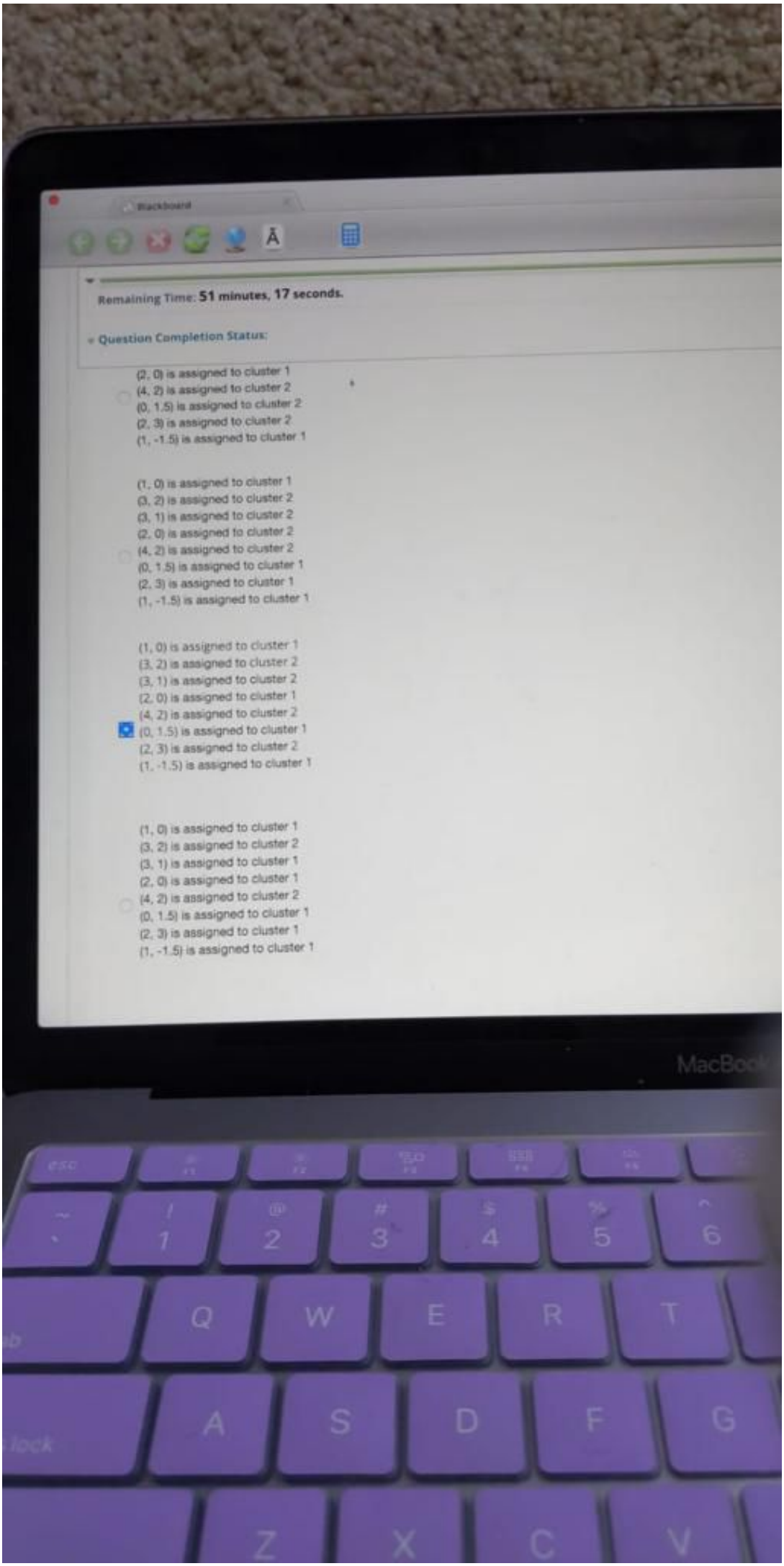

Blackboard
Remaining Time: 51 minutes, 17 seconds.
Question Completion Status:
(2, 0) is assigned to cluster 1
(4, 2) is assigned to cluster 2
(0, 1.5) is assigned to cluster 2
(2, 3) is assigned to cluster 2
(1, -1.5) is assigned to cluster 1
(1, 0) is assigned to cluster 1
(3, 2) is assigned to cluster 2
(3, 1) is assigned to cluster 2
(2, 0) is assigned to cluster 2
(4, 2) is assigned to cluster 2
(0, 1.5) is assigned to cluster 1
(2, 3) is assigned to cluster 1
(1, -1.5) is assigned to cluster 1
(1, 0) is assigned to cluster 1
(3, 2) is assigned to cluster 2
(3, 1) is assigned to cluster 2
(2, 0) is assigned to cluster 1
(4, 2) is assigned to cluster 2
(0, 1.5) is assigned to cluster 1
(2, 3) is assigned to cluster 2
(1, -1.5) is assigned to cluster 1
(1, 0) is assigned to cluster 1
(3, 2) is assigned to cluster 2
(3, 1) is assigned to cluster 1
(2, 0) is assigned to cluster 1
(4, 2) is assigned to cluster 2
(0, 1.5) is assigned to cluster 1
(2, 3) is assigned to cluster 1
(1, -1.5) is assigned to cluster 1
MacBook

Blackboard

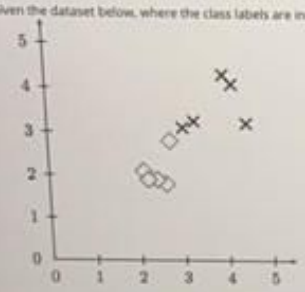
Remaining Time: 1 hour, 49 minutes, 27 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 2

Given the dataset below, where the class labels are indicated by the two different shapes, indicate which of the following classifiers would be able to obtain zero training error on this dataset.



- 3-Nearest Neighbor
- 9-Nearest Neighbor
- SVM with quadratic kernel
- Logistic Regression
- Perceptron

Moving to the next question prevents changes to this answer.

MacBook Pro

esc tab caps lock shift fn control option command

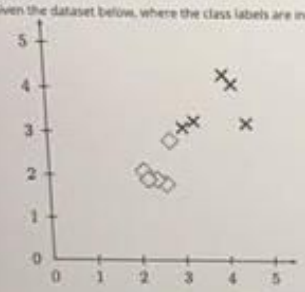
Remaining Time: 1 hour, 49 minutes, 27 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 2

Given the dataset below, where the class labels are indicated by the two different shapes, indicate which of the following classifiers would be able to obtain zero training error on this dataset.



- 1-Nearest Neighbor
- 9-Nearest Neighbor
- SVM with quadratic kernel
- Logistic Regression
- Perceptron

Moving to the next question prevents changes to this answer.

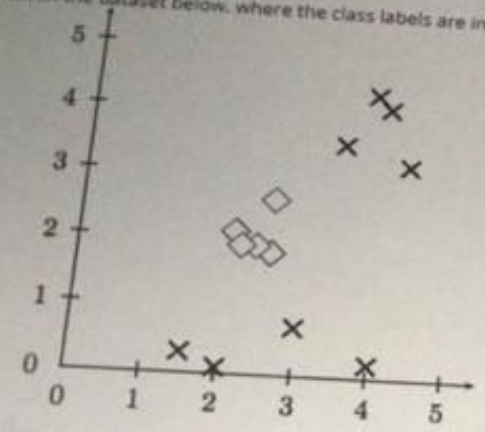
36 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 5

Given the dataset below, where the class labels are indicated by the two different shapes, indicate which of the following classifiers would be able to obtain zero training error on this dataset.



- [ ] Logistic Regression
- [ ] SVM with quadratic kernel
- [ ] Adaboost with decision tree stumps
- [x] 3-Nearest Neighbor
- [ ] Decision tree of maximum depth 2

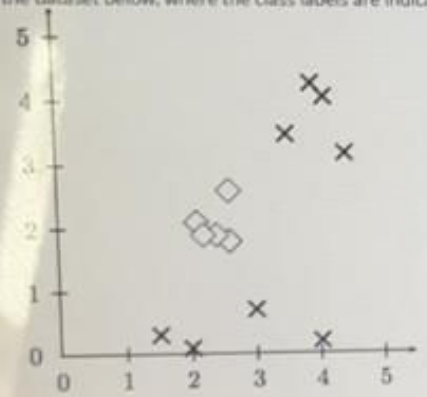
Moving to the next question prevents changes to this answer.

