Question 1

(a)

$$\text{Info(D)} = -\frac{1}{2}log_2\frac{1}{2} - \frac{1}{2}log_2\frac{1}{2} = 1$$

$$\text{Info}_{univ}\text{(D)} = \frac{5}{12}\left(-\frac{3}{5}log_2\frac{3}{5} - \frac{2}{5}log_2\frac{2}{5}\right) + \frac{3}{12}\left(-\frac{2}{3}log_2\frac{2}{3} - \frac{1}{3}log_2\frac{1}{3}\right) + \frac{4}{12}\left(-\frac{1}{4}log_2\frac{1}{4} - \frac{3}{4}log_2\frac{3}{4}\right) = 0.9046$$

$$\text{Gain(univ)} = \text{Info(D)} - \text{Info}_{univ}\text{(D)} = 0.0954$$

(b)

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

$$\text{gini}_{published}\text{(D)} = \frac{5}{12}(1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2) + \frac{7}{12}(1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2) = 0.4857$$

$$\Delta\text{gini(univ)} = \text{gini(D)} - \text{gini}_{published}\text{(D)} = 0.0143$$

Question 2

(a)

P(accepted=yes) =0.5

P(accepted=no) =0.5

(b)

P(GPA=4.0|accepted=yes)=0.5

P(GPA=3.7|accepted=yes)=0.5

P(GPA=3.5|accepted=yes)=0

P(univ=top-10|accepted=yes)=0.5

P(univ=top-20|accepted=yes)=1/3

P(univ=top-30|accepted=yes)=1/6

P(published=yes|accepted=yes)=0.5

P(published=no|accepted=yes)=0.5

P(recommendation=good|accepted=yes)=5/6

P(recommendation=normal |accepted=yes)=1/6

(c)

P(GPA=4.0|accepted=no)=0

P(GPA=3.7|accepted=no)=1/3

P(GPA=3.5|accepted=no)=2/3

P(univ=top-10|accepted=no)=1/3

P(univ=top-20|accepted=no)=1/6

P(univ=top-30|accepted=no)=1/2

P(published=yes|accepted=no)=1/3

P(published=no|accepted=no)=2/3

P(recommendation=good|accepted=no)=1/2

P(recommendation=normal |accepted=no)=1/2

(d)

X1= (GPA=3.7, university=top-20, published=yes, recommendation=good)

P(X1|accepted=yes) =0.5*1/3*0.5*5/6=0.0694

P(X1|accepted=no) =1/3*1/6*1/3*1/2=0.00925

P(X1)= P(X1|accepted=yes)*P(accepted=yes)+ P(X1|accepted=no)*P(accepted=no)

=0.0393

$$\mathrm{P}(\text{accpeted} = \text{yes}|\mathrm{X1}) = \frac{\mathrm{P}(\mathrm{X1}|\text{accepted} = \text{yes}) * \mathrm{P}(\text{accepted} = \text{yes})}{P(X1)}$$
$$= \frac{0.0694 * 0.5}{0.0393} = 0.8829$$

X2= (GPA=3.7, university=top-30, publication=no, recommendation=normal)

P(X2|accepted=yes) =0.5*1/6*0.5*1/6=0.00694

P(X2|accepted=no)=1/3*1/2*2/3*1/2=0.0555

P(X2)= P(X2|accepted=yes)*P(accepted=yes)+ P(X2|accepted=no)*P(accepted=no)

=0.03122

$$\mathrm{P}(\text{accpeted} = \text{yes}|\mathrm{X2}) = \frac{\mathrm{P}(\mathrm{X2}|\text{accepted} = \text{yes}) * \mathrm{P}(\text{accepted} = \text{yes})}{P(X2)}$$
$$= \frac{0.00694 * 0.5}{0.03122} = 0.1111$$

Question 3

(a)

C1:

a:4 b:4 c:3 d:3 e:3 f:3

L1:

a:4 b:4 c:3 d:3 e:3 f:3

(b)

C2:

aa:1 ab:2 ac:2 ad:2 ae:3 af:3

ba:1 bb:1 bc:1 bd:3 be:3 bf:3

ca:2 cb:1 cd:3 ce:3 cf:3

da:1 db:1 de:2 df:3

ef:3

(ab):1 (ad):1 (bc):2 (de):1

L2:

ae:3 af:3 bd:3 be:3 bf:3 cd:3 ce:3 cf:3 df:3 ef:3

(c)

C3:

aef:3 bdf:3 bef:3 cdf:3 cef:3

L3:

aef:3 bdf:3 bef:3 cdf:3 cef:3

Then, the algorithm terminates.