Example 1:

Object identifier	test-1	test-2	test-3
	Nominal	ordinal	numeric
1	code A	excellent	45
2	code B	fair	22
3	code C	good	64
4	code D	excellent	28

Attribute Type	Dissimilarity	Similarity
Nominal	$d = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$	$s = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$
Ordinal	$d = \lvert x - y \rvert / (n-1)$ (values mapped to integers 0 to $n-1$, where n is the number of values)	$s = 1 - d$
Interval or Ratio	$d = \lvert x - y \rvert$	$s = -d$, $s = \frac{1}{1+d}$, $s = e^{-d}$, $s = 1 - \frac{d - min\_d}{max\_d - min\_d}$

We will have 3 matrices one for the Nominal attribute, one for the ordinal, one for numeric.

Nominal

0			
1	0		
1	1	0	
0	1	1	0

Ordinal

0			
1.0	0		
0.5	0.5	0	
0	1.0	0.5	0

Let use look at the ordinal attribute, fair, good, and excellent. Mf = 3, three states. Let us replace each value by its rank, ranks. Fair is 1, good is 2 and excellent is 3. Let us use the following normalization formula Zif=(rif – 1) /Mf – 1 .

A database may contain all attribute types

- Nominal, symmetric binary, asymmetric binary, numeric, ordinal

- One may use a weighted formula to combine their effects

Object identifier	test-1	test-2	test-3
	Nominal	ordinal	numeric
1	code A	excellent	45
2	code B	fair	22
3	code C	good	64 $maxhxhf = 64$ $minhxhf = 22$,
4	code D	excellent	28

Let us compute the dissimilarity matrix for the numeric attribute

numeric

0			
0.55	0		
0.45	1	0	
0.40	0.14	0.86	0

$d^{3}_{(1,2)}$=|45- 22|/42 =0.55,

$d^{3}_{(1,3)}$=|45- 64|/42 =0.45,

$d^{3}_{(1,4)}$=|45- 24|/42 =0.40,

We can now use the formula $$d(i,j) = \frac{\sum_{f=1}^{p} \delta_{ij}^{(f)} d_{ij}^{(f)}}{\sum_{f=1}^{p} \delta_{ij}^{(f)}}$$ **to compute dissimilarity matrix using the**

d(3,1) = (1 * 1 + 1 * 0.50 + 1 * 0.45)/ 3 = 0.65

0			
0.85	0		
0.65	0.83	0	
0.13	0.71	0.79	0

Example 2: Given the following dataset

Object identifier **Symmetric** **Asymmetric**

name	gender	fever	cough	Test-1	Test-2	Test-3	Test-4
jack	M (1)	Y (1)	N (0)	P (1)	N (0)	N (0)	N (0)
jim	M (1)	Y (1)	Y (1)	N (0)	N (0)	N (0)	N (0)
mary	F (0)	Y (1)	N (0)	P (1)	N (0)	P (1)	N (0)
….							
…							

d(I,j) = r + s/q+r+s+t for symmetric binary, d(I,j) = r + s/q+r+s for asymmetric binary

		Object j		
		1	0	sum
	1	q	r	q + r
Object I	0	s	t	s +t
	Sum	q+ s	r + t	p

Example #1 q=1 s=1 r=1

jack	**Y (1)**	**N (0)**	**P (1)**	**N (0)**	**N (0)**	**N (0)**
jim	Y (1)	Y (1)	N (0)	N (0)	N (0)	N (0)

Want to compute d(jack,Jim) = 1 + 1/ 1 + 1 + 1 = 0.67

Example # 2 **q=1** **q=1**

jack	Y (1)	N (0)	P (1)	N (0)	N (0)	N (0)
mary	Y (1)	N (0)	P (1)	N (0)	P (1)	N (0)

d(Jack,Mary)= 0+ 1/2+0 + 1= 0.33

Example # 3

d(Jim,Mary) = 1 + 2/ 1 + 1 + 2= 0.75

Jim and Mary are unlikely to have similar disease because they have the highest dissimilarity.

Jack and Mary are the most likely to have a similar disease