Blackboard

Remaining Time: 45 minutes, 55 seconds.

Question Completion Status:

Given the dataset below, where the class labels are indicated by the two different shapes. Indicate which of the following classifiers would be able to obtain zero training error on this dataset.



- [x] 3-Nearest Neighbor
- [x] SVM with quadratic kernel
- [ ] Adaboost with decision tree stumps
- [x] Decision tree of maximum depth 2
- [ ] Logistic Regression

Moving to the next question prevents changes to this answer.

Question 17 of 30

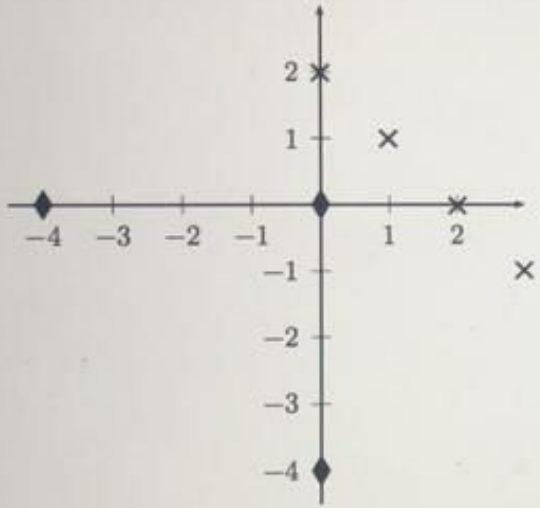
Remaining Time: 1 hour, 40 minutes, 53 seconds.

Question Completion Status:

Question 4

3 points Save Answer

Consider the 2-dimensional dataset shown below. It has two classes represented by the two different shapes in the plot. You would like to train a linear SVM on this dataset.
a. What will be the training set error expressed as a percentage?
b. If you perform 7-fold cross-validation, what will be the total error expressed as a percentage with 2 decimal points?



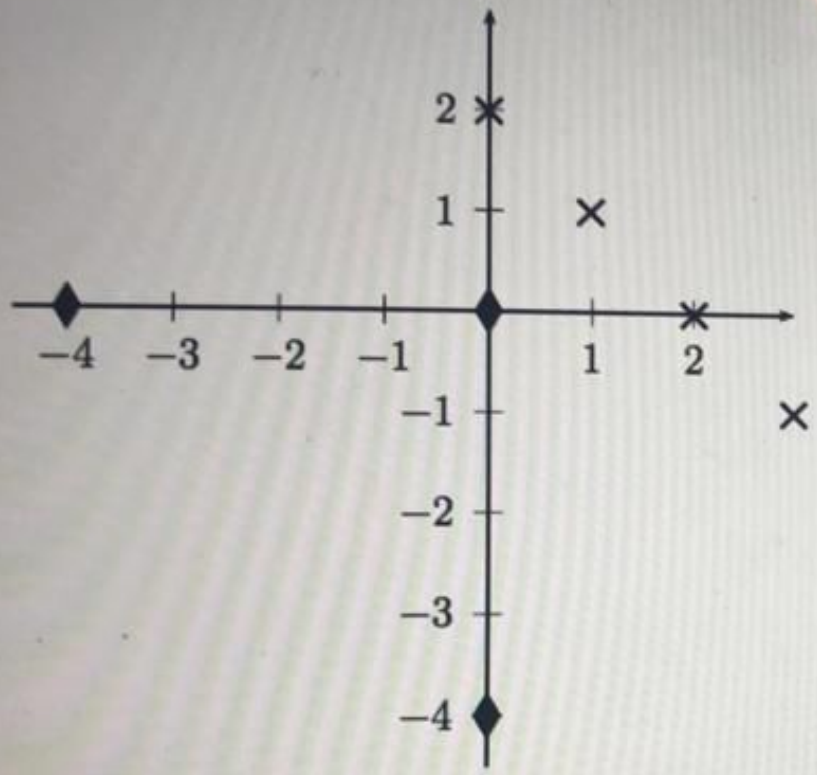
a. 0%
b. 14.28%

a. 14.28%
b. 14.28%

a. 14.28%
b. 28.57%

a. 0%
b. 0%

...nted by the two different shapes in the plot. You would like to train a linear SVM on this d



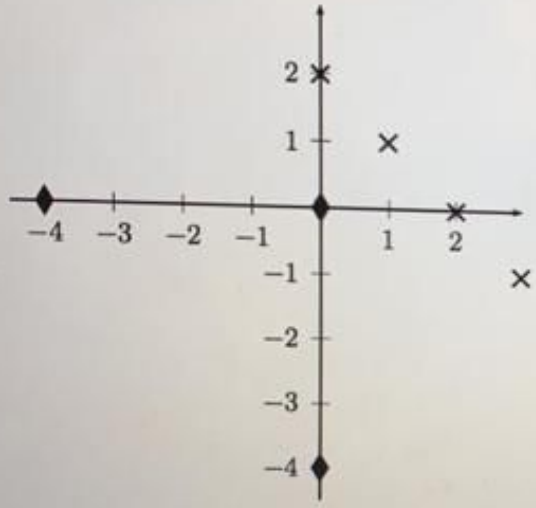
Which of the following could be the equation for the linear SVM dividing surface?
* You can assume that the two axes are x1 and x2 *

- x1+x2 +1 = 0
- x1+x2 -1 = 0
- $x1 + x2 - \sqrt{3} = 0$

Remaining Time: 1 hour, 03 minutes, 09 seconds.

Question Completion Status:

Consider the 2-dimensional dataset shown below. It has two classes represented by the two different shapes in the plot. You would like to train a linear SVM on this dataset.
a. What will be the training set error expressed as a percentage?
b. If you perform 7-fold cross-validation, what will be the total error expressed as a percentage with 2 decimal points?



a. 14.28%
b. 28.57%

a. 0%
b. 0%

a. 0%
b. 14.28%

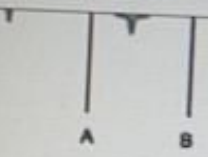
a. 14.28%
b. 14.28%

MacBook Pro

Blackboard

Remaining Time: 1 hour, 45 minutes, 35 seconds.

Question Completion Status:



After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.286$
$\alpha_1 = 0.457$

After the first iteration, there will be one misclassified point and 6 correctly classified points.

After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.143$
$\alpha_1 = 0.895$

After the first iteration, the normalized weights would be as follows:
Each misclassified point would have a weight of 0.501
Each correctly classified point would have a weight of 0.083

In the second iteration, if we choose stump C as the hypothesis, we would get an error of 0.083

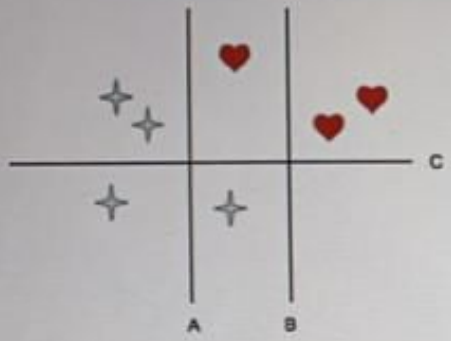
In the second iteration, if we choose stump B as the hypothesis, we would get an error of 0.083

In the second iteration, we would choose stump B as the hypothesis as it would lead to a lower error.

After the first iteration, the normalized weights would be as follows:
Each misclassified point would have a weight of 0.350
Each correctly classified point would have a weight of 0.083

PREMIUM SOUND

Question Completion Status:



After the first iteration, there will be one misclassified point and 6 correctly classified points.

In the second iteration, if we choose stump B as the hypothesis, we would get an error of 0.083

In the second iteration, we would choose stump B as the hypothesis as it would lead to a lower error.

