Queen Mary
University of London

BSc Examination by course unit

28th May 2013 10:00am–12:30pm

ECS611P Artificial Intelligence Duration: 2 hours 30 minutes

THIS PAPER INCLUDES ANSWERS!!!

YOU ARE NOT PERMITTED TO READ THE CONTENTS OF THIS QUESTION PAPER UNTIL INSTRUCTED TO DO SO BY AN INVIGILATOR.

Answer ALL FOUR Questions

CALCULATORS ARE PERMITTED IN THIS EXAMINATION.

COMPLETE ALL ROUGH WORKINGS IN THE ANSWER BOOK AND CROSS THROUGH ANY WORK THAT IS NOT TO BE ASSESSED.

IMPORTANT NOTE: THE ACADEMIC REGULATIONS STATE THAT POSSESSION OF UNAUTHORISED MATERIAL AT ANY TIME WHEN A STUDENT IS UNDER EXAMINATION CONDITIONS IS AN ASSESSMENT OFFENCE AND CAN LEAD TO EXPULSION FROM QMUL.

PLEASE CHECK NOW TO ENSURE YOU DO NOT HAVE ANY NOTES, MOBILE PHONES OR UNAUTHORISED ELECTRONIC DEVICES ON YOUR PERSON. IF YOU HAVE ANY, THEN PLEASE RAISE YOUR HAND AND GIVE THEM TO AN INVIGILATOR IMMEDIATELY. PLEASE BE AWARE THAT IF YOU ARE FOUND TO HAVE HIDDEN UNAUTHORISED MATERIAL ELSEWHERE, INCLUDING TOILETS AND CLOAKROOMS, IT WILL BE TREATED AS BEING FOUND IN YOUR POSSESSION. UNAUTHORISED MATERIAL FOUND ON YOUR MOBILE PHONE OR OTHER ELECTRONIC DEVICE WILL BE CONSIDERED THE SAME AS BEING IN POSSESSION OF PAPER NOTES. MOBILE PHONES CAUSING A DISRUPTION IS ALSO AN ASSESSMENT OFFENCE.

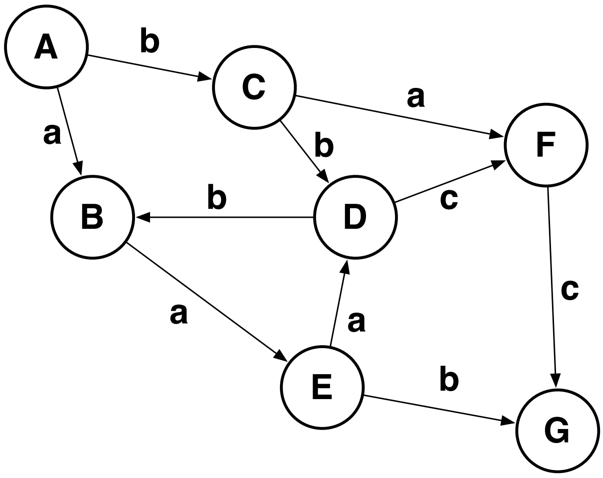
EXAM PAPERS CANNOT BE REMOVED FROM THE EXAM ROOM

Examiners: Prof. Geraint A. Wiggins; Prof. Edmund Robinson

Answer ALL FOUR of the following questions.

1. **Search strategies**

(a) Consider the following state space graph. States are represented as nodes in the graph (labelled A, B, ...). The initial state is A; there is exactly one goal state, G. There are three operators (labelled a, b, c). You can refer to particular arcs by the names of their ends: for example, AB is the arc from A to B. So arc AB represents the application of operator a to state A, resulting in state B.



A search algorithm is needed for this state space. Library functions are provided, as follows.

`expand` takes the label of a state, and returns an ordered list of all the states that are reachable in one step from it, applying the operators in alphabetical order. So `expand`(C) returns [F,D].

