Feedback — Homework 3

Help Center

You submitted this quiz on **Sun 3 Jan 2016 2:41 PM PST**. You got a score of **300.00** out of **400.00**. However, you will not get credit for it, since it was submitted past the deadline.

Question 1

Decision Tree

Impurity functions play an important role in decision tree branching. For binary classification problems, let μ_+ be the fraction of positive examples in a data subset, and $\mu_- = 1 - \mu_+$ be the fraction of negative examples in the data subset.

The Gini index is $1 - \mu_+^2 - \mu_-^2$. What is the maximum value of the Gini index among all $\mu_+ \in [0, 1]$?

Your Answer		Score	Explanation
◯ 1			
◯ 0.25			
◯ 0			
◉ 0.5	✔	20.00	
◯ 0.75			
Total		20.00 / 20.00	

Question 2

Following Question 1, there are four possible impurity functions below. We can normalize each impurity function by dividing it with its maximum value among all $\mu_+ \in [0, 1]$. For instance, the classification error is simply $\min(\mu_+, \mu_-)$ and its maximum value is 0.5. So the normalized classification error is $2\min(\mu_+, \mu_-)$. After normalization, which of the following impurity function is equivalent to the normalized Gini index?

Your Answer	Score	Explanation
◯ the entropy, which is $-\mu_+ \ln \mu_+ - \mu_- \ln \mu_-$, with		

$0 \log 0 \equiv 0$.

(selected) the squared regression error (used for branching in classification data sets), which is by definition $\mu_+(1-(\mu_+ - \mu_-))^2 + \mu_-(-1-(\mu_+ - \mu_-))^2$.	✔	20.00
the closeness, which is $1 - \lvert\mu_+ - \mu_-\rvert$.		
none of the other choices		
the classification error $\min(\mu_+, \mu_-)$.		
Total		20.00 / 20.00

Question 3

Random Forest

If bootstrapping is used to sample $N' = pN$ examples out of N examples and N is very large. Approximately how many of the N examples will not be sampled at all?

Your Answer		Score	Explanation
$(1 - e^{-1/p}) \cdot N$			
$e^{-1/p} \cdot N$			
$e^{-1} \cdot N$			
(selected) $e^{-p} \cdot N$	✔	20.00	
$(1 - e^{-p}) \cdot N$			
Total		20.00 / 20.00	

Question 4

Consider a Random Forest G that consists of three binary classification trees $\{g_k\}_{k=1}^{3}$, where each tree is of test 0/1 error $E_{\text{out}}(g_1) = 0.1$, $E_{\text{out}}(g_2) = 0.2$, $E_{\text{out}}(g_3) = 0.3$. Which of the following is the exact possible range of $E_{\text{out}}(G)$?

Your Answer	Score	Explanation

◉ $0 \leq E_{\text{out}}(G) \leq 0.3$ ✔	20.00	
○ $0.1 \leq E_{\text{out}}(G) \leq 0.3$		
○ $0.2 \leq E_{\text{out}}(G) \leq 0.3$		
○ $0.1 \leq E_{\text{out}}(G) \leq 0.6$		
○ $0 \leq E_{\text{out}}(G) \leq 0.1$		
Total	20.00 / 20.00	

Question 5

Consider a Random Forest G that consists of K binary classification trees $\{g_k\}_{k=1}^{K}$, where K is an odd integer. Each g_k is of test 0/1 error $E_{\text{out}}(g_k) = e_k$. Which of the following is an upper bound of $E_{\text{out}}(G)$?

Your Answer	Score	Explanation
◉ $\frac{2}{K+1} \sum_{k=1}^{K} e_k$ ✔	20.00	
○ $\frac{1}{K+1} \sum_{k=1}^{K} e_k$		
○ $\max_{1 \leq k \leq K} e_k$		
○ $\frac{1}{K} \sum_{k=1}^{K} e_k$		
○ $\min_{1 \leq k \leq K} e_k$		
Total	20.00 / 20.00	

Question 6

Gradient Boosting

Let ϵ_t be the weighted 0/1 error of each g_t as described in the AdaBoost algorithm (Lecture 208), and $U_t = \sum_{n=1}^{N} u_n^{(t)}$ be the total example weight during AdaBoost. Which of the following equation expresses U_{T+1} by ϵ_t?

Your Answer	Score	Explanation
○ $\prod_{t=1}^{T} (2\sqrt{\epsilon_t(1-\epsilon_t)})$		

$\sum_{t=1}^{T} \epsilon_t$		
$\sum_{t=1}^{T} (2\sqrt{\epsilon_t(1-\epsilon_t)})$		
(selected) $\prod_{t=1}^{T} \epsilon_t$	✗	-5.00
none of the other choices		
Total		-5.00 / 20.00

Question 7

For the gradient boosted decision tree, if a tree with only one constant node is returned as g_1, and if $g_1(\mathbf{x}) = 2$, then after the first iteration, all s_n is updated from 0 to a new constant $\alpha_1 g_1(\mathbf{x}_n)$. What is s_n?