In the second iteration, if we choose stump C as the hypothesis, we would get an error of 0.083

After the first iteration, the normalized weights would be as follows:
Each misclassified point would have a weight of 0.501
Each correctly classified point would have a weight of 0.083

After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.286$
$\alpha_1 = 0.457$

After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.143$
$\alpha_1 = 0.895$

After the first iteration, the normalized weights would be as follows:
Each misclassified point would have a weight of 0.350
Each correctly classified point would have a weight of 0.083

Blackboard

Remaining Time: **1 hour, 21 minutes, 05 seconds.**

Question Completion Status:

- [x] After the first iteration, there will be one misclassified point and 6 correctly classified points.
- [x] In the second iteration, if we choose stump B as the hypothesis, we would get an error of 0.083
- [x] In the second iteration, we would choose stump B as the hypothesis as it would lead to a lower error.
- [ ] In the second iteration, if we choose stump C as the hypothesis, we would get an error of 0.083
- [x] After the first iteration, the normalized weights would be as follows:
  Each misclassified point would have a weight of 0.501
  Each correctly classified point would have a weight of 0.083
- [ ] After the first iteration, the following values will be obtained:
  $\epsilon_1 = 0.286$
  $\alpha_1 = 0.457$
- [x] After the first iteration, the following values will be obtained:
  $\epsilon_1 = 0.143$
  $\alpha_1 = 0.895$
- [ ] After the first iteration, the normalized weights would be as follows:
  Each misclassified point would have a weight of 0.350
  Each correctly classified point would have a weight of 0.083

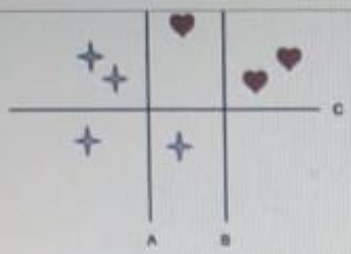
deserve

MacBook Pro

Blackboard

Remaining Time: 20 minutes, 02 seconds.

Question Completion Status:



After the first iteration, there will be one misclassified point and 6 correctly classified points.

After the first iteration, the normalized weights would be as follows:
Each misclassified point would have a weight of 0.350
Each correctly classified point would have a weight of 0.083

In the second iteration, we would choose stump B as the hypothesis as it would lead to a lower error.

After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.143$
$\alpha_1 = 0.895$

In the second iteration, if we choose stump B as the hypothesis, we would get an error of 0.083.

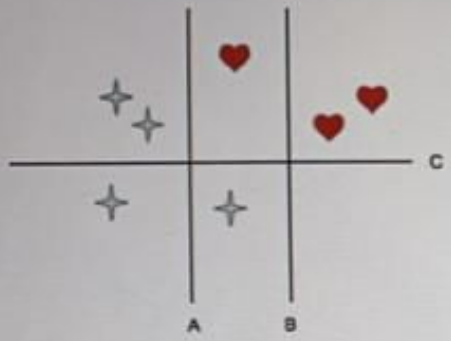
In the second iteration, if we choose stump C as the hypothesis, we would get an error of 0.883.

After the first iteration, the normalized weights would be as follows:
Each misclassified point would have a weight of 0.501.
Each correctly classified point would have a weight of 0.083.

After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.286$

DELL

Question Completion Status:



After the first iteration, there will be one misclassified point and 6 correctly classified points.

In the second iteration, if we choose stump B as the hypothesis, we would get an error of 0.083

In the second iteration, we would choose stump B as the hypothesis as it would lead to a lower error.

In the second iteration, if we choose stump C as the hypothesis, we would get an error of 0.083

After the first iteration, the normalized weights would be as follows:
Each misclassified point would have a weight of 0.501
Each correctly classified point would have a weight of 0.083

After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.286$
$\alpha_1 = 0.457$

After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.143$
$\alpha_1 = 0.895$

After the first iteration, the normalized weights would be as follows:
Each misclassified point would have a weight of 0.350
Each correctly classified point would have a weight of 0.083

Remaining Time: 1 hour, 45 minutes, 43 seconds.

Question Completion Status:

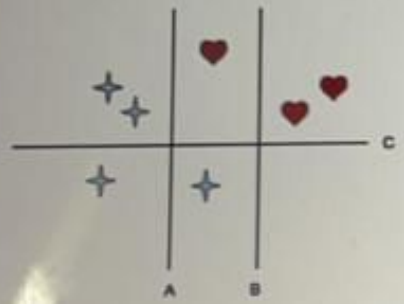
Question 3 of 30

Moving to the next question prevents changes to this answer.

Question 3

4 points Save Answer

You would like to train the AdaBoost ensemble classification algorithm to the dataset below, where the classes are represented by the two different shapes. For each iteration of AdaBoost, you will use a weak classifier in the form of a decision tree stump.
Further, you have a choice of only 3 stumps - A, B, or C - as shown below. In the first iteration of the algorithm, you use stump A as the weak classifier. Which of the following statements are true?



After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.286$
$\alpha_1 = 0.457$

After the first iteration, there will be one misclassified point and 6 correctly classified points.

After the first iteration, the following values will be obtained:
$\epsilon_1 = 0.143$
$\alpha_1 = 0.895$

After the first iteration, the normalized weights would be as follows:

Blackboard

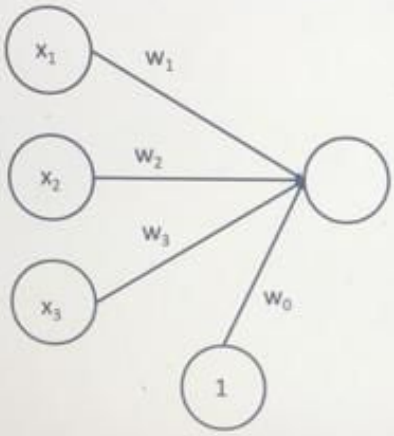
Remaining Time: 49 minutes, 53 seconds.

Question Completion Status:

Consider the following dataset consisting of 3 Boolean attributes - X1, X2, and X3, and a class output (Y) which could be +1 or -1

X1	X2	X3	Y
0	0	0	+1
0	0	1	-1
0	1	0	+1
0	1	1	+1
1	0	0	+1
1	0	1	-1
1	1	0	+1
1	1	1	+1

We would like to represent this dataset using the following neural network:



$Y = \text{sign}(w_0 + w_1x_1 + w_2x_2 + w_3x_3)$

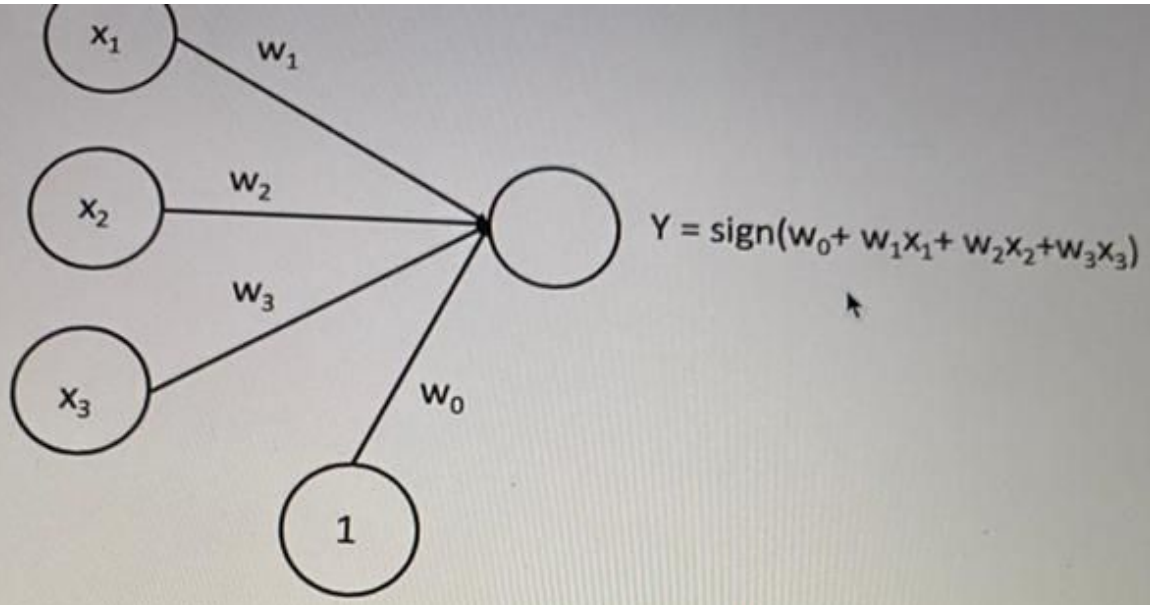
Which of the following could be a valid choice for the set of weights: $[w_0, w_1, w_2, w_3]$?

