace through how a fuzzy control inference engine would use your rules to choose settings for the fan and compressor actuators. (Assume that all arithmetic uses real numbers with no special concern for implementation efficiency.)

(e) [5 points] "Fuzzy controllers have no state." Explain what this statement means. Would this cause difficulties in automotive air conditioning control?

(Question 3) [20 points] [This question is based on Exercise 12.17 in *Object-oriented AI using C++* by Kim W. Tracy and Peter Bouthoorn, W. H. Freeman and Co., 1997.]

(a) [5 points] Sketch the overall design of a multimedia case-based reasoning system that will help people choose which house to buy. The software will take as input the house features that a customer wants, such as the number of bathrooms. The case library will consist of houses for sale.

(b) [5 points] Discuss what knowledge should be represented formally in this system, and what knowledge should remain informal in multimedia documents.

(c) [5 points] Outline a formal representation for cases using the object-oriented features of CLIPS.

(d) [5 points] Using your representation of cases, develop an appropriate metric for deciding how similar a customer's requirements are to an existing house. Sketch how to code this metric as a decision tree or as CLIPS rules.

(Question 4) [15 points] (a) [5 points] What is a "closed-world assumption"? Give a logically precise and general explanation, and then give an illustrative example.

(b) [5 points] Is it possible for a closed-world assumption to destroy the soundness of a knowledge base? Is it possible for it to destroy completeness? For each question, if yes, give an example, and if no, explain your answer with precision.

(c) [5 points] Consider a CLIPS-style knowledge base where all facts (but not rules) have probabilities associated with them instead of being just true or false. What sorts of closed-world assumptions make sense for such a knowledge base? (Hint: what is the ontology of such a knowledge base?)

(Question 5) [15 points] Do you think that there are great opportunities in industry and business to apply AI techniques more than is currently done? Or do you think that applied AI is a niche technology that is already reasonably close to its maximum level of deployment?

Base your answer on what you have learned in CSE255 and on your general knowledge and work experience. Be sure to give arguments and examples to support your claims. Answer based on existing, proven AI techniques—do not speculate about techniques that may be refined or invented in the future.