Your Answer		Score	Explanation
none of the other choices			
$\min_{1 \le n \le N} y_n$			
2			
$\max_{1 \le n \le N} y_n$			
(selected) $\frac{1}{N} \sum_{n=1}^{N} y_n$	✔	20.00	
Total		20.00 / 20.00	

Question 8

For the gradient boosted decision tree, after updating all s_n in iteration t using the steepest η as α_t, what is the value of $\sum_{n=1}^{N} s_n g_t(\mathbf{x}_n)$?

Your Answer		Score	Explanation
0			
$\sum_{n=1}^{N} y_n s_n$			
(selected) $\sum_{n=1}^{N} y_n g_t(\mathbf{x}_n)$	✔	20.00	

- $\sum_{n=1}^{N} y_n^2$
- none of the other choices

Total	20.00 / 20.00

Question 9

Neural Network

Consider Neural Network with $\text{sign}(s)$ instead of $\tanh(s)$ as the transformation functions. That is, consider Multi-Layer Perceptrons. In addition, we will take $+1$ to mean logic TRUE, and -1 to mean logic FALSE. Assume that all x_i below are either $+1$ or -1. Which of the following perceptron

$$g_A(\mathbf{x}) = \text{sign}\left(\sum_{i=0}^{d} w_i x_i\right).$$

implements

$$\text{OR}(x_1, x_2, \ldots, x_d).$$

Your Answer		Score	Explanation
$(w_0, w_1, w_2, \cdots, w_d) = (-d+1, -1, -1, \cdots, -1)$			
$(w_0, w_1, w_2, \cdots, w_d) = (-d+1, +1, +1, \cdots, +1)$			
$(w_0, w_1, w_2, \cdots, w_d) = (d-1, +1, +1, \cdots, +1)$	✔	20.00	
none of the other choices			
$(w_0, w_1, w_2, \cdots, w_d) = (d-1, -1, -1, \cdots, -1)$			
Total		20.00 / 20.00	

Question 10

Continuing from Question 9, among the following choices of D, which D is the smallest for some 5-D-1 Neural Network to implement $\text{XOR}\left(x_1, x_2, x_3, x_4, x_5\right)$?

Your Answer	Score	Explanation

○ 9		
○ 7		
○ 5		
◉ 3	✖	-5.00
○ 1		
Total		-5.00 / 20.00

Question 11

For a Neural Network with at least one hidden layer and $\tanh(s)$ as the transformation functions on all neurons (including the output neuron), what is true about the gradient components (with respect to the weights) when all the initial weights $w_{ij}^{(\ell)}$ are set to 0?

Your Answer		Score	Explanation
○ none of the other choices			
○ only the gradient components with respect to $w_{01}^{(L)}$ may be non-zero, all other gradient components must be zero			
○ only the gradient components with respect to $w_{0j}^{(\ell)}$ for $j > 0$ may non-zero, all other gradient components must be zero			
○ all the gradient components are zero			
◉ only the gradient components with respect to $w_{j1}^{(L)}$ for $j > 0$ may be non-zero, all other gradient components must be zero	✖	-5.00	
Total		-5.00 / 20.00	

Question 12

For a Neural Network with one hidden layer and $\tanh(s)$ as the transformation functions on all neurons (including the output neuron), what is always true about the backprop algorithm when all the initial weights $w_{ij}^{(\ell)}$ are set to 1?

Your Answer		Score	Explanation
none of the other choices			
$w_{ij}^{(1)} = w_{(i+1)j}^{(1)}$ for $1 \leq i < d^{(0)} - 1$ and all j			
all $w_{j1}^{(2)}$ for $j > 0$ are different			
the gradient components with respect to all $w_{ij}^{(\ell)}$ are zero			
$w_{ij}^{(1)} = w_{i(j+1)}^{(1)}$ for all i and $1 \leq j < d^{(1)} - 1$	✔	20.00	
Total		20.00 / 20.00	

Question 13

Experiments with Decision Tree

Implement the simple C&RT algorithm without pruning using the Gini index as the impurity measure as introduced in the class. For the decision stump used in branching, if you are branching with feature i and direction s, please sort all the $x_{n,i}$ values to form (at most) $N + 1$ segments of equivalent θ, and then pick θ within the median of the segment. Run the algorithm on the following set for training:

hw3_train.dat

and the following set for testing:

hw3_test.dat

How many internal nodes (branching functions) are there in the resulting tree G?