[1.0, 1.0, -1.0, -1.5]

Live
56:42

our assessment.

issues? See our F

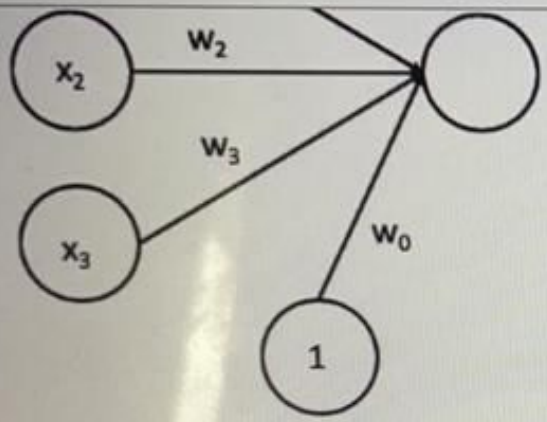


Which of the following could be a valid choice for the set of weights: $[w_0, w_1, w_2, w_3]$?

- ○ [1.0, 0.0, -1.0, -1.5]
- ○ [1.0, 0.0, 1.0, -1.5]
- ○ [1.0, 1.0, -1.0, -1.5]
- ○ [-1.0, 0.0, -2.0, 1.5]

Remaining Time: 09 minutes, 24 seconds.

Question Completion Status:



$$Y = \text{sign}(w_0 + w_1x_1 + w_2x_2 + w_3x_3)$$

Which of the following could be a valid choice for the set of weights: $[w_0, w_1, w_2, w_3]$?

- [1.0, 0.0, -1.0, -1.5]
- [-1.0, 0.0, -2.0, 1.5]
- [1.0, 0.0, 1.0, -1.5]
- [1.0, 1.0, -1.0, -1.5]

⚠ Moving to the next question prevents changes to this answer.

Questio

Close

Remaining Time: 1 hour, 56 minutes, 01 second.

Question Completion Status:

⚠ Moving to the next question prevents changes to this answer.

Question 1

Which of the following statements are true?

- [x] Bagging is a technique that can reduce variance without increasing bias significantly.
- [x] As we run more iterations of the backpropagation algorithm on a neural network, we are increasing the variance.
- [ ] If there is a model that always outputs a constant value, it would have 0 bias and high variance.
- [x] If there is a model that always outputs a constant value, it would have high bias and 0 variance.
- [x] If we post-prune a fully grown decision tree, we would reduce variance
- [x] Consider two models:
  M1 uses a polynomial of degree 9 to fit the training data
  M2 uses a polynomial of degree 2 to fit the training data

  M1 would have a lower bias and greater variance than M2

⚠ Moving to the next question prevents changes to this answer.

our, 57 minutes, 16 seconds.

pletion Status:

on 1

Suppose you have n points x1, x2, ...,xn that are independently drawn from the following probability distribution, where θ is a parameter.

$$Pr(x) = \frac{\theta^x e^{-5\theta}}{x!}$$

What would be the maximum likelihood estimate for θ

$$\theta_{MLE} = \frac{\sum_{i=1}^{n} x_i}{5n}$$

$$\theta_{MLE} = \frac{\prod_{i=1}^{n} x_i}{5n}$$

$$\theta_{MLE} = \frac{\prod_{i=1}^{n} x_i}{n}$$

$$\theta_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Remaining Time: **37 minutes, 08 seconds.**

Question Completion Status:

Suppose you have n points x1, x2, ...,xn that are independently drawn from the following probability distribution, where θ is a parameter.

$$Pr(x) = \frac{\theta^x e^{-5\theta}}{x!}$$

What would be the maximum likelihood estimate for θ

$$\theta_{MLE} = \frac{\sum_{i=1}^{n} x_i}{5n}$$

$$\theta_{MLE} = \frac{\prod_{i=1}^{n} x_i}{n}$$

$$\theta_{MLE} = \frac{\prod_{i=1}^{n} x_i}{5n}$$

$$\theta_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$$

▾ **Question Completion Status:**

Close Window

⚠ Moving to the next question prevents changes to this answer.

Question 2 of 30

Question 2

5 points Save Answer

You are given following 8 data points.
(1, 0), (3,2), (3,1), (2,0), (4, 2), (0, 1.5), (2, 3), (1, -1.5)
You decide to run k-means algorithm on them using k=2 and intial centroids as the following two points (1,0) (called cluster1) and (3,2) (called cluster 2). You will be using simple Manhattan distance for calculations.

How many total iterations will the k-means algorithm take before converging?
Also, let's say we define the total error of clustering as:

$$\sum_i d(x_i, \mu_j)$$

i.e. the distance of each point (x_i) to the center of the cluster (μ_j) to which it is assigned to. You will use Manhattan distance for the calculations. What will be the final value of the total error?

Definition of convergence: If the clustering doesn't change after the n^{th} iteration, then number of iterations required is equal to n

- Number of iterations = 2
  Total Error = 9.0
- Number of iterations = 1
  Total Error = 7.0
- Number of iterations = 1
  Total Error = 9.0
- Not enough information is available to compute these values.

Remaining Time: 1 hour, 50 minutes, 29 seconds.

Question Completion Status:

Close Window

Moving to the next question prevents changes to this answer.

Question 3 of 30

Question 3

5 points Save Answer

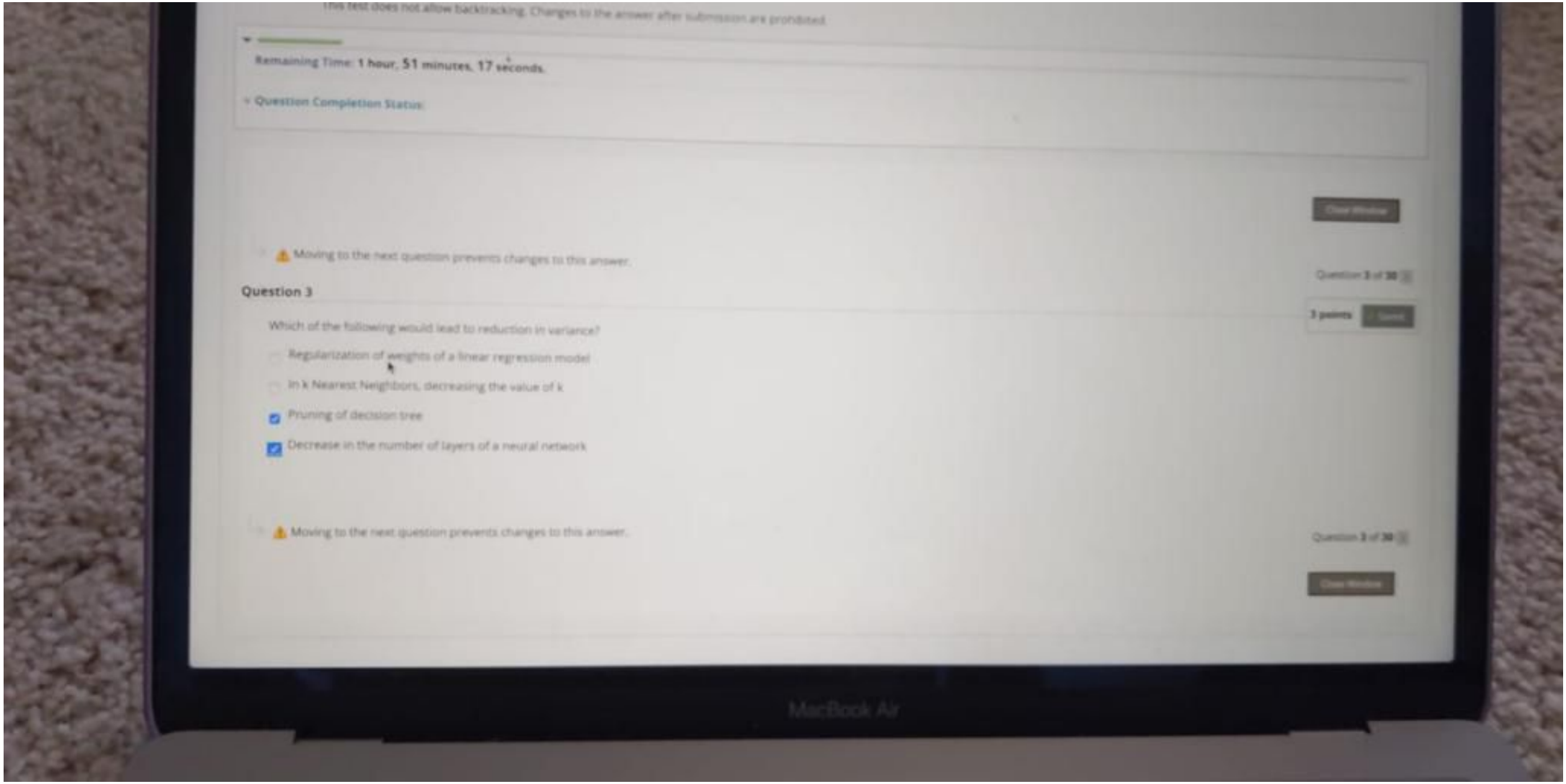
Suppose X is a random variable from uniform distribution between 0 and 1 i.e. Uniform(0, 1) and Y is a random variable from uniform distribution between 0 and 2 i.e. Uniform(0, 2). What would be the value of variance of the random variable Z defined as: Z = X + Y