`append` takes two ordered lists, and returns a new one formed by joining the two arguments together, the first one going first; the order of the original lists is maintained. So `append`([D,E], [A,B]) returns [D,E,A,B].

`first` takes a non-empty ordered list, and returns its first element. So `first`([A,B,C]) returns A.

`rest` takes a non-empty ordered list, and removes the first element and returns the rest. So `rest`([A,B,C]) returns [B,C].

`nonEmpty` takes an ordered list, and returns `true` if it is non-empty, `false` otherwise.

The very basic algorithm to be used is as follows (in pseudocode); it is missing a crucial part, labelled `?????`.

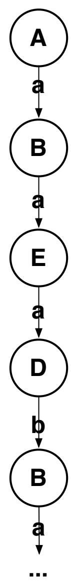
```
Function search( initialState ) returns Boolean
   Ordered-List agenda ← [initialState]
   Ordered-List route ← []

   while nonEmpty( Agenda )
      if first( Agenda ) == G
         return True
      agenda ← ?????
   return False
```

Now answer the following questions, given that `expand` and `append`, as defined above, are to be used in the implementation of the search algorithm.

i. What pseudocode must be inserted at `?????` to implement Depth First Search? **[2 marks]**

   `append( expand( first( agenda )), rest( agenda ))`
   [2]

ii. Draw the search tree for this state space for Depth First Search. At each depth, write the nodes in order of selection, left to right. Include only the nodes expanded in an actual application of the search algorithm as you have completed it. **[5 marks]**

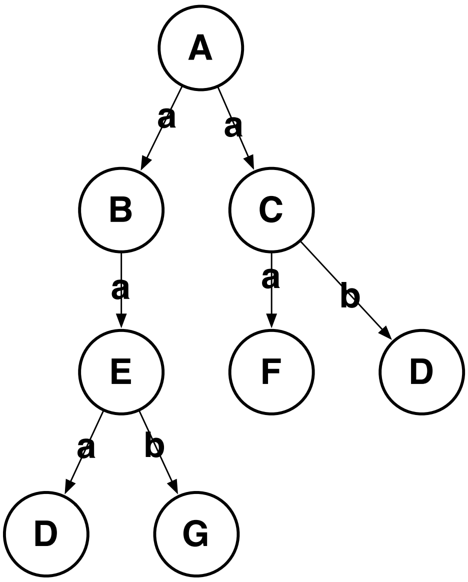


   The tree is infinite, looping back to B. [Deduct half a mark for each error. Do not penalise knock-on errors.]

iii. What pseudocode must be inserted at `?????` to implement Breadth First Search? **[2 marks]**

   `append( rest( agenda ), expand( first( agenda )))`
   [2]

iv. Draw the search tree for this state space for Breadth First Search. At each depth, write the nodes in order of selection, left to right. Include only the nodes expanded in an actual application of the search algorithm as you have completed it. **[5 marks]**



[Deduct 1 mark for each error. Do not penalise knock-on errors.]

(b) Explain what is a *heuristic*, and exactly what it does in search algorithms. **[2 marks]**

A heuristic is a rule of thumb that is used to guide search algorithms [1]. It estimates the cost of traversal of the search space to the nearest solution [1].

(c) Explain the property that makes a heuristic *admissible*. **[1 mark]**

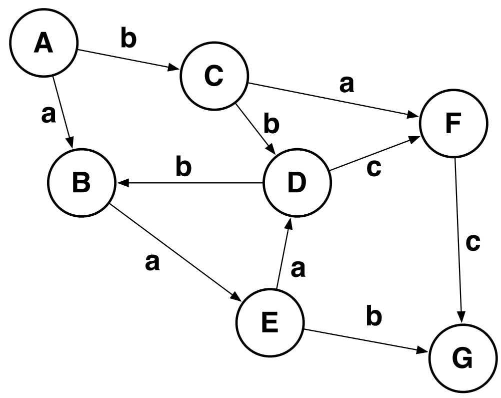
An admissible heuristic is one that never over-estimates the distance between the current node and the nearest goal [1].

