

MSBD5002: Data Mining and Knowledge Discovery (MicroMasters)  
Exercise 4 (Suggested Solution)  
Monothetic Approach

**Q1**

Consider the correlation between A and B.

<b>B\A</b>	<b>1</b>	<b>0</b>
<b>1</b>	2	0
<b>0</b>	1	1

$$X_{AB}^2 = 1.33$$

Consider the correlation between A and C.

<b>C\A</b>	<b>1</b>	<b>0</b>
<b>1</b>	1	1
<b>0</b>	2	0

$$X_{AC}^2 = 1.33$$

Consider the correlation between B and C.

<b>C\B</b>	<b>1</b>	<b>0</b>
<b>1</b>	0	2
<b>0</b>	2	0

$$X_{BC}^2 = 4$$

For attribute A,

$$X_{AB}^2 + X_{AC}^2 = 1.33 + 1.33 = 2.66$$

For attribute B,

$$X_{AB}^2 + X_{BC}^2 = 1.33 + 4 = 5.33$$

For attribute C,

$$X_{AC}^2 + X_{BC}^2 = 1.33 + 4 = 5.33$$

We choose attribute B for splitting since it has the largest value.

We divide the data into two groups, namely {1, 2} and {3, 4}.

Dendrogram:

