CIS-445 DATA MINING

Test #4 (100 points)

Name of student and signature: **Damon Quire**

This is a take-home, closed book, and closed notes test. You can have a 3×5" cheat card filled in with anything you want on one side of the card. You may use a calculator and/or Excel. You have 80 minutes to complete the test and turn in in via e-mail only to jozef.zurada@louisville.edu by Friday, December 8, Midnight. You cannot use any other help on the test. Hand-write your answers on this test, convert to a pdf file, and e-mail the pdf file to me. Name the file as YourFirstLastName_Test4.pdf.

By putting my name and signature above I certify that I have adhered to an honor code and the above instructions while writing this test.

I. Fill out blanks – 34 points (30 regular points, 4 extra credit points)

1. In a fuzzy system, knowledge is carried both in its **fuzzy rules and fuzzy sets**

2. Three Darwin's basic principles that genetic algorithms use to create solutions for problems are:

   - **Survival of the fittest**
   - **Cross-breeding**
   - **Mutation**

3. What class of problems/tasks are genetic algorithms good for? **Optimization problems**

4. A **Membership Function** is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1.

5. The antecedent of a fuzzy rule has three parts combined by the AND operators. Parts one, two, and three of the antecedent are true to the degree of 0.5, 0.1, and 0.9, respectively. To what degree is the consequent of the rule true, if the minimum operator is used? **.9**

6. The antecedent of a fuzzy rule has three parts combined by the OR operator. Parts one, two, and three of the antecedent are true to the degree of 0.5, 0.1, and 0.9, respectively. To what degree is the consequent of the rule true, if the maximum operator is used? **.9**

7. Genetic algorithms always guarantee an optimal solution (True/False) **False**

8. Inputs to and outputs from the fuzzy logic system are crisp numbers (True/False) **True**

9. A salesman is going to visit 8 cities. He can visit each city only once and start travel from any city. What is the total number of the possible routes/permutations?
   (7*6*5*4*3*2*1)/2=2520

   What is the number of the unique/distinct routes/combinations? **1260**

10. Give an example of a crisp rule
If student gets a 90% or higher in his final class, student graduates.

11. Give an example of a fuzzy rule
If food is good, then tip is average.

12. **Fuzzy Inference Process** is the process of formulating the mapping from a given input to an output using fuzzy logic (2 points – **extra credit**).

13. In the vector game, 1001101 is a string to guess. What is the average number of attempts that would allow one to guess the string? **2^7=128/2=64** (2 points – **extra credit**)

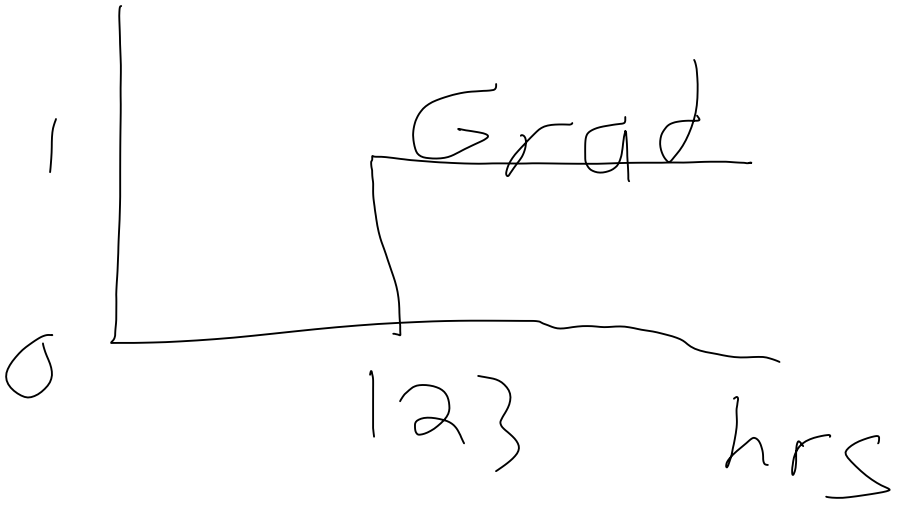
II. Short essay questions (10 points each). Answer any three out of the four questions. Circle the questions you have chosen to answer. If you do not, the first three questions will be graded. (30 points)

1. Discuss why some problems such as the Traveling Salesman Problem (TSP) require the use of genetic algorithms. Explain why is the TSP problem called NP-complete problem?

2. Explain the difference between a crisp set and a fuzzy set. Draw a membership function for each of the 2 sets and briefly explain.

A fuzzy set has no crisp or abrupt boundaries. It also allows for partial membership as well. An object can possibly belong to more than one fuzzy set and have a partial degree of membership. They can also overlap each other: in this example the height of a building can belong to 2 different fuzzy sets at once.

Crisp sets have abrupt boundaries and allow for no partial memberships or overlapping sets. There is only a specific result based on the input: In this example if a student as completed > or = to 123 hours they will graduate.



