

Student Number:

Student Name:

Date:

CS5483 Review exercise 10

Given the following training data set

Training instance	Income range	Credit card insurance	Sex	Age
1	30-40k	Yes	Male	30-39
2	30-40k	No	Female	40-49
3	50-60k	No	Female	30-39
4	20-30k	No	Female	50-59
5	20-30k	No	Male	20-29
6	30-40k	No	Male	40-49

Describe the steps needed to apply unsupervised genetic learning to cluster the instances of the credit card promotion database?

Model answer:

Step 1: Transform the data to a form suitable for a genetic algorithm.

Step 2: Define the number of the classes of the clusters requested. In this case, it should be the class of the instances that accept the credit card promotion by use of genetic learning.

Step 3: Set up a set of initial solutions, that is, the medoid (center pointer) of each cluster.

Step 4: Perform fitness function of the solution against a given training data set by use of Euclidean distance between $(X_{i1}, X_{i2} \dots X_{ip})$ and $(X_{j1}, X_{j2} \dots X_{jp})$ as shown below:

$$d(i, j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)}$$

Step 5: Revise the solution by crossover or mutation if the fitting score does not meet the users' requirements.

Step 6: Repeat step 4 and step 5 until termination conditions of all elements in the solution meet the criteria or reach the maximum iterations of the genetic learning.

Step 7: Use the resultant solution to classify new data instance into each cluster.