Question1

(a)

$$Info(D) = -\sum_i p_i \log p_i$$

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} Info(D_j)$$

$$Gain(A) = Info(D) - Info_A(D)$$

$$Info(D) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$$

$$Info_A(D) = \frac{5}{12} * (-\frac{3}{5}\log\frac{3}{5} - \frac{2}{5}\log\frac{2}{5}) - \frac{3}{12} * (-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}) - \frac{4}{12} * (-\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}) = 0.904$$

$$Gain(A) = 1 - 0.904 = 0.096$$

(b)

$$gini(D) = 1 - \sum_i p_i^2$$

$$gini_A(D) = \frac{|D_1|}{|D|}gini(D_1) + \frac{|D_2|}{|D|}gini(D_2)$$

$$\Delta gini(A) = gini(D) - gini_A(D)$$

$$gini(D) = 1 - (\frac{1}{2})^2 - (\frac{1}{2})^2 = 0.5$$

$$gini(D)_A = \frac{5}{12} * (1 - (\frac{3}{5})^2 - (\frac{2}{5})^2) + \frac{7}{12} * (1 - (\frac{3}{7})^2 - (\frac{4}{7})^2) = 0.485$$

$$\Delta gini(A) = 0.5 - 0.485 = 0.015$$

Question2

(a)
$$P(accept = yes) = \frac{1}{2} \quad , \quad P(accept = no) = \frac{1}{2}$$

(b)
$$P(GPA = 4.0 \,|\, accept = yes) = \frac{3}{6} = \frac{1}{2}$$

$$P(GPA = 3.7 \,|\, accept = yes) = \frac{1}{2}$$

$$P(GPA = 3.5 \,|\, accept = yes) = 0$$

$$P(univ = top10 \,|\, accept = yes) = \frac{1}{2}$$

$$P(univ = top20 \,|\, accept = yes) = \frac{1}{3}$$

$$P(univ = top30 \,|\, accept = yes) = \frac{1}{6}$$

$$P(published = yes \,|\, accept = yes) = \frac{1}{2}$$

$$P(published = no \,|\, accept = yes) = \frac{1}{2}$$

$$P(recommendation = good \,|\, accept = yes) = \frac{5}{6}$$

$$P(recommendation = normal \,|\, accept = yes) = \frac{1}{6}$$

(c)
$$P(GPA = 4.0 \,|\, accept = no) = 0$$

$$P(GPA = 3.7 \,|\, accept = no) = \frac{1}{3}$$

$$P(GPA = 3.5\,|\,accept = no) = \frac{2}{3}$$

$$P(univ = top10\,|\,accept = no) = \frac{1}{3}$$

$$P(univ = top20\,|\,accept = no) = \frac{1}{6}$$

$$P(univ = top30\,|\,accept = no) = \frac{1}{2}$$

$$P(published = yes\,|\,accept = no) = \frac{1}{3}$$

$$P(published = no\,|\,accept = no) = \frac{2}{3}$$

$$P(recommendation = good\,|\,accept = no) = \frac{1}{2}$$

$$P(recommendation = normal\,|\,accept = no) = \frac{1}{2}$$

(d)

define X: (GPA=3.7, univ=top-20, published=yes, recommendation=good)

$$P(accept = yes\,|\,X) = P(X\,|\,accept = yes) * P(accept = yes)$$

$$= \frac{1}{2} * \frac{1}{3} * \frac{1}{2} * \frac{5}{6} * \frac{1}{2} = 0.0347$$

define Y: (GPA=3.7, univ=top-30, published=no, recommendation=normal)

$$P(accept = yes\,|\,X) = P(X\,|\,accept = yes) * P(accept = yes)$$

$$= \frac{1}{2} * \frac{1}{6} * \frac{1}{2} * \frac{1}{6} * \frac{1}{2} = 0.0035$$

Question3

(a)

C1: <a>, <b>, <c>, <d>, <e>, <f>

L1: <a>:3, <b>:3, <c>:3, <d>:3, <e>:3, <f>:3

(b)

C2: Total of 51 Candidates

	<a>	<b>	<c>	<d>	<e>	<f>
<a>	<aa>	<ab>	<ac>	<ad>	<ae>	<af>
<b>	<ba>	<bb>	<bc>	<bd>	<be>	<bf>
<c>	<ca>	<cb>	<cc>	<cd>	<ce>	<cf>
<d>	<da>	<db>	<dc>	<dd>	<de>	<df>
<e>	<ea>	<eb>	<ec>	<ed>	<ee>	<ef>
<f>	<fa>	<fb>	<fc>	<fd>	<fe>	<ff>

	<a>	<b>	<c>	<d>	<e>	<f>
<a>		<(ab)>	<(ac)>	<(ad)>	<(ae)>	<(af)>
<b>			<(bc)>	<(bd)>	<(be)>	<(bf)>
<c>				<(cd)>	<(ce)>	<(cf)>
<d>					<(de)>	<(df)>
<e>						<(ef)>
<f>						

L2: Total of 10 length-2 sequence pattern with same support 3

<ae>, <af>, <bd>, <be>, <bf>, <cd>, <ce>, <cf>, <df>, <ef>

(c)

C3: <aef>, <bdf>, <bef>, <cdf>, <cef>

	<ae>	<af>	<bd>	<be>	<bf>	<cd>	<ce>	<cf>	<df>	<ef>
<ae>										<aef>
<af>										
<bd>									<bdf>	
<be>										<bef>
<bf>										
<cd>									<cdf>	
<ce>										<cef>
<cf>										
<df>										
<ef>										

L3: <aef>:3, <bdf>:3, <bef>:3, <cdf>:3, <cef>:3

(d)

C4: {}

	<aef>	<bdf>	<bef>	<cdf>	<cef>
<aef>					
<bdf>					
<bef>					
<cdf>					
<cef>					

L4: {}

The algorithm terminates.