Review Test Submission: Decision Tree

User	Fei Shen
Course	CS-584-Parent.17S
Test	Decision Tree
Started	2/7/17 7:18 PM
Submitted	2/7/17 8:46 PM
Due Date	2/7/17 11:59 PM
Status	Completed
Attempt Score	140 out of 140 points
Time Elapsed	1 hour, 28 minutes out of 2 hours
Results Displayed	All Answers, Submitted Answers, Correct Answers

Question 1

10 out of 10 points

We have a dataset D, which we will use in 12-fold cross validation. What is the fraction of D that overlaps between training sets of two distinct folds. Please enter your answer as a decimal number.

Selected Answer: 0.8333

Correct Answer: 0.83333 ± 0.001

Question 2

40 out of 40 points

Given the confusion matrix below, calculate the following metrics. Note that, A is positive and B is negative.

		Predicted A	Predicted B
True	A	60	10
	B	20	30

Note: If you find a repeating decimal, round it to 3 digits after the decimal point. (E.g. 0.5555555 -> 0.556)

Accuracy = **[a]**

Precision = **[b]**

Recall = **[c]**

F1-Score = **[d]**

Specified Answer for: a 0.75

Specified Answer for: b 0.75

Specified Answer for: c 0.857

Specified Answer for: d 0.8

Correct Answers for: a		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.75(0)?	
Correct Answers for: b		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.75(0)?	
Correct Answers for: c		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.(86|857|8571|85714|857143|8571429|85714286|857142857)	
Correct Answers for: d		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.8(0|00)?	

Question 3

50 out of 50 points

Given the data set below, please calculate following values:

X	Y
R	1
G	1
B	1
R	0
R	0
G	1
B	1
R	1
B	0
B	0
G	1

Please round each value you calculate to three digits after the decimal point, e.g. enter 0.987654321 as 0.988.

Entropy of Y, i.e. H(Y): **[a]**

Weighted average entropy of Y given X, i.e. H(Y|X): **[b]**

Information Gain of X on Y, i.e. IG(Y, X): **[c]**

Gini impurity of Y, i.e. G(Y): **[d]**

Gini impurity of Y after splitting with X (i.e. weighted average of Gini index of Y given X) : **[e]**

Specified Answer for: a 0.946

Specified Answer for: b 0.727

Specified Answer for: c 0.219

Specified Answer for: d 0.231

Specified Answer for: e 0.182

Correct Answers for: a		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.946(0)?	
Correct Answers for: b		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.727(0)?	
Correct Answers for: c		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.21(8|9)(0)?	
Correct Answers for: d		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.231(0)?|0?.463(0)?	
Correct Answers for: e		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.182(0|00)? |0?.364(0|00)?	

Question 4

30 out of 30 points

We have the list of transactions as given below. Each transaction shows items bought.

1. milk, soda, tea, bread, yogurt, eggs, lemon
2. milk, soda, tea
3. soda, tea
4. soda, tea, bread, yogurt, lemon
5. milk, tea, eggs, lemon
6. tea, bread, yogurt
7. tea, yogurt, lemon
8. soda, yogurt, eggs
9. milk, soda, bread, yogurt, eggs
10. soda, tea, yogurt
11. soda, tea, bread, lemon
12. soda, tea, bread, yogurt, eggs
13. soda, tea, bread, yogurt
14. milk, lemon
15. bread, yogurt, lemon
16. milk, bread, yogurt

Based on this transaction list, we would like to compute association metrics for soda and tea being X, and eggs being Y. Please calculate each of the metrics below.

Support({soda, tea} => {eggs}): **[a]**

Confidence({soda, tea} => {eggs}): **[b]**

Lift({soda, tea} => {eggs}): **[c]**

Specified Answer for: a 0.125

Specified Answer for: b 0.25

Specified Answer for: c 0.8

Correct Answers for: a		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.125(0|00|000)?	
Correct Answers for: b		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.25(0|00|000)?	
Correct Answers for: c		
Evaluation Method	**Correct Answer**	**Case Sensitivity**
Pattern Match	0?.8(0|00|000)?	

Question 5

10 out of 10 points

X is a numerical variable, Y is a binary class variable. Given the data below, find the best split point using information gain:

	X	Y
0	80	1
1	90	1
2	100	0
3	110	1
4	120	0
5	130	0
6	140	1
7	150	0
8	160	0

Selected Answer: 95

Correct Answer: 95

Answer range +/- 0 (95.0 - 95.0)

Friday, April 28, 2017 5:40:28 PM CDT

← OK