

Student Number:

Student Name:

Date

CS5483 Tutorial question 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

After transforming the input data into numeric such as yes=1, no=2, male=1, female=2, 20-29=1, 30-39=2, 40-49=3, 50-59=4, 20-30k=1, 30-40k=2, 40-50k=3, 50-60k=4, the training data set becomes:

$T(1)=(2,1,1,2)$

$T(2)=(2,2,2,3)$

$T(3)=(4,2,2,2)$

$T(4)=(1,2,2,4)$

$T(5)=(1,2,1,1)$

$T(6)=(2,2,1,3)$

Assume there are two set of initial population for two clusters as:

Solution 1 of 2 clusters centers: $K1(1,1,1,1), (4,2,2,4)$

Solution 2 of 2 clusters centers: $K2(4,4,4,4), (2,2,1,1)$

Choose the best solution based on their fitness function score by use of unsupervised genetic learning.

Model answer:

$$\begin{aligned}
F(K1) &= (\min (\text{Squareeroot}(|1-2|^2+|1-1|^2+|1-1|^2+|1-2|^2), (\text{Squareeroot}(|4-2|^2+|2-1|^2+|2-1|^2+|4-2|^2)) \\
&+ (\min (\text{Squareeroot}(|1-2|^2+|1-2|^2+|1-2|^2+|1-3|^2), (\text{Squareeroot}(|4-2|^2+|2-2|^2+|2-2|^2+|4-3|^2)) \\
&+ (\min (\text{Squareeroot}(|1-4|^2+|1-2|^2+|1-2|^2+|1-2|^2), (\text{Squareeroot}(|4-4|^2+|2-2|^2+|2-2|^2+|4-2|^2)) \\
&+ (\min (\text{Squareeroot}(|1-1|^2+|1-2|^2+|1-2|^2+|1-4|^2), (\text{Squareeroot}(|4-1|^2+|2-2|^2+|2-2|^2+|4-4|^2)) \\
&+ (\min (\text{Squareeroot}(|1-1|^2+|1-2|^2+|1-1|^2+|1-1|^2), (\text{Squareeroot}(|4-1|^2+|2-2|^2+|2-1|^2+|4-1|^2)) \\
&+ (\min (\text{Squareeroot}(|1-2|^2+|1-2|^2+|1-1|^2+|1-3|^2), (\text{Squareeroot}(|4-2|^2+|2-2|^2+|2-1|^2+|4-3|^2)) \\
&= 1.414 + 2.236 + 2 + 3 + 1 + 2.449 \\
&= 12.099
\end{aligned}$$

$$\begin{aligned}
F(K2) &= (\min (\text{Squareeroot}(|4-2|^2+|4-1|^2+|4-1|^2+|4-2|^2), (\text{Squareeroot}(|2-2|^2+|2-1|^2+|1-1|^2+|1-2|^2)) \\
&+ (\min (\text{Squareeroot}(|4-2|^2+|4-2|^2+|4-2|^2+|4-3|^2), (\text{Squareeroot}(|2-2|^2+|2-2|^2+|1-2|^2+|1-3|^2)) \\
&+ (\min (\text{Squareeroot}(|4-4|^2+|4-2|^2+|4-2|^2+|4-2|^2), (\text{Squareeroot}(|2-4|^2+|2-2|^2+|1-2|^2+|1-2|^2)) \\
&+ (\min (\text{Squareeroot}(|4-1|^2+|4-2|^2+|4-2|^2+|4-4|^2), (\text{Squareeroot}(|2-1|^2+|2-2|^2+|1-2|^2+|1-4|^2)) \\
&+ (\min (\text{Squareeroot}(|4-1|^2+|4-2|^2+|4-1|^2+|4-1|^2), (\text{Squareeroot}(|2-1|^2+|2-2|^2+|1-1|^2+|1-1|^2)) \\
&+ (\min (\text{Squareeroot}(|4-2|^2+|4-2|^2+|4-1|^2+|4-3|^2), (\text{Squareeroot}(|2-2|^2+|2-2|^2+|1-1|^2+|1-3|^2)) \\
&= 1.414 + 2.23 + 2.44 + 3.31 + 1 + 2 \\
&= 12.368
\end{aligned}$$

Since F(K1) has a lower fitness score (i.e. closer neighborhood), F(K1) is a better solution than F(K2).