Proximity - Nominal attributes.

ID	Test Result
1	Code A
2	Code B
3	Code C
4	Code A.

p - no. of attributes

m - matches.

$$d(i,j) = \frac{p-m}{p}$$

p=1 (Test Result is the only attribute)

$$d(2,1) = \frac{1-0}{1} = 1$$

$$d(3,1) = \frac{1-0}{1} = 1$$

$$d(3,2) = \frac{1-0}{1} = 1$$

$$d(4,1) = \frac{1-1}{1} = 0$$

$$d(4,2) = \frac{1-0}{1} = 1$$

$$d(4,3) = \frac{1-0}{1} = 1$$

	1	2	3	4
1	0			
2	1	0		
3	1	1	0	
4	0	1	1	0

∴ (4,1) — similar

proximity - Ordinal attributes

Step 1:- Count states - m_f

Step 2:- Rank the state values

Step 3:- normalize the ranking

$$z_{if} = \frac{r_{if} - 1}{m_{if} - 1}$$

ID	Test 2 (ordinal)
1	Excellent
2	fair
3	Good
4	Excellent

Ranking: Fair (1)
Good (2)
Excellent (3)

z_{if} = fair(1) = $\frac{1-1}{3-1} = 0$

Good(2) = $\frac{2-1}{3-1} = 0.5$

Excellent(3) = $\frac{3-1}{3-1} = 1$

	1	2	3	4
1	0			
2	1	0		
3	0.5	0.5	0	
4	0	1	0.5	0

New table

ID	Test 2 (ordinal)
1	1
2	0
3	0.5
4	1

Use manhattan distance for this problem

$d(2,1) = 0-1 = 1$

$d(3,1) = 0.5-1 = 0.5$

$d(3,2) = 0.5-1 = 0.5$

$d(4,1) = 1-1 = 0$

$d(4,2) = 1-0 = 1$

$d(4,3) = 1-0.5 = 0.5$

$\therefore$ (4,1) are similar

Numerical Attribute

Normalize to [0-1]

Id	Test 3
1	45
2	22
3	64
4	28

$$d_{ij} = \frac{|x_{ij} - x_{ji}|}{\text{Max} - \text{min}}$$

Calculate nominal, ordinal, numerical separately

	1	2	3	4
1	0			
2	0.55	0		
3	0.45	1	0	
4	0.40	0.14	0.86	0

$$d(2,1) = \frac{|45-22|}{64-22} = 0.55$$

$$d(3,1) = \frac{64-45}{64-22} = 0.45$$

Mixed attribute.

	1	2	3	4
1	0			
2	0.85	0		
3	0.65	0.83	0	
4	0.13	0.71	0.79	0

$$d(i,j) = \frac{\sum_{i=1}^{p} \delta_{ij} d_{ij}}{\sum_{i=1}^{p} \delta_{ij}}$$

$\delta_{ij} = 0$, if any data is missing or zero.

$\delta_{ij} = 1$, otherwise

$$d(2,1) = \frac{(1 \times 1) + (1 \times 1) + (0.55 \times 1)}{1+1+1} = 0.85$$

$$d(3,1) = \frac{(1 \times 1) + (1 \times 0.5) + (0.45 \times 1)}{1+1+1} = 0.65$$

Dissimilarity b/w binary Attributes

$$\begin{array}{cc|cc|c} & & \multicolumn{2}{c|}{\text{obj } j} & \\ & & 1 & 0 & \\ \hline \text{obj } i & 1 & q & r & q+r \\ & 0 & s & t & s+t \\ \hline & \text{sum} & q+s & r+t & p \end{array}$$

$p = q+r+s+t$

$$d(i,j) = \frac{r+s}{q+r+s+t}$$ ← Symmetric binary: two state are equally important

$$d(i,j) = \frac{r+s}{q+r+s}$$ ← Asymmetric binary: only one state (0–0), is not important

Name	Test 1	T2	T3	T4	T5	T6
Jack	1	0	1	0	0	0
Jim	1	0	1	0	1	0
Mary	1	1	0	0	0	0

$$d(i,j) = \frac{m_{10} + m_{01}}{m_{11} + m_{10} + m_{01}}$$

$$d(\text{Jim}, \text{Jack}) = \frac{1+0}{2+1+0} = 1/3 = 0.33$$

$d(\text{Jack}, \text{Mary}) =$

Higher dissimilarity is b/w Jim and Mary.

Case on Asymmetric binary Attributes

	Jack	Jim	Mary
Jack	0	0.33	
Jim	0.33	0	
Mary	0.67	0.75	0

1114 do it for Symmetric binary.