(d) Briefly, compare and contrast UCS, Greedy search, Algorithm A and Algorithm A*. In particular, why is Algorithm A* superior to the others? You may assume the definitions you gave in Parts 1b and 1c, above. **[8 marks]**

UCS uses the cost-so-far of a path through the search space and attempts to minimise it [1]. It guarantees to find a solution if there is one, but may search the entire state space to find it [1]. Greedy search attempts to estimate the distance to the nearest solution using a heuristic, and limits the number of operator application at each step to the best, usually just one [1]. It does not guarantee to find a solution, because of the limit on the number of operators considered [1]. Algorithm A uses a combination of known cost and a heuristic estimate of cost to the best solution, so it is a combination of the other two [1]. It is guaranteed to find a solution if there is one, and it may search less of the space than UCS [1]. Given an admissible heuristic, A becomes A* becomes optimal, because an admissible heuristic guarantees that the search will use the fewest possible nodes possible [1], given the problem formulation [1].

2. **UCS Search and Probabilistic Reasoning**

(a) Again, consider the following state space (it's the same as in Question 1). States are represented as nodes in the graph (labelled A, B, ...). The initial state is A; there is exactly one goal state, G. There are three operators (labelled a, b, c). You can refer to particular arcs by the names of their ends: for example, AB is the arc from A to B. So arc AB represents the application of operator a to state A, resulting in state B.



This time, also consider the following costs, which are incurred by each operator, respectively.

a	5
b	3
c	6

Write down the successive values of the agenda in a Uniform-Cost search implemented in the same basic search algorithm, using these costs, starting from A, and ending at G. Write exactly one agenda per line in your answer, indicating the cost of each state in the agenda by writing the relevant number above the state, as shown for the starting state. **[12 marks]**

$$[\overset{0}{A}]$$

$$[\overset{0}{A}]$$
$$[\overset{3}{C}, \overset{5}{B}]$$
$$[\overset{5}{B}, \overset{6}{D}, \overset{8}{F}]$$
$$[\overset{6}{D}, \overset{8}{F}, \overset{10}{E}]$$
$$[\overset{8}{F}, \overset{9}{B}, \overset{10}{E}, \overset{12}{F}]$$
$$[\overset{9}{B}, \overset{10}{E}, \overset{12}{F}, \overset{14}{G}]$$
$$[\overset{10}{E}, \overset{12}{F}, \overset{14}{G}, \overset{14}{E}]$$
$$[\overset{12}{F}, \overset{13}{G}, \overset{14}{G}, \overset{14}{E}, \overset{15}{D}]$$
$$[\overset{13}{G}, \overset{14}{G}, \overset{14}{E}, \overset{15}{D}, \overset{18}{G}]$$

[Deduct half a mark for each error. Do not penalise for knock-on errors.]

(b) Students taking exams fail, pass (without distinction) or pass with distinction. For a given class, students who study have a likelihood of .2 of passing with distinction and a likelihood of .6 of passing without distinction. Students who don't study have a likelihood of .4 of passing, and no chance of passing with a distinction. 50% of the class does not study.

i. Tabulate these likelihoods and the associated marginal probabilities. **[6 marks]**

	Fail	*Pass*	*Distinction*	*Marginal*
Study	*.1*	*.3*	*.1*	*.5*
¬ Study	*.4*	*.1*	*0*	*.5*
Marginal	*.5*	*.4*	*.1*	*1*

[2 marks per line of table.]

ii. Show how Bayes' Rule can be used to compute the likelihood that a student who passed without distinction took time to study. Show your working in full; solutions without working will be penalised. Bayes' Rule is defined as follows:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}.$$