3. What is the function of the Filter Node in SAS Enterprise Miner 13.1?

The filter node is used to filter out outliers so they will not skew the results of the next nodes. This is to help ensure an accurate representation of a population if using sampling, or simply filter out uniquely high or low or flat out different values to eliminate them from effecting the results negatively.

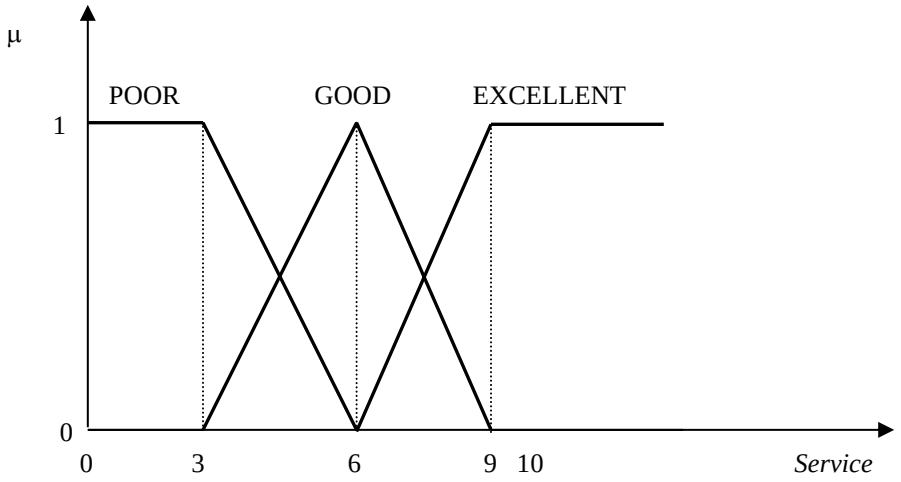
4. Give at least four examples of optimization problems/tasks.

- **Planning a schedule**
- **Deciding which route to take on a multi-stop vacation**
- **Scheduling classes**
- **How to structure an application's code**
- **When to leave for a night out of the house**

All of these problems/tasks require the need to minimize something or gain efficiency.

III. Problem questions. (40 points)

1. There are three fuzzy sets representing POOR, GOOD, and EXCELLENT *Service* you have received in a restaurant on the scale from 1 to 10. They are described by the three membership functions shown on the chart below. Compute to what degrees *Service* = 2, 4, and 8 belongs to the three fuzzy sets. In addition, to what degree *Service*=10 belongs to the fuzzy set POOR. (Note that an item may belong to more than one fuzzy set at the same time.) (20 points)



Service	μ POOR (*Service*)	μ GOOD (*Service*)	μ EXCELLENT (*Service*)
2	**1**	**0**	**0**
4	**.667**	**.333**	**0**
8	**0**	**.333**	**.667**
10	**0**		

2. Using genetic algorithms you are trying to find the maximum value of a function $y=127x-x^2$, where x is a whole number within the range [0,127]. You need 7 bits to represent numbers from within this range. The function peaks at the integer values 63 and 64 represented as binary 0111111 and 1000000. You have generated at random an initial population of 5 chromosomes shown in the table below. (20 points)

Calculate the value of x, the value of the fitness function, and the percentage of total population fitness.

Generation 1 of solutions

Chromosome #	Chromosome	X	Fitness	Percentage of total population fitness
1	1010111	**87**	**3480**	**21.58**
2	0010101	**21**	**2226**	**13.809**
3	1010011	**83**	**3652**	**22.65**
4	0111001	**57**	**3990**	**24.75**
5	1100011	**99**	**2772**	**17.19**

Average: **3224**

If you spin the roulette wheel for generation 1 of solutions

What chromosome is the most likely to be chosen for generation 2 of solutions?
4

What chromosome is the least likely to be chosen for generation 2 of solutions?
2

For generation 1 of solutions:

- Perform crossover for chromosomes 1 and 2 after bit 3 (count the leftmost bit as bit number 1) to create two offsprings, chromosomes 6 and 7. Replace chromosomes 1 and 2 with the two offsprings in generation 2 of solutions.

101|0111|
001|0101|

- Perform mutation at bit 2 for chromosome 5. Replace chromosome 5 after mutation in generation 2 of solutions to create chromosome 8.

1000011

- Chromosomes 3 and 4 will be unchanged. Compute x, the fitness, and the percentage of total population fitness for each chromosome in generation 2 of solutions.

Generation 2 of solutions

Chromosome #	Chromosome	X	Fitness	Percentage of total population fitness
6	1010101	**85**	**3570**	**20.25**
7	0010111	**23**	**2392**	**13.57**
3	1010011	**83**	**3652**	**20.7**
4	0111001	**57**	**3990**	**22.64**
8	1000011	**67**	**4020**	**22.81**

Average: **3524.8**

Is the average fitness of generation 2 of solutions better than for generation 1 of solutions?
yes because it has the higher value and we are looking for the max value.