UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Engineering, Built Environment and IT
Department of Computer Science

COS 314

Tutorial/Practical 6 Solutions

26 May 2022

Questions

Age	Vision	Astig	Tears	Class
1	1	1	1	3
1	1	1	2	2
1	1	2	1	3
1	1	2	2	1
1	2	1	1	3
1	2	1	2	2
1	2	2	1	3
1	2	2	2	1
2	1	1	1	3
2	1	1	2	2
2	1	2	1	3
2	2	2	2	3
3	1	1	1	3
3	1	2	1	3
3	2	1	1	3
3	2	1	2	2
3	2	2	2	3
3	1	2	2	1

Age = 1 - young , middle , elderly **Vision** = 1-short, 2-long, **Astigmatic** = 1 -no, 2 -yes,
Tears = 1- reduced, 2- normal**, Class (contact lenses)** - 1- hard, 2-soft , 3-none

Figure 1: Contact Lenses Data

1. Given the following dataset which relates to the classification of whether an individual qualifies for contact lenses or not. Use the Information Gain Decision Tree induction method to determine which parameter will be the root node of the decision tree. (10 Marks)
   Sol: First calculate the entropy of the complete dataset.
   $\mathbf{E}_{max} = -\frac{11}{18}\ \mathbf{log_2}(\frac{11}{18}) - \frac{4}{18}\ \mathbf{log_2}(\frac{4}{18}) - \frac{3}{18}\ \mathbf{log_2}(\frac{3}{18}) = \mathbf{1.35.}$
   Then the entropy for attribute Age
   $\mathbf{E}_{age1} = -\frac{2}{8}\ \mathbf{log_2}(\frac{2}{8}) - \frac{2}{8}\ \mathbf{log_2}(\frac{2}{8}) - \frac{4}{8}\ \mathbf{log_2}(\frac{4}{8}) = \mathbf{1.50}$
   $\mathbf{E}_{age2} = -\frac{0}{4}\ \mathbf{log_2}(\frac{0}{4}) - \frac{1}{4}\ \mathbf{log_2}(\frac{1}{4}) - \frac{3}{4}\ \mathbf{log_2}(\frac{3}{4}) = \mathbf{0.811,}$
   $\mathbf{E}_{age3} = -\frac{1}{6}\ \mathbf{log_2}(\frac{1}{6}) - \frac{1}{6}\ \mathbf{log_2}(\frac{1}{6}) - \frac{4}{6}\ \mathbf{log_2}(\frac{4}{6}) = \mathbf{1.25}$

E_{age} = $\frac{8}{18}$ (1.5)+$\frac{4}{18}$ (0.8)+$\frac{6}{18}$(1.25) = 1.26.
Gain = E_{max} - E_{age} = 1.35 - 1.26 = 0.09
The process is repeated for the other attributes and the attribute with the highest gain is selected as the root node.

2. Given the following measurements pertaining to 16 objects. Using the following values as the initial centroids **C1 - (3.8 ,9.9) C2-(7.8,12.2) and C3 (6.2, 18.5)**. Perform three iterations of the K-means algorithm showing values of the centroid updates at each iteration. (9 Marks)

x	y
6.8	12.6
0.8	9.8
1.2	11.6
2.8	9.6
3.8	9.9
4.4	6.5
4.8	1.1
6.0	19.9
6.2	18.5
7.6	17.4
7.8	12.2
6.6	7.7
8.2	4.5
8.4	6.9
9.0	3.4
9.6	11.1

Figure 2: Object Metrics

	Initial		First Iteration		Second Iteration	
	x	y	x	y	x	y
Centroid 1	3.8	9.9	4.6	7.1	5.0	3.8
Centroid 2	7.8	12.2	8.2	10.7	8.1	12
Centroid 3	6.2	18.5	6.6	18.6	6.6	18.6

Figure 3: Solution 1st 2 iterations

3. Given the following algorithms are used to solve the TSP problem:
(i) Ant Colony Optimisation
(ii) Particle Swarm Optimisation
You are tasked with selecting an Evolutionary algorithm to configure the two given Optimisation algorithms. Based on your selection answer the following questions for each algorithm.
   (a) Describe the representation you will use. (your description should be algorithm and problem specific). (3)
   Sol: The representation should be able to represent the components or parameters of each of the algorithms. Eg GA chromosome will be a design of the algorithm where each gene represents an attribute of ACO or PSO
   (b) What components of the configured algorithms will make up an individual of the EA. (3)
   Sol: If a GA is selected for PSO a chromosome will be made of attributes that pertain to the

global algorithm e.g w, l_1, r_1, l_2 *and* r_2. not the lower level individual attributes for the solution space such as velocity or position

(c) Describe how you will evaluate the fitness of each individual (3)
Fitness of the individual of the high level will be the global best solution produced by that design.

(d) How will you update the population. (3)
Steady state update is the best update strategy as the population size of the designs will be low

(e) Define the stopping criteria. (3)
Sol: When there is no global improvement