```
+=====================================================================+
|              Indian Institute of Technology Kharagpur               |
|           Department of Computer Science and Engineering            |
|---------------------------------------------------------------------|
|                      CS60050 : Machine Learning                     |
|    Spring 2021   |   Short Test 1 (ALL Q&A)   |   Marks : 20        |
| Date : 01-Feb-2021 (Monday)  |  Time : 8:30pm-9:00pm (30 min)       |
+=====================================================================+
```

~~~~~~

Question-1:

Assume that, you have a concept learning problem where the input space is a subset of the two-dimensional (x-y) plane with 0 ≤ x ≤ 9, and 0 ≤ y ≤ 9 (i.e., a 10x10 plane in the all positive quadrant). The hypothesis space consists of axis parallel rectangles that lie completely within the input space and has corners with integer co-ordinate values. Rectangles touching the boundaries of input space are also considered to be within the input space. More precisely, hypotheses are of the form, <(a,b);(c,d)>, indicating (a,b) and (c,d) are the co-ordinates of two diagonal corners of a rectangle (with a ≤ c and b ≤ d) and hence any positively labeled point (x,y) satisfies both the conditions, (a ≤ x ≤ c) and (b ≤ y ≤ d), where a, b, c and d can be integers lying inside [0,9].

We are provided with the following six training instances having the form (x,y,CLASS), where CLASS ∈ {+,-}: { (3,3,+), (5,5,+), (1,2,-), (4,8,-), (7,2,+), (8,5,-) }

(i) What is the cardinality (size) of the total hypothesis space? Only write an integer number as answer. [2 marks]

Ans: 3026

Explanation:

Number of pairs <(a,b);(c,d)> where a ≤ c and b ≤ d
= For every fixation of the point (c,d) = (i,j),
  you may select (a,b) points in [(i+1) x (j+1)] ways
= Σi Σj [(i+1) x (j+1)] (here i,j both go from [0-9] in the summation)
= 55 x 55 = 3025

Therefore, the size of hypothesis space = 3025 + 1 (nothing as hypothesis)

(Alternate) Explanation:

Number of pairs <(a,b);(c,d)> where a ≤ c and b ≤ d
= Number of pairs <(a,b);(c,d)> where a < c and b < d
  (i.e. number of axis-parallel rectangles)
+ Number of pairs <(a,b);(c,d)> where a = c and b < d
  (i.e number of vertical line segments)
+ Number of pairs <(a,b);(c,d)> where a < c and b = d
  (i.e number of horizontal line segments)
+ Number of pairs <(a,b);(c,d)> where a = c and b = d
  (i.e. number of junction points)
+ 1 (nothing as hypothesis)
= [10C2 x 10C2] + [(9+8+...+2+1) x 10] + [(9+8+...+2+1) x 10] + [10 x 10] + 1
= 2025 + 450 + 450 + 100 + 1 = 3026

(ii) What is the most general boundary (G) of the version space created for the above set of training instances using CANDIDATE-ELIMINATION algorithm? Choose the correct option. [2 marks]

Options:

(a) G = { <(1,2);(8,8)> }
(b) G = { <(2,0);(7,7)> }
(c) G = { <(3,2);(7,5)> }
~~~~~~

(d) G = { <(2,2);(7,7)> }

Ans: (b) G = { <(2,0);(7,7)> }
Explanation:
(2,0) and (7,7) are the corner points of the largest rectangle that does not inc
lude -ve labelled points.

(iii) What is the most specific boundary (S) of the version space created for th
e above set of training instances using CANDIDATE-ELIMINATION algorithm? Choose
the correct option. [2 marks]
Options:
(a) S = { <(3,2);(7,5)> }
(b) S = { <(2,2);(7,7)> }
(c) S = { <(3,3);(3,3)> ; <(5,5);(5,5)> ; <(7,2);(7,2)> }
(d) S = { <(1,2);(8,8)> }

Ans: (a) S = { <(3,2);(7,5)> }
Explanation:
(3,2) and (7,5) are the corner points of the smallest rectangle that includes al
l +ve labelled points.

(iv) Suppose the learner can now suggest a new (x,y) instance and ask the traine
r for its classification. Which of the following query the learner would like to
suggest so that it is guaranteed to reduce the size of the version space, regar
dless of how the trainer classifies it. Choose the correct option. [1 mark]
Options:
(a) (4,4)
(b) (7,7)
(c) (3,2)
(d) (2,7)

Ans: (b) (7,7) and (d) (2,7)
Explanation:
For (7,7,+) and (2,7+), the specific boundary expands; For (7,7,-) and (2,7,-),
the general boundary contracts. Hence, the version space reduces in both cases.

(v) Now assume that you are a teacher, attempting to teach a particular target c
oncept <(4,3);(6,7)>. What is the smallest number of training examples you can p
rovide so that CANDIDATE-ELIMINATION algorithm will perfectly learn the target c
oncept (i.e. only one final hypothesis will be learned)? Only write an integer n
umber as answer. [1 mark]

Ans: 4
Explanation:
Minimum four points are required to find any axis-parallel rectangle as hypothes
is. In particular, two + and two - labelled points to determine the outer bounda
ry of the axis-parallel rectangle.

~~~~~~

Question-2:
At the beginning of an examination, you try to predict whether each problem is e
asy or difficult (say, D = +, if it is difficult and D = –, if it is easy). Let
us assume that you use two observable problem attributes:
L = The text length of the problem (say, L = 1, if it is long and L = 0, other
wise)
M = The amount of math in the text (say, M = 1, if a lot of math is there and
M = 0, otherwise)
For training data, assume that you have examined 12 previous problems from the h
omeworks, and have collected the following data:
~~~~~~

L	M	D	#
0	0	−	4
0	0	+	1
0	1	−	0
0	1	+	3
1	0	−	1
1	0	+	2
1	1	−	1
1	1	+	0

In the above table,
The first line says: 4 problems for which L = 0 and M = 0 were not difficult (D = −).
The second line says: 1 problem for which L = 0 and M = 0 was difficult (D = +).
... and so on. Note that, in your training data, you observed no problem for which L = 0 and M = 1, or L = 1 and M = 1.