- 1/3
- 1/4
- 1/12
- 5/12

Moving to the next question prevents changes to this answer.

Question 3 of 30

Close Window


This test does not allow backtracking. Changes to the answer after submission are prohibited.
Remaining Time: 1 hour, 51 minutes, 17 seconds.
Question Completion Status:
Moving to the next question prevents changes to this answer.
Question 3
Which of the following would lead to reduction in variance?
Regularization of weights of a linear regression model
In k Nearest Neighbors, decreasing the value of k
Pruning of decision tree
Decrease in the number of layers of a neural network
Moving to the next question prevents changes to this answer.
MacBook Air

You are given a training dataset below where each instance consists of 3 features - F1, F2, and F3. Each feature can take one of three values - {a, b, c}. Each instance a

F1	F2	F3	Class Label
c	c	c	+
a	a	b	+
a	c	c	-
b	a	a	-
c	c	b	-

You have to train a 3-Nearest Neighbor (3-NN) algorithm on this dataset and find the predicted class label of the following test data point:
X(F1 = a, F2 = c, F3 = b

For calculating distance of two data points, you will use the Hamming distance, which is defined as the fraction of the features that two data points differ on.

$$d_{Hamming}(x1, x2) = \frac{Number\ of\ features\ where\ x1\ and\ x2\ differ}{Total\ number\ of\ features}$$

Which of the following statements are true?

- [ ] The distance of the test point to the second training point is 2/3
- [x] The distance of the test point to the first training point is 2/3
- [x] The distance of the test point to the third training point is 1/3
- [ ] The predicted class label of the test data point would be class +
- [x] The predicted class label of the test data point would be class -
- [x] The distance of the test point to the last training point is 1/3

Moving to the next question prevents changes to this answer.

This test does not allow backtracking. Changes to the answer after submission are prohibited.

Show Timer

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 4 of 30

Question 4

5 points

Professor Ann spends 60% of her time in her office. The rest of her time is spent elsewhere. When Ann is in her office, half the time her light is off (when she is trying to hide from students and get research done). When she is not in her office, she leaves her light on only 5% of the time. 80% of the time she is in her office, Ann is logged onto the computer. Because she sometimes logs onto the computer from home, 10% of the time she is not in her office, she is still logged onto the computer.
Suppose a student checks Ann's login status and sees that she is logged on. What effect does this have on the student's belief that her light is on?

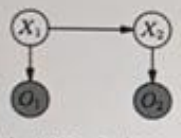
- 0.5
- 0.32
- 0.465
- 0.92

Question 4 of 30

Moving to the next question prevents changes to this answer.

Remaining Time: 1 hour, 42 minutes, 35 seconds.

Question Completion Status:



In the HMM, the hidden variable X can be in any of two states 0 or 1. The observed variable O can have two possible values A or B.
The emission probabilities are presented in the table below.

X	O	P(O|X)
0	A	0.9
0	B	0.1
1	A	0.5
1	B	0.5

The state transition probabilities are shown below:

From(below) / To (right)	0	1
0	0.4	0.6
1	0.8	0.2

The initial state probabilities are shown below:

X_1	$P(X_1)$
0	0.3
1	0.7

Calculate the values of $P(O_1 = A, O_2 = B)$ using the forward algorithm

$P(O_1 = A, O_2 = B) = 0.1548$

$P(O_1 = A, O_2 = B) = 0.0918$

$P(O_1 = A, O_2 = B) = 0.0388$

$P(O_1 = A, O_2 = B) = 0.611$

MacBook Pro

Blackboard

Remaining Time: 42 minutes, 44 seconds.

Question Completion Status:

In the HMM, the hidden variable X can be in any of two states 0 or 1. The observed variabl
The emission probabilities are presented in the table below.

X	O	P(O|X)
0	A	0.9
0	B	0.1
1	A	0.5
1	B	0.5

The state transition probabilities are shown below:

From(below) / To (right)	0	1
0	0.4	0.6
1	0.8	0.2

The initial state probabilities are shown below:

X_1	$P(X_1)$
0	0.3
1	0.7

Calculate the values of $P(O_1 = A, O_2 = B)$ using the forward algorithm

- $P(O_1 = A, O_2 = B) = 0.1548$
- $P(O_1 = A, O_2 = B) = 0.0918$
- $P(O_1 = A, O_2 = B) = 0.0388$
- $P(O_1 = A, O_2 = B) = 0.611$

Moving to the next question prevents changes to this answer.

Blackboard

Remaining Time: **14 minutes, 14 seconds.**

Question Completion Status:

In the HMM, the hidden variable X can be in any of two states 0 or 1. The observed variable O can have two possible values A or B. The emission probabilities are presented in the table below.

X	O	P(O|X)
0	A	0.9
0	B	0.1
1	A	0.5
1	B	0.5

The state transition probabilities are shown below:

From(below) / To (right)	0	1
0	0.4	0.6
1	0.8	0.2

The initial state probabilities are shown below:

X_1	$P(X_1)$
0	0.3
1	0.7

Calculate the values of $P(O_1 = A, O_2 = B)$ using the forward algorithm

$P(O_1 = A, O_2 = B) = 0.1548$

$P(O_1 = A, O_2 = B) = 0.0918$

$P(O_1 = A, O_2 = B) = 0.0388$

$P(O_1 = A, O_2 = B) = 0.611$

Blackboard

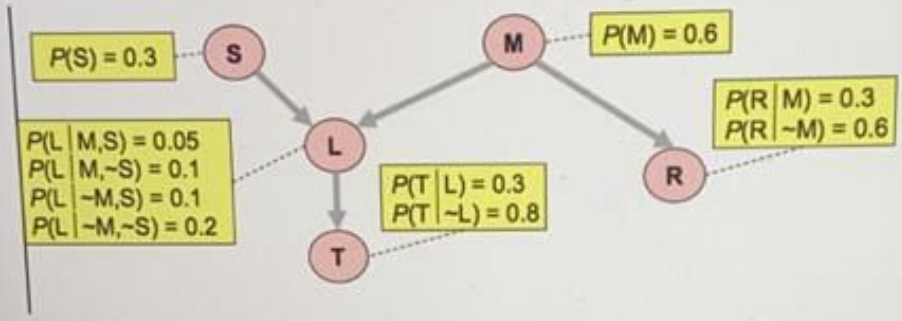
Remaining Time: 1 hour, 44 minutes, 29 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 4

Given the following Bayes net:



Note: ~ symbol means NOT.

