Assignment 3 24133101g WU Yifan

1. $\frac{1}{2} / \frac{1}{2}+\frac{\theta}{4} = \frac{2}{2+\theta}$, $\frac{\theta}{4} / \frac{1}{2}+\frac{\theta}{4} = \frac{\theta}{2+\theta}$

$X_1 \sim (y_1, \frac{2}{2+\theta})$ $X_2 \sim (y_1, \frac{\theta}{2+\theta})$

$P(X,Y|\theta) = P(X,|\theta) = \frac{(\sum x_i)!}{x_1! \cdots x_5!}(\frac{1}{2})^{x_1}(\frac{\theta}{4})^{x_2}(\frac{1-\theta}{4})^{x_3+x_4}(\frac{\theta}{4})^{x_5}$

Since Y is only a function of X

$$P(X|Y,\theta) = \binom{y_1}{x_1}\left(\frac{2}{2+\theta}\right)^{x_1}\left(\frac{\theta}{2+\theta}\right)^{x_2}$$

$$Q(\theta,\theta^{(k)}) = \int \log p(X,Y|\theta)\, p(X|y,\theta^{(k)})\, dx$$

$$= \int \left[\log\left(\frac{(\sum x_i)!}{x_1! \cdots x_5!}\right) + x_1 \log(\tfrac{1}{2}) + x_2 \log(\tfrac{\theta}{4}) + (x_3+x_4)\log(\tfrac{1-\theta}{4}) + x_5 \log(\tfrac{\theta}{4})\right] P(X|y,\theta^{(k)})\, dx$$

$$= \int \left[C + x_2 \log\theta + (x_3+x_4)\log(1-\theta) + x_5 \log\theta\right] P(X|y,\theta^{(k)})\, dx$$

$$= E(C|Y,\theta^{(k)}) + (y_2+y_3)\log(1-\theta) + \left(\frac{\theta^{(k)} y_1}{\theta^{(k)}+2} + y_4\right)\log\theta$$

$$\frac{\partial Q(\theta,\theta^{(k)})}{\partial\theta} = -\frac{y_2+y_3}{1-\theta} + \frac{1}{\theta}\left(\frac{\theta^{(k)} y_1}{\theta^{(k)}+2} + y_4\right) = 0$$

$$M(\theta) = \frac{2y_4 + (y_1+y_4)\theta^{(k)}}{2(y_2+y_3+y_4) + (\sum y_i)\theta^{(k)}} = \theta \quad \text{i.e.}$$

$$\boxed{M(\theta) = \frac{110 + 230\,\theta^{(k)}}{220 + 285\,\theta^{(k)}} = \theta}$$

The Final iteration equation is

$$\boxed{\theta^{(k+1)} = \frac{22 + 46\,\theta^{(k)}}{44 + 57\,\theta^{(k)}}}$$

2. with the definition of single linkage, $dist(C_1, C_2) = \min_{x \in C_1} \min_{y \in C_2} dist(x, y)$