[3 marks]

$$\begin{aligned} P(Study \mid Pass) &= \frac{P(Pass \mid Study)P(Study)}{P(Pass)} \\ &= \frac{.6 \times .5}{.4} \\ &= .75 \end{aligned}$$

[Deduct 1 for each error. Do not penalise knock-on errors. Must use Bayes' rule, not the table.]

iii. What is the probability that not studying implies failure? Show your working. **[4 marks]**

$$\begin{aligned} P(\neg Study \rightarrow Fail) &= P(\neg\neg Study \vee Fail) \\ &= P(Study \vee Fail) \\ &= P(Study) + P(Fail \wedge \neg Study) \\ &= .5 + .4 \\ &= .9 \end{aligned}$$

[Deduct 1 for each error. Do not penalise knock-on errors.]

3. **Logic and Knowledge Representation**

Consider the following situation:

Sam is a man has with blue eyes and blond hair. He is 1.8m tall. He likes dogs. Sally is a woman with green eyes and auburn hair. She is 1.6m tall. She owns a dog. Sally likes everyone who likes her dog.

(a) Write down the ontology required to express this situation in predicate calculus, identifying the three different kinds of syntactic object that can appear in an ontology, and saying which is which. You may assume real numbers and associated tests (i.e., you don't need to write them down). **[5 marks]**

Predicates:	*Constants:*	*Functions:*
Man	*Sam*	*(None)*
EyeColour	*Blue*	
HairColour	*Blond*	
Height	*Sally*	
Likes	*Green*	
Dog	*Auburn*	
Woman		
Owns		

[Half off for each error.]

(b) Translate the situation into predicate calculus, using your ontology. **[8 marks]**

$$Man(Sam) \qquad EyeColour(Sam, Blue) \qquad [1]$$
$$HairColour(Sam, Blond) \qquad Woman(Sally) \qquad [1]$$
$$EyeColour(Sally, Green) \qquad HairColour(Sally, Auburn) \qquad [1]$$
$$Height(Sam, 1.8) \qquad Height(Sally, 1.6) \qquad [1]$$
$$\forall x.Dog(x) \rightarrow Likes(Sam, x) \qquad [1]$$
$$\exists d.Dog(d) \wedge Owns(Sally, d) \qquad [1]$$
$$\forall x.\forall d.Dog(d) \wedge Owns(Sally, d) \wedge Likes(x, d) \rightarrow Likes(Sally, x) \qquad [2]$$

(c) Give a resolution proof, showing and explaining your working, of the claim **[12 marks]**

Sally likes Sam.

In resolution, use the negated goal: $\neg Likes(Sally, Sam)$*.[1]*

Convert non-normal clauses to CNF. First, remove $\rightarrow$ *[1]:*

$$\forall x.\neg Dog(x) \vee Likes(Sam, x)$$
$$\exists d.Dog(d) \wedge Owns(Sally, d)$$
$$\forall x.\forall d.\neg(Dog(d) \wedge Owns(Sally, d) \wedge Likes(x, d)) \vee Likes(Sally, x)$$

Then reduce scope of negations [1]:

$$\forall x.\neg Dog(x) \vee Likes(Sam, x)$$
$$\exists d.Dog(d) \wedge Owns(Sally, d)$$
$$\forall x.\forall d.\neg Dog(d) \vee \neg Owns(Sally, d) \vee \neg Likes(x, d) \vee Likes(Sally, x)$$

Then standardise variables apart [1]:

$$\forall x. \neg Dog(x) \vee Likes(Sam, x)$$
$$\exists d. Dog(d) \wedge Owns(Sally, d)$$
$$\forall y. \forall z. \neg Dog(z) \vee \neg Owns(Sally, z) \vee \neg Likes(y, z) \vee Likes(Sally, y)$$

Then Skolemise existentials and drop universals [1]:

$$\neg Dog(x) \vee Likes(Sam, x)$$
$$Dog(D) \wedge Owns(Sally, D)$$
$$\neg Dog(z) \vee \neg Owns(Sally, z) \vee \neg Likes(y, z) \vee Likes(Sally, y)$$