What is the joint probability of the following: S, ~M, L, ~R, T

- 0.0014
- 0.00336
- 0.00216
- 0.00504

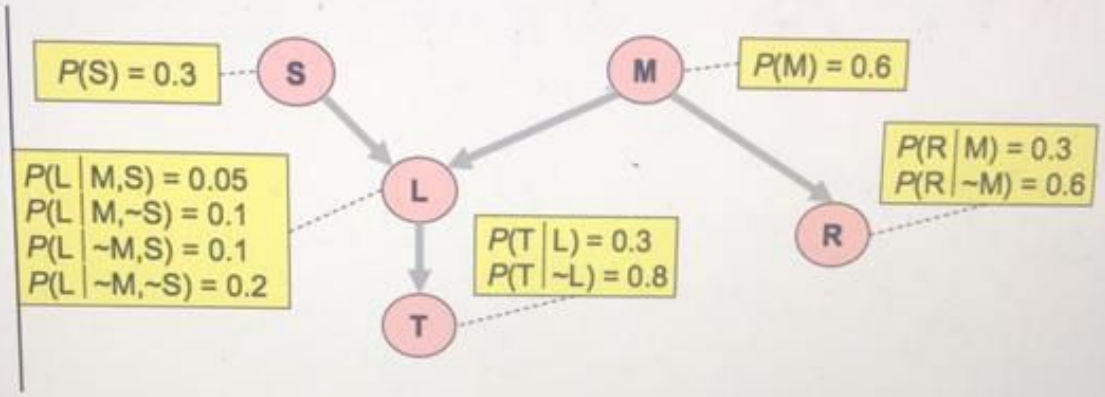
Moving to the next question prevents changes to this answer.

Blackboard

Remaining Time: 03 minutes, 12 seconds.

Question Completion Status:

Given the following Bayes net:



Note: ~ symbol means NOT.

What is the joint probability of the following: S, ~M, L, ~R, T

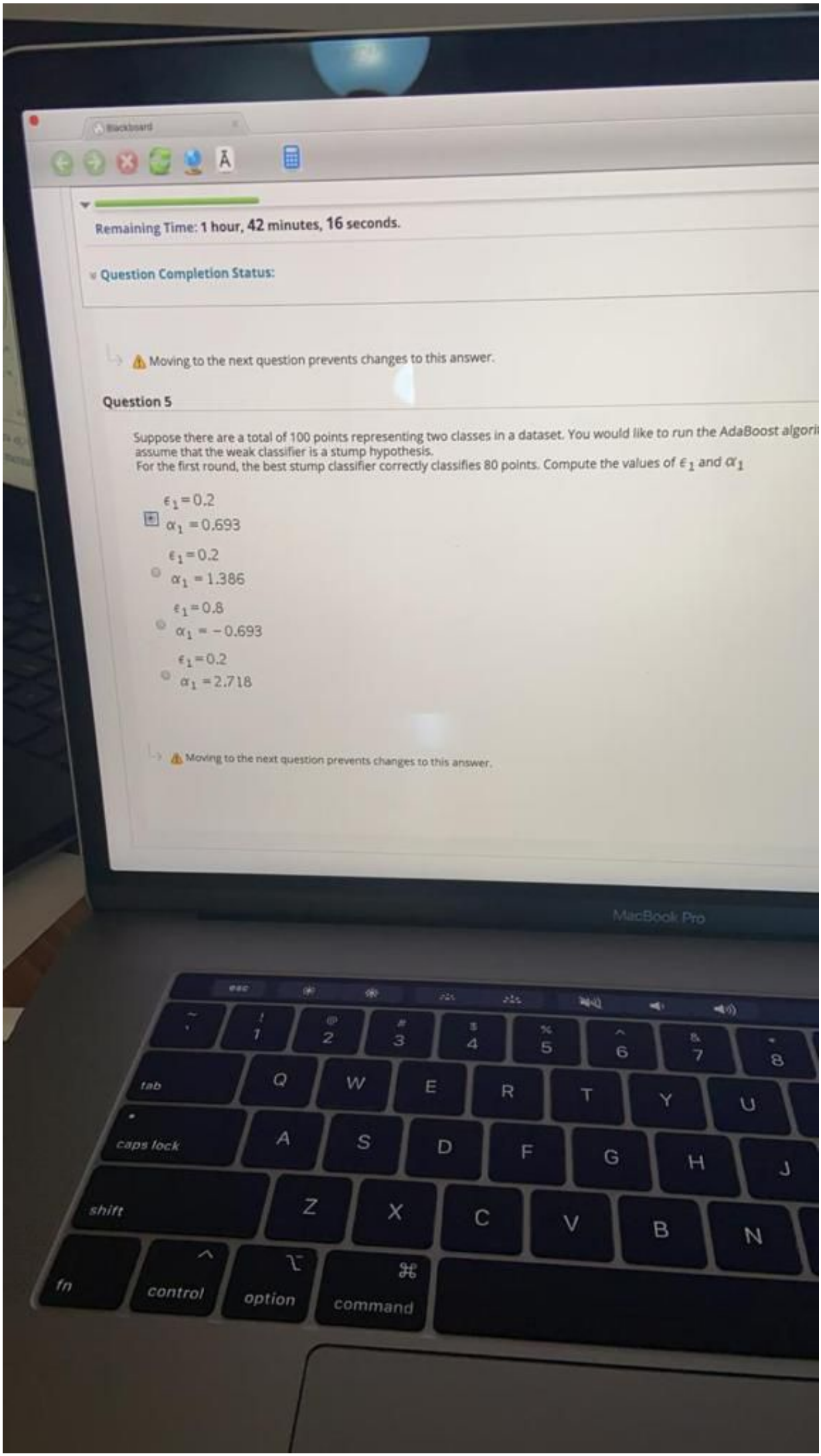
0.00504

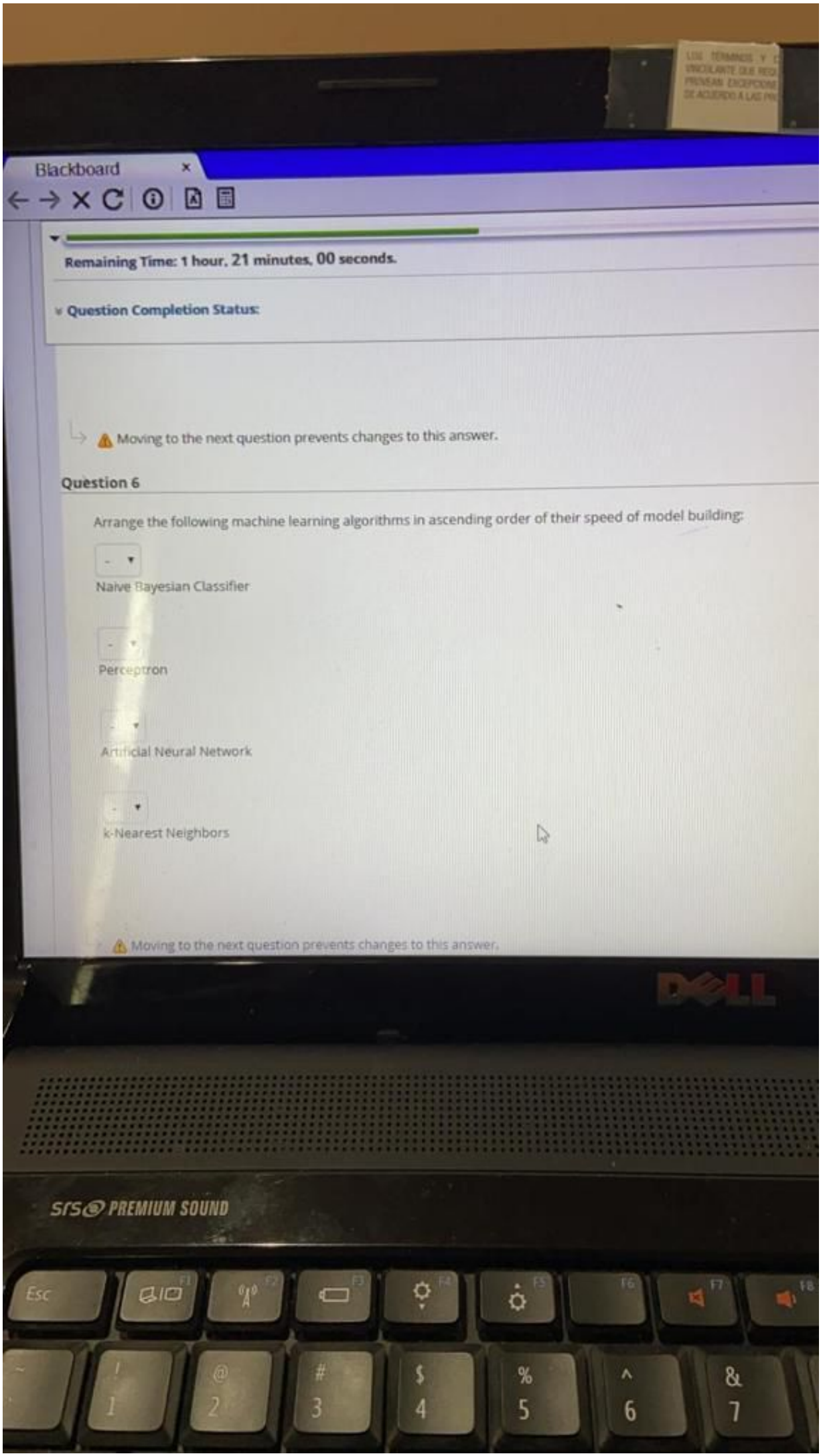
0.00336

0.00144

0.00216

Click **Submit** to complete this assessment.