Step 1 $C_1 = \{5, 2\}$

$C_2 = 1$

$C_3 = 3$

$C_4 = 4$

	C_1	C_2	C_3	C_4
C_1	0	2.33	1.30	1.50
C_2	2.33	0	3.15	1.90
C_3	1.30	3.15	0	3.70
C_4	1.50	1.90	3.70	0

at level 0.47

Step 2 $C_5 = \{C_1, C_3\}$

$C_6 = \{C_2\}$

$C_7 = \{C_4\}$

	C_5	C_6	C_7
C_5	0	2.33	1.50
C_6	2.33	0	1.90
C_7	1.50	1.90	0

at level 1.30

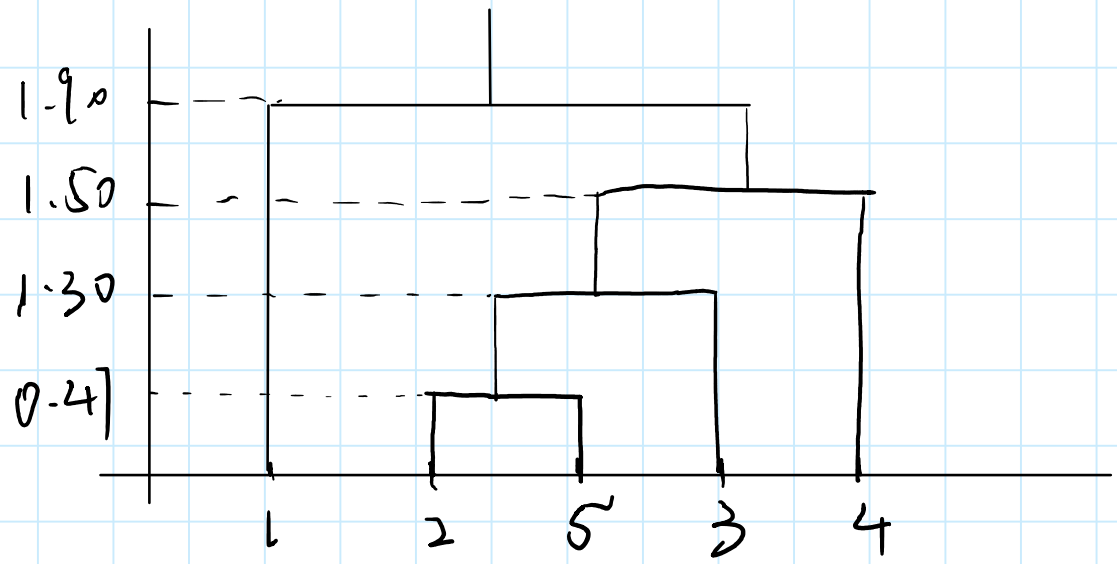
Step 3 $C_8 = \{C_5, C_7\}$

$C_9 = \{C_6\}$

	C_8	C_9
C_8	0	1.90
C_9	1.90	0

at level 1.50

The cluster tree result



3.

(a) support $\{e\} = \frac{8}{10} = \boxed{0.8000}$

support $\{b,d\} = \frac{2}{10} = \boxed{0.2000}$

support $\{b,d,e\} = \frac{2}{10} = \boxed{0.2000}$

(b) confidence $\{b,d\} \rightarrow \{e\} = \frac{\sigma\{b,d,e\}}{\sigma\{b,d\}} = \boxed{\frac{2}{2} = 1}$

confidence $\{e\} \rightarrow \{b,d\} = \frac{\sigma\{b,d,e\}}{\sigma\{e\}} = \boxed{\frac{2}{8} = 0.2500}$

confidence is not a symmetric measure

(c) $\{a,b\} \rightarrow \{e\}$ confidence $= \frac{\sigma\{a,b,e\}}{\sigma\{a,b\}} = \frac{3}{4} = 0.7500 \geq 0.5000$

$\{b,e\} \rightarrow \{a\}$ confidence $= \frac{\sigma\{a,b,e\}}{\sigma\{b,e\}} = \frac{3}{5} = 0.6000 \geq 0.5000$

$\{a,e\} \rightarrow \{b\}$ confidence $= \frac{\sigma\{a,b,e\}}{\sigma\{b\}} = \frac{3}{6} = 0.5000 \geq 0.5000$

$\{a\} \rightarrow \{b,e\}$ confidence $= \frac{\sigma\{a,b,e\}}{\sigma\{a\}} = \frac{3}{7} = 0.4286 < 0.5000$

$\{b\} \rightarrow \{a,e\}$ confidence $= \frac{\sigma\{a,b,e\}}{\sigma\{b\}} = \frac{3}{6} = 0.5000 \geq 0.5000$

$\{e\} \rightarrow \{a,b\}$ confidence $= \frac{\sigma(a,b,e)}{\sigma\{e\}} = \frac{3}{8} = 0.3750 < 0.5000$

All association confidence is not less than 50% as follows

$\boxed{\begin{array}{l}\{b\} \rightarrow \{a,e\} \\ \{a,b\} \rightarrow \{e\} \\ \{b,e\} \rightarrow \{a\} \\ \{a,e\} \rightarrow \{b\}\end{array}}$