Question 1

(1) FALSE ("ai" occurs 3 times)
(2) TRUE (if both child nodes have the same majority class, then the parent node will also have that same majority class)
(3) FALSE (they are equivalent if they have the same skeleton and *v-structures*)
(4) FALSE (each split)
(5) TRUE
(6) TRUE (candidates are generated by taking a frequent tree, and adding a node with a frequent label using right-most extension)
(7) TRUE
(8) FALSE
(9) FALSE (Counterexample: r = AD, s = ABD, d = ABCD)
(10) TRUE

Question 2

x <= 10, x <= 14, x <= 18

Question 3

g(t) = (8/100-0)/(5-1) = 2/100 = 0.02

Question 4

Most unfavorable case for the root node is if each class has relative frequency 1/3.
Then the resubstitution error is 2/3 and total cost of the root is 2/3 + a.
Most favorable case for Tmax is when 2 splits are sufficient to get pure leaf nodes.
Then total cost of Tmax is 0 + 3a. Total cost is equal when 2/3 + a = 3a, hence for a = 1/3.
Therefore the root node is guaranteed to have the same or lower total cost than Tmax for a >= 0.33.

Question 5

(a) The maximum likelihood fitted counts are given by the formula: n(A,B,D)n(C,D)/n(D).
(b) The number of u-terms is 10.

Question 6

Level 2 generators: BD, CD
Closed frequent item sets: ABC, BC, CD, BCD, D.

Question 7

(a) 1/7
(b) 361/655

Question 8

(a) 1/(1+exp(-0.4)) = 0.5986877, so the answer is: 0.60
(b) -0.4 + 0.05x = 0, ==> x = 8, so 8 minutes.
(c) exp(10*0.05)= 1.648721, so 65 percent.

Question 9

(a) 1721 * log(1721/2000) + 279 * log(279/2000) = -808.1096, so the answer is -808.11
(b) 429 * log(429/495) + 66 * log(66/495) + 1292 * log(1292/1505) + 213 * log(213/1505) = -808.0045, so the answer is -808.00
(c) There are 29 parameters in the model, at log(2000)/2 = 3.800451 each, so the total size of the complexity penalty is 29 * 3.800451 = 110.21

Question 10

remove(A --> D) and remove(C --> D)