Blackboard
Remaining Time: 1 hour, 42 minutes, 16 seconds.
Question Completion Status:
Moving to the next question prevents changes to this answer.
Question 5
Suppose there are a total of 100 points representing two classes in a dataset. You would like to run the AdaBoost algori
assume that the weak classifier is a stump hypothesis.
For the first round, the best stump classifier correctly classifies 80 points. Compute the values of ϵ_1 and α_1
$\epsilon_1 = 0.2$
$\alpha_1 = 0.693$
$\epsilon_1 = 0.2$
$\alpha_1 = 1.386$
$\epsilon_1 = 0.8$
$\alpha_1 = -0.693$
$\epsilon_1 = 0.2$
$\alpha_1 = 2.718$
Moving to the next question prevents changes to this answer.
MacBook Pro
esc
tab
caps lock
shift
fn
control
option
command


Blackboard
Remaining Time: 1 hour, 21 minutes, 00 seconds.
Question Completion Status:
Moving to the next question prevents changes to this answer.
Question 6
Arrange the following machine learning algorithms in ascending order of their speed of model building:
Naive Bayesian Classifier
Perceptron
Artificial Neural Network
k-Nearest Neighbors
Moving to the next question prevents changes to this answer.
DELL
SRS PREMIUM SOUND
Esc

Remaining Time: 1 hour, 29 minutes, 37 seconds.

Question Completion Status:

Question 7

5 points

You are given the dataset below for 1 attribute (X) and its associated class:

Dataset:

X	-0.1	0.7	1.0	1.6	2.0	2.5	3.2	3.5	4.1	4.9
Class	-	+	+	-	+	+	-	-	+	+

You decide to use k-Nearest Neighbor approach on this dataset. To find the best value of k, you decide to use 10-fold cross validation and Manhattan distance. Which of the following are true?

- [x] Using k=1, the total error for 10-fold cross validation is 40%
- [ ] Using k=1, the total error for 10-fold cross validation is 20%
- [ ] Using k=1, the total error for 10-fold cross validation is 80%
- [ ] Using k=3, the total error for 10-fold cross validation is 40%
- [x] Using k=3, the total error for 10-fold cross validation is 80%
- [ ] Using k=3, the total error for 10-fold cross validation is 20%

⚠ Moving to the next question prevents changes to this answer.

Question 7 of 30

Blackboard

Remaining Time: 1 hour, 21 minutes, 54 seconds.

Question Completion Status:

Question 9

Suppose we are developing a logistic regression model which is expressed as:

$$h_\theta(x) = P(Y=1|X;\theta) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$$

Note that θ is the weights vector. Suppose for two-attribute case, you obtain the weights vector as $\theta_0 = 6, \theta_1 = -1, \theta_2 = 0$.

Which one of the following would be true for the decision boundary in the attribute plane?

Image shown below

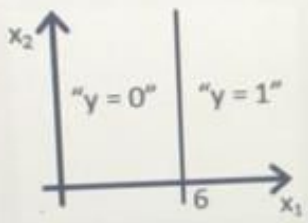


Image shown below

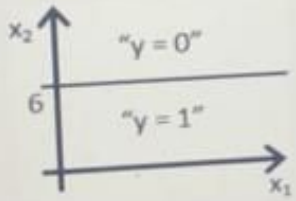


Image shown below

Honorlock Proc

ctoring session in

Live
02:19

ment.

Remaining Time: 1 hour, 57 minutes, 40 seconds.

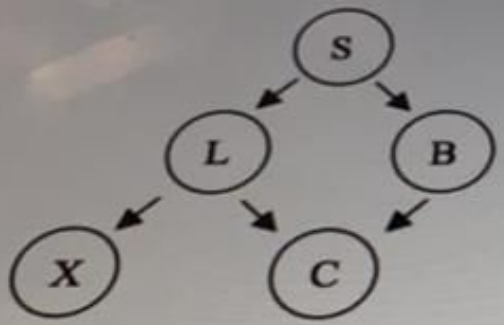
Question Completion Status:

QUESTION 4

3 points

Save Answer

Given the following Bayesian net:



Assuming you have evidence about L and C nodes, then which one(s) of the following will NOT be conditionally independent?

- ☐ S and X
- ☐ S and B
- ☐ L and B
- ☐ S and C

QUESTION 5

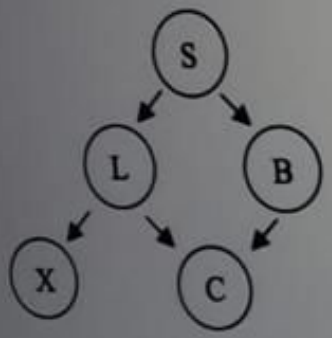
If a model outputs constant value irrespective of data, it would indicate which of the following scenarios: (Choose all correct choices)

Remaining Time: 1 hour, 23 minutes, 30 seconds.

Question Completion Status:

Question 9

Given the following Bayesian net:



Assuming you have evidence about L and C nodes, then which one(s) of the following will NOT be conditionally independent?

- S and X
- S and C
- L and B
- S and B

Moving to the next question prevents changes to this answer.

Remaining Time: 1 hour, 13 minutes, 03 seconds.

Question Completion Status:

Close Window

Moving to the next question prevents changes to this answer.

Question 10 of 30

Question 10

5 points Save Answer

Professor Ann spends 60% of her time in her office. The rest of her time is spent elsewhere. When Ann is in her office, half the time her light is off (when she is trying to hide from students and get research done). When she is not in her office, she leaves her light on only 5% of the time. 80% of the time she is in her office, Ann is logged onto the computer. Because she sometimes logs onto the computer from home, 10% of the time she is not in her office, she is still logged onto the computer.
Suppose a student checks Ann's login status and sees that she is logged on. What effect does this have on the student's belief that her light is on?

- 0.92
- 0.32
- 0.5
- 0.465

Question 10 of 30

Moving to the next question prevents changes to this answer.

Close Window

⚠ Moving to the next question prevents changes to this answer.

Question 12

In which of the following cases, will k-means NOT work well (More than one answers can be correct)

- When means of data points cannot be defined clearly
- When data contains lots of outliers
- When you have a good prior estimate of the number of clusters
- When data is non-globular (non-spherical) in shape

⚠ Moving to the next question prevents changes to this answer.

Question 12

5 points

You are given a training dataset below where each instance consists of 3 features - F1, F2, and F3. Each feature can take one of three values - {a, b, c}. Each instance is labeled as either + or -.

F1	F2	F3	Class Label
c	c	c	+
a	a	b	+
a	c	c	-
b	a	a	-
c	c	b	-

You have to train a 3-Nearest Neighbor (3-NN) algorithm on this dataset and find the predicted class label of the following test data point:
X(F1 = a, F2 = c, F3 = b
For calculating distance of two data points, you will use the Hamming distance, which is defined as the fraction of the features that two data points differ on.

$$d_{Hamming}(x1, x2) = \frac{Number\ of\ features\ where\ x1\ and\ x2\ differ}{Total\ number\ of\ features}$$

Which of the following statements are true?