Now, based on this training data, you want to compute a representation of a difficult problem (D) in the form of a decision tree using the two binary attributes L and M.

(i) Which attribute will you choose first to build a decision tree model using ID3 algorithm with entropy-based information gain measures? Only write the attribute name. [1 mark]

Ans: M
Explanation:
Information Gain (M) = 0.09375 > 0.0 = Information Gain (L)
(calculations are given in next answer)

(ii) What is the information gain of your first chosen attribute in the previous question? Only write the real value as answer. [3 marks]
Helper Data: log(3) = 1.585 and log(5) = 2.323 (all logs are considered in base 2)

Ans: 0.09375
Explanation:

```
Information Gain at Root Node (for attribute M)
    = [- (6/12).log(6/12) - (6/12).log(6/12)]
      - (8/12)[- (3/8).log(3/8) - (5/8).log(5/8)]
      - (4/12).[- (3/4).log(3/4) - (1/4).log(1/4)]
    = 0.09375
Information Gain at Root Node (for attribute L)
    = [- (6/12).log(6/12) - (6/12).log(6/12)]
      - (8/12)[- (4/8).log(4/8) - (4/8).log(4/8)]
      - (4/12).[- (2/4).log(2/4) - (2/4).log(2/4)]
    = 0.0
```

(iii) For classifying the difficult problems, i.e. D = + class, what is the Boolean formula (rule) learnt by the complete decision tree formed with the above training data using ID3 algorithm? For a mixed class leaf-node, majority rule is a dopted for class decision. Choose the correct option. [2 marks]

Options:

(a) (M==0) ∧ (L==1)
(b) (M==1) ∧ (L==0)
(c) ((M==0) ∧ (L==1)) V ((M==1) ∧ (L==0))
(d) ((M==0) ∧ (L==0)) V ((M==1) ∧ (L==1))

Ans: (c) ((M==0) ∧ (L==1)) V ((M==1) ∧ (L==0))

Explanation:

The decision tree is:

```
     M
  0/   \1        So, the decision tree has two paths,
  L     L        (¬M ∧ L) and (M ∧ ¬L),
0/ \1 0/ \1      to reach to the D = + outcome.
-   + +   -
```

~~~~~~

Question-3:

Imagine that you are given the following set of training examples. Each feature (F1, F2, F3) can take on one of three nominal values: a, b, or c.

F1	F2	F3	Category
a	c	a	+
c	a	c	+
a	a	c	-
b	c	a	-
c	c	b	-

Here, in the Category, "+" means "Approve" and "-" means "Reject".

How would a Naive Bayes Classifier system categorize the following three test examples?

Test Example 1:   F1 = a, F2 = c, F3 = b
Test Example 2:   F1 = a, F2 = a, F3 = a
Test Example 3:   F1 = c, F2 = c, F3 = c

Choose the correct option. [3 x 2 = 6 marks]

Options:

(a) + , + , +
(b) - , + , +
(c) - , - , +
(d) - , - , -

Ans: (b) - , + , +

Explanation:

```
  Ex1: P(+|F1=a,F2=c,F3=b) = P(+).P(F1=a,F2=c,F3=b|+)/P(F1=a,F2=c,F3=b)
     = P(+).P(F1=a|+).P(F2=c|+).P(F3=b|+) /
       [P(+).P(F1=a|+).P(F2=c|+).P(F3=b|+) + P(-).P(F1=a|-).P(F2=c|-).P(F3=b|-)]
     = (2/5).(1/2).(1/2).(0) / [(2/5).(1/2).(1/2).(0) + (3/5).(1/3).(2/3).(1/3)]
     = 0
   So, P(-|F1=a,F2=c,F3=b) = 1
  Ex2: P(+|F1=a,F2=a,F3=a) = P(+).P(F1=a,F2=a,F3=a|+)/P(F1=a,F2=a,F3=a)
     = P(+).P(F1=a|+).P(F2=a|+).P(F3=a|+) /
       [P(+).P(F1=a|+).P(F2=a|+).P(F3=a|+) + P(-).P(F1=a|-).P(F2=a|-).P(F3=a|-)]
     = (2/5).(1/2).(1/2).(1/2) / [(2/5).(1/2).(1/2).(1/2) + (3/5).(1/3).(1/3).(1
```
~~~~~~

```
/3)]
     = 9/13 > 1/2
   So, P(-|F1=a,F2=a,F3=a) = 4/13 < 1/2
  Ex2: P(+|F1=c,F2=c,F3=c) = P(+).P(F1=c,F2=c,F3=c|+)/P(F1=c,F2=c,F3=c)
     = P(+).P(F1=c|+).P(F2=c|+).P(F3=c|+) /
       [P(+).P(F1=c|+).P(F2=c|+).P(F3=c|+) + P(-).P(F1=c|-).P(F2=c|-).P(F3=c|-)]
     = (2/5).(1/2).(1/2).(1/2) / [(2/5).(1/2).(1/2).(1/2) + (3/5).(1/3).(2/3).(1
/3)]
     = 9/17 > 1/2
   So, P(-|F1=c,F2=c,F3=c) = 8/17 < 1/2

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```