Split into disjunctive clauses [1]:

$$\neg Dog(x) \vee Likes(Sam, x)$$
$$Dog(D)$$
$$Owns(Sally, D)$$
$$\neg Dog(z) \vee \neg Owns(Sally, z) \vee \neg Likes(y, z) \vee Likes(Sally, y)$$

Resolution proof: derive contradiction from [6. Do not penalise knock-on errors.]:

1. $Man(Sam)$
2. $EyeColour(Sam, Blue)$
3. $HairColour(Sam, Blond)$
4. $Woman(Sally)$
5. $EyeColour(Sally, Green)$
6. $HairColour(Sally, Auburn)$
7. $Height(Sam, 1.8)$
8. $Height(Sally, 1.6)$
9. $\neg Dog(x) \vee Likes(Sam, x)$
10. $Dog(D)$
11. $Owns(Sally, D)$
12. $\neg Likes(Sally, Sam)$
13. $\neg Dog(z) \vee \neg Owns(Sally, z) \vee \neg Likes(y, z) \vee Likes(Sally, y)$

Resolve 12 with last literal in 13, unifying y with Sam.

14. $\neg Dog(z) \vee \neg Owns(Sally, z) \vee \neg Likes(Sam, z)$

Resolve 10 with first literal in 14, unifying z with D.

15. $\neg Owns(Sally, D) \vee \neg Likes(Sam, D)$

Resolve 11 with first literal in 15.

16. $\neg Likes(Sam, D)$

Resolve 9 with 16.

17. $\neg Dog(D)$

Resolve 10 with 17.

18. $\perp$

[Do not penalise for knock-on errors from previous part.]

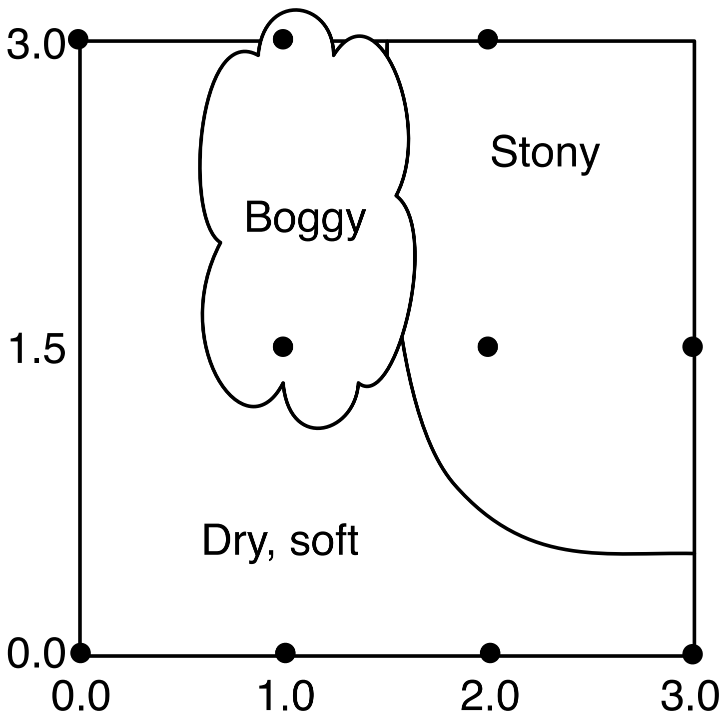
4. **The ID3 algorithm**

(a) Briefly explain the purpose of the ID3 algorithm, and the advantage it confers over simple search. **[4 marks]**

ID3 automatically builds a decision tree [1] which classifies data [1]. It computes the most efficient path through the attributes of the data [1], which are used as successive discriminators between categories [1]. This optimises the search to categorise an item whose attributes are known.