- [x] The predicted class label of the test data point would be class -
- [ ] The predicted class label of the test data point would be class +
- [x] The distance of the test point to the first training point is 2/3
- [ ] The distance of the test point to the second training point is 2/3
- [x] The distance of the test point to the third training point is 1/3
- [x] The distance of the test point to the last training point is 1/3

Force Completion

This test can be ...

This test does not allow backtracking. Changes to the answer after submission are p...

Remaining Time: 1 hour, 03 minutes, 04 seconds.

Question Completion Status:

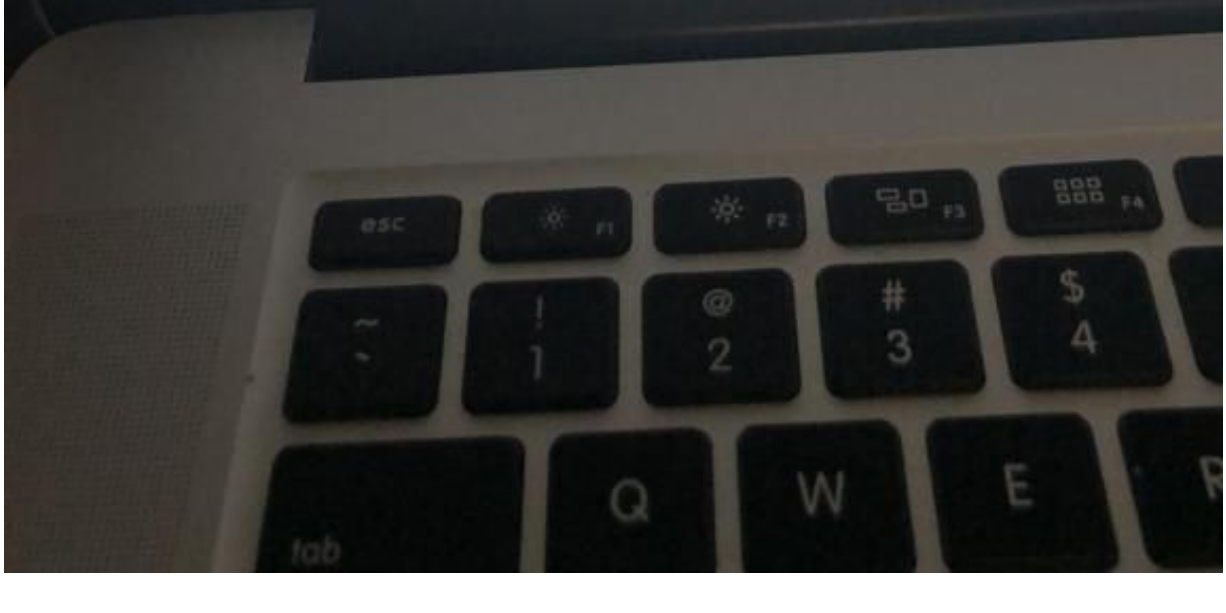
⚠ Moving to the next question prevents changes to this answer.

Question 13

Which of the following would lead to reduction in variance?

- Decrease in the number of layers of a neural network
- In k Nearest Neighbors, decreasing the value of k
- Pruning of decision tree
- Regularization of weights of a linear regression model

⚠ Moving to the next question prevents changes to this answer.



⚠ Moving to the next question prevents changes to this answer.

Question 1

Question 14

3 points

Which of the following statements are true?

Consider two models:
M1 uses a polynomial of degree 9 to fit the training data
M2 uses a polynomial of degree 2 to fit the training data

M1 would have a lower bias and greater variance than M2

If there is a model that always outputs a constant value, it would have 0 bias and high variance.

If there is a model that always outputs a constant value, it would have high bias and 0 variance.

Bagging is a technique that can reduce variance without increasing bias significantly.

If we post-prune a fully grown decision tree, we would reduce variance

As we run more iterations of the backpropagation algorithm on a neural network, we are increasing the variance.

Question 14 of

⚠ Moving to the next question prevents changes to this answer.

Close Window

emaining Time: **54 minutes, 23 seconds.**

Question Completion Status:

Close Window

Question 14 of 30

⚠ Moving to the next question prevents changes to this answer.

3 points Save Answer

Question 14

Suppose there are a total of 100 points representing two classes in a dataset. You would like to run the AdaBoost algorithm on this dataset, using a single weak classifier at each step. You can assume that the weak classifier is a stump hypothesis. For the first round, the best stump classifier correctly classifies 80 points.
What would be the updated normalized weights of the misclassified and the correctly classified points after the first round?
Hint: Normalized weights for all the points in the dataset sum up to 1.

- Normalized weight of each misclassified point = 0.0333
  Normalized weight of each correctly classified point = 0.0833
- Normalized weight of each misclassified point = 0.025
  Normalized weight of each correctly classified point = 0.0833
- Normalized weight of each misclassified point = 0.025
  Normalized weight of each correctly classified point = 0.00625
- Normalized weight of each misclassified point = 0.02
  Normalized weight of each correctly classified point = 0.005

Question 14 of 30

Close Window

⚠ Moving to the next question prevents changes to this answer.

⚠ Moving to the next question prevents changes to this answer.

Question 1

Question 14

3 points

Which of the following statements are true?

Consider two models:
M1 uses a polynomial of degree 9 to fit the training data
M2 uses a polynomial of degree 2 to fit the training data

- [ ] M1 would have a lower bias and greater variance than M2
- [ ] If there is a model that always outputs a constant value, it would have 0 bias and high variance.
- [ ] If there is a model that always outputs a constant value, it would have high bias and 0 variance.
- [ ] Bagging is a technique that can reduce variance without increasing bias significantly.
- [ ] If we post-prune a fully grown decision tree, we would reduce variance
- [ ] As we run more iterations of the backpropagation algorithm on a neural network, we are increasing the variance.

Question 14 of

⚠ Moving to the next question prevents changes to this answer.

Close Window

QUESTION 9

Suppose there are a total of 100 points representing two classes in a dataset. You would like to run the AdaBoost algorithm on this dataset, using a single weak classifier at each step. You can assume that the weak classifier is a stump hypothesis. For the first round, the best stump classifier correctly classifies 80 points.
What would be the updated normalized weights of the misclassified and the correctly classified points after the first round?
Hint: Normalized weights for all the points in the dataset sum up to 1.

- Normalized weight of each misclassified point = 0.025
  Normalized weight of each correctly classified point = 0.00625
- Normalized weight of each misclassified point = 0.02
  Normalized weight of each correctly classified point = 0.005
- Normalized weight of each misclassified point = 0.0333
  Normalized weight of each correctly classified point = 0.0833
- Normalized weight of each misclassified point = 0.025
  Normalized weight of each correctly classified point = 0.0833

QUESTION 10

Suppose there are a set of n discrete random variables (x1, x2, ..., xn) that can each take m distinct values. For the cases indicates on the left, find out the total number of parameters needed to specify their full joint probability distribution i.e. p(x1, x2, ..., xn)

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 15

5 points Save

Question 15

You are given the dataset below for 1 attribute (X) and its associated class:

Dataset:

X	-0.1	0.7	1.0	1.6	2.0	2.5	3.2	3.5	4.1	4.9
Class	-	+	+	-	+	+	-	-	+	+

You decide to use k-Nearest Neighbor approach on this dataset. To find the best value of k, you decide to use 10-fold cross validation and Manhattan distance. Which of the following are true?

- Using k=1, the total error for 10-fold cross validation is 40%
- Using k=1, the total error for 10-fold cross validation is 20%
- Using k=1, the total error for 10-fold cross validation is 80%
- Using k=3, the total error for 10-fold cross validation is 40%
- Using k=3, the total error for 10-fold cross validation is 80%
- Using k=3, the total error for 10-fold cross validation is 20%

Moving to the next question prevents changes to this answer.

Remaining Time: **56 minutes, 35 seconds.**

Question Completion Status:

Close Window

Question 16 of 30

Moving to the next question prevents changes to this answer.

5 points

Save Answer

Question 16

Suppose there are a set of n discrete random variables (x1, x2, ... xn) that can each take m distinct values. For the cases indicates on the left, find out the total number of parameters needed to specify their full joint probability distribution i.e. p(x1, x2, ...,xn)

A. Assuming they are all dependent on each other

C. Assuming they are completely independent

Assuming they form a Bayes net such that