Your Answer		Score	Explanation
12			
10			
6			
8	✖	-5.00	
14			
Total		-5.00 / 20.00	

Question 14

Continuing from Question 13, which of the following is closest to the E_{in} (evaluated with 0/1 error) of the tree?

Your Answer		Score	Explanation
0.3			
0.1			
0.0	✔	20.00	
0.4			
0.2			
Total		20.00 / 20.00	

Question 15

Continuing from Question 13, which of the following is closest to the E_{out} (evaluated with 0/1 error) of the tree?

Your Answer		Score	Explanation
0.05			
0.25			
0.35			
0.15	✔	20.00	
0.00			
Total		20.00 / 20.00	

Question 16

Now implement the Bagging algorithm with $N' = N$ and couple it with your decision tree above to make a preliminary random forest G_{RS}. Produce $T = 300$ trees with bagging. Repeat the

experiment for 100 times and compute average E_{in} and E_{out} using the 0/1 error.
Which of the following is true about the average $E_{\text{in}}(g_t)$ for all the 30000 trees that you have generated?

Your Answer		Score	Explanation
$0.09 \le$ average $E_{\text{in}}(g_t) < 0.12$			
$0.03 \le$ average $E_{\text{in}}(g_t) < 0.06$	✔	20.00	
$0.12 \le$ average $E_{\text{in}}(g_t) < 0.50$			
$0.00 \le$ average $E_{\text{in}}(g_t) < 0.03$			
$0.06 \le$ average $E_{\text{in}}(g_t) < 0.09$			
Total		20.00 / 20.00	

Question 17

Continuing from Question 16, which of the following is true about the average $E_{\text{in}}(G_{RF})$?

Your Answer		Score	Explanation
$0.12 \le$ average $E_{\text{in}}(G_{RF}) < 0.50$			
$0.00 \le$ average $E_{\text{in}}(G_{RF}) < 0.03$	✔	20.00	
$0.03 \le$ average $E_{\text{in}}(G_{RF}) < 0.06$			
$0.06 \le$ average $E_{\text{in}}(G_{RF}) < 0.09$			
$0.09 \le$ average $E_{\text{in}}(G_{RF}) < 0.12$			
Total		20.00 / 20.00	

Question 18

Continuing from Question 16, which of the following is true about the average $E_{\text{out}}(G_{RF})$?

Your Answer	Score	Explanation
$0.03 \le$ average $E_{\text{out}}(G_{RF}) < 0.06$		

◉ $0.06 \le$ average $E_{\text{out}}(G_{RF}) < 0.09$	✔	20.00
○ $0.00 \le$ average $E_{\text{out}}(G_{RF}) < 0.03$		
○ $0.09 \le$ average $E_{\text{out}}(G_{RF}) < 0.12$		
○ $0.12 \le$ average $E_{\text{out}}(G_{RF}) < 0.50$		
Total		20.00 / 20.00

Question 19

Now, `prune' your decision tree algorithm by restricting it to have one branch only. That is, the tree is simply a decision stump determined by Gini index. Make a random `forest' G_{RS} with those decision stumps with Bagging like Questions 16-18 with $T = 300$. Repeat the experiment for 100 times and compute average E_{in} and E_{out} using the 0/1 error.
Which of the following is true about the average $E_{\text{in}}(G_{RS})$?

Your Answer		Score	Explanation
◉ $0.09 \le$ average $E_{\text{in}}(G_{RS}) < 0.12$	✔	20.00	
○ $0.06 \le$ average $E_{\text{in}}(G_{RS}) < 0.09$			
○ $0.03 \le$ average $E_{\text{in}}(G_{RS}) < 0.06$			
○ $0.12 \le$ average $E_{\text{in}}(G_{RS}) < 0.50$			
○ $0.00 \le$ average $E_{\text{in}}(G_{RS}) < 0.03$			
Total		20.00 / 20.00	

Question 20

Continuing from Question 19, which of the following is true about the average $E_{\text{out}}(G_{RS})$?

Your Answer	Score	Explanation
○ $0.06 \le$ average $E_{\text{out}}(G_{RS}) < 0.09$		
○ $0.09 \le$ average $E_{\text{out}}(G_{RS}) < 0.12$		
○ $0.03 \le$ average $E_{\text{out}}(G_{RS}) < 0.06$		
○ $0.00 \le$ average $E_{\text{out}}(G_{RS}) < 0.03$		

$0.12 \leq \text{average } E_{\text{out}}(G_{RS}) < 0.50$	✔	20.00
Total		20.00 / 20.00