(b) The lawn in my garden has a boggy patch at one end, where previous owners had put an ornamental fish pond, which is now filled in. There's an area of stony ground, too, where there used to be a rockery. I recently purchased an advanced robot lawn mower, which can extend different sets of wheels for dry, earthy ground, boggy ground, and stony ground. It can sense where it is in the garden, but not what kind of ground it's on.

So that my lawn mower can plan its work, I can give it a set of data about the type of earth at various points in the garden, as shown in the diagram and table. It then uses the ID3 algorithm to classify the area and to decide which wheels to use at which point.



x	y	*Ground type*
0.0	0.0	D
0.0	3.0	D
1.0	0.0	D
1.0	1.5	B
1.0	3.0	B
2.0	0.0	D
2.0	1.0	S
2.0	3.0	S
3.0	0.0	D
3.0	1.5	S

Given the formula for entropy,

$$H(S) = -\Sigma_{i=1}^{n} p_i \log_2 p_i$$

where S is a sample of n items and p_i is the frequency of each item in S, write down the process of reasoning through which my new robot goes as it makes its decisions. Give real numbers to 2 decimal places; alternatively, use fractions where appropriate. **[17 marks]**

Start by calculating overall entropy. $p(D) = 0.50, p(B) = 0.20, p(S) = 0.30; H(S) = 1.49$. *[1]*
Now, try each possible division along y *and* x *axes.*

Split	$H(S_{\leq})$	$\mid S_{\leq} \mid$	$H(S_{>})$	$\mid S_{>} \mid$	ΔH
$y = 0.75$	$H(\langle p(D) = 1.00, p(B) = 0.00, p(S) = 0.00\rangle) = 0.00$	4	$H(\langle p(D) = 0.17, p(B) = 0.33, p(S) = 0.50\rangle) = 1.46$	6	$\Delta H = 1.49 - 0.00 \times \frac{4}{10} - 1.46 \times \frac{6}{10} = 0.61$
$y = 1.25$	$H(\langle p(D) = 0.57, p(B) = 0.14, p(S) = 0.29\rangle) = 1.38$	7	$H(\langle p(D) = 0.33, p(B) = 0.33, p(S) = 0.33\rangle) = 1.58$	3	$\Delta H = 1.49 - 1.38 \times \frac{7}{10} - 1.58 \times \frac{3}{10} = 0.05$
$x = 0.50$	$H(\langle p(D) = 1.00, p(B) = 0.00, p(S) = 0.00\rangle) = 0.00$	2	$H(\langle p(D) = 0.38, p(B) = 0.25, p(S) = 0.38\rangle) = 1.56$	8	$\Delta H = 1.49 - 0.00 \times \frac{2}{10} - 1.56 \times \frac{8}{10} = 0.24$
$x = 1.50$	$H(\langle p(D) = 0.60, p(B) = 0.40, p(S) = 0.00\rangle) = 0.97$	5	$H(\langle p(D) = 0.40, p(B) = 0.00, p(S) = 0.60\rangle) = 0.97$	5	$\Delta H = 1.49 - 0.97 \times \frac{5}{10} - 0.97 \times \frac{5}{10} = 0.52$
$x = 2.50$	$H(\langle p(D) = 0.50, p(B) = 0.25, p(S) = 0.25\rangle) = 1.50$	8	$H(\langle p(D) = 0.50, p(B) = 0.00, p(S) = 0.50\rangle) = 1.00$	2	$\Delta H = 1.49 - 1.50 \times \frac{8}{10} - 1.00 \times \frac{2}{10} = 0.09$

[5; 1 per line.] So choose $y = 0.75$ *as first split, because this yields biggest information gain. This gives one set (*$S_{y\leq 0.75}$*) which is fully classified as* D *(*$H = 0$*). [1]*
Now repeat process, with reduced set, $S_{y>0.75}$*. Start by calculating overall entropy.* $p(D) = 0.17, p(B) = 0.33, p(S) = 0.50; H(S) = 1.46$. *[1]*
Next, try each possible division along y *and* x *axes.*

Split	$H(S_{\leq})$	$\mid S_{\leq} \mid$	$H(S_{>})$	$\mid S_{>} \mid$	ΔH
$y = 1.25$	$H(\langle p(D) = 0.33, p(B) = 0.33, p(S) = 0.33\rangle) = 1.58$	3	$H(\langle p(D) = 0.00, p(B) = 0.33, p(S) = 0.67\rangle) = 0.92$	3	$\Delta H = 1.46 - 1.58 \times \frac{3}{6} - 0.92 \times \frac{3}{6} = 0.21$
$x = 0.50$	$H(\langle p(D) = 1.00, p(B) = 0.00, p(S) = 0.00\rangle) = 0.00$	1	$H(\langle p(D) = 0.00, p(B) = 0.40, p(S) = 0.60\rangle) = 0.97$	5	$\Delta H = 1.46 - 0.00 \times \frac{1}{6} - 0.97 \times \frac{5}{6} = 0.65$
$x = 1.50$	$H(\langle p(D) = 0.33, p(B) = 0.67, p(S) = 0.00\rangle) = 0.92$	3	$H(\langle p(D) = 0.00, p(B) = 0.00, p(S) = 1.00\rangle) = 0.00$	3	$\Delta H = 1.46 - 0.92 \times \frac{3}{6} - 0.00 \times \frac{3}{6} = 0.54$
$x = 2.50$	$H(\langle p(D) = 0.20, p(B) = 0.40, p(S) = 0.40\rangle) = 1.52$	5	$H(\langle p(D) = 0.00, p(B) = 0.00, p(S) = 1.00\rangle) = 0.00$	1	$\Delta H = 1.46 - 1.52 \times \frac{5}{6} - 0.00 \times \frac{1}{6} = 0.19$

[4; 1 per line.] This selects $x = 0.5$ *as the next split. This divides* $S_{y>0.75}$ *into two sets. First, consider* $S_{y>0.75,x\leq 0.5}$*; it is fully classified as* D*. [1] So consider* $S_{y>0.75,x>0.5}$*. Initial entropy is* $H(\langle p(D) = 0.00, p(B) = 0.40, p(S) = 0.60\rangle) = 0.97$.

Split	$H(S_{\leq})$	$\mid S_{\leq} \mid$	$H(S_{>})$	$\mid S_{>} \mid$	ΔH
$y = 1.25$	$H(\langle p(D) = 0.00, p(B) = 0.50, p(S) = 0.50\rangle) = 1.00$	2	$H(\langle p(D) = 0.00, p(B) = 0.33, p(S) = 0.67\rangle) = 0.92$	3	$\Delta H = 0.97 - 1.00 \times \frac{2}{5} - 0.92 \times \frac{3}{5} = 0.02$
$x = 1.50$	$H(\langle p(D) = 0.00, p(B) = 1.00, p(S) = 0.00\rangle) = 0.00$	2	$H(\langle p(D) = 0.00, p(B) = 0.00, p(S) = 1.00\rangle) = 0.00$	3	$\Delta H = 0.97 - 0.00 \times \frac{2}{5} - 0.00 \times \frac{3}{5} = 0.97$

[3; 1 per line.] Split on $x \leq 1.50$*. At this point, everything is fully classified. [1]*
[Small penalty only for slips in calculations; most of the credit to be awarded for getting the process right and following it through objectively.]

(c) Write out in pseudocode the structure of `if/then/else` statements that reflects an optimal classification. **[4 marks]**

```
if y ≤ 0.75 then Dry
        else if x ≤ 0.5 then Dry
                    else if x ≤ 1.5 then Boggy
                                else Stony
```

[1 mark per correct line. Other optimal formulations are acceptable.]

End of Paper