Supplementary Notes #04

Data Mining and Data Warehousing

Solution to Exercises on Fuzzy Data Mining

Question 1

		Water Melon		
		Small	Large	Total
Milk	Small	1.7	2.1	3.8
	Medium	1.5	1.2	2.7
	Large	1.0	1.5	2.5
	Total	4.2	4.8	9.0

The support of the association "Small Milk $\rightarrow$ Large Mellon" is given by:

$$support = \Pr(\text{Small Milk}, \text{Large Mellon}) = \frac{2.1}{9.0} = 0.23 .$$

Since the support is less than 0.25 (the user-specified threshold), the association is uninteresting.

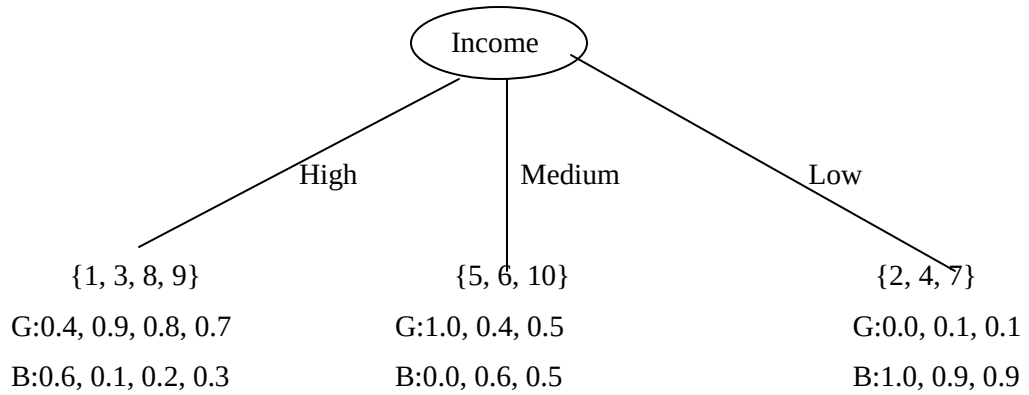
Question 2

$$\Pr(\ Credit\ \ = Good\ \ ) = \frac{0.4 + 0.9 + 0.1 + 1 + 0.4 + 0.1 + 0.8 + 0.7 + 0.5}{10} = 0.49$$

$$\Pr(\ Credit\ \ = Bad\ \ ) = \frac{0.6 + 1 + 0.1 + 0.9 + 0.6 + 0.9 + 0.2 + 0.3 + 0.5}{10} = 0.51$$

$$U_{all}\ (Credit\ \ ) = -0.49\ \log(\ 0.49\ ) - 0.51\ \log(\ 0.51\ ) = 0.999$$

Record No.	Income level	Frequency of Call	Credit Rating
1	High	Frequent	{(Good, 0.4), (Bad, 0.6)}
2	Low	Frequent	Bad
3	High	Not Frequent	{(Good, 0.9), (Bad, 0.1)}
4	Low	Frequent	{(Good, 0.1), (Bad, 0.9)}
5	Medium	Not Frequent	Good
6	Medium	Frequent	{(Good, 0.4), (Bad, 0.6)}
7	Low	Frequent	{(Good, 0.1), (Bad, 0.9)}
8	High	Not Frequent	{(Good, 0.8), (Bad, 0.2)}
9	High	Not Frequent	{(Good, 0.7), (Bad, 0.3)}
10	Medium	Frequent	{(Good, 0.5), (Bad, 0.5)}

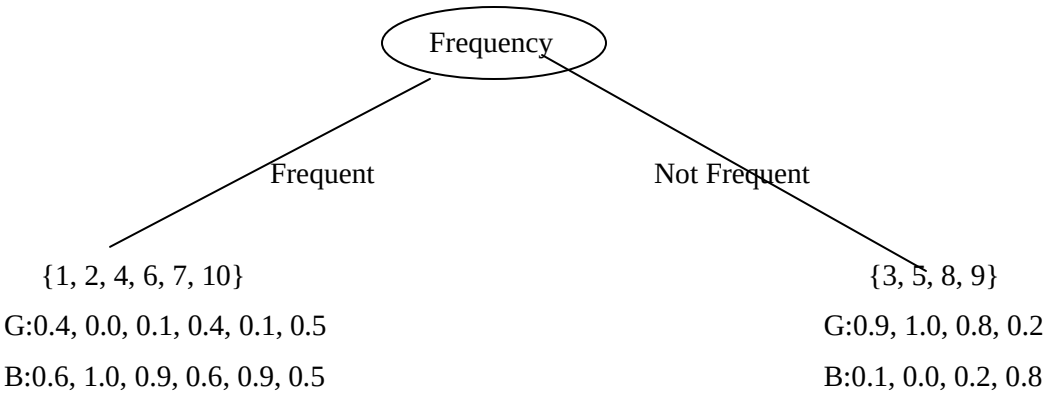


$$U_{incom\ =high}\ (Credit\ \ ) = -(2.8/4.0)\log(\ 2.8/4.0) - (1.2/4.0)\log(\ 1.2/4.0) = 0.36\ + 0.52\ = 0.88$$

$$U_{incom\ =Medium}\ (Credit\ \ ) = -(1.9/3.0)\log(\ 1.9/3.0) - (1.1/3.0)\log(\ 1.1/3.0) = 0.42\ + 0.53\ = 0.95$$

$$U_{incom\ =low}\ (Credit\ \ ) = -(0.2/3.0)\log(\ 0.2/3.0) - (2.8/3.0)\log(\ 2.8/3.0) = 0.26\ + 0.09\ = 0.35$$

$$Average_{\ income}\ (Credit\ \ ) = (4/10)\times 0.88\ + (3/10)\times 0.95\ + (3/10)\times 0.35\ = 0.742$$



$$U_{frequency=f}(Credit) = -(1.5/6.0)\log(1.5/6.0) - (4.5/6.0)\log(4.5/6.0) = 0.811$$

$$U_{frequencey=nf}(Credit) = -(2.9/4.0)\log(2.9/4.0) - (1.1/4.0)\log(1.1/4.0) = 0.85$$

$$Average_{frequency}(Credit) = (6/10)\times 0.811 + (4/10)\times 0.85 = 0.83$$

Therefore, we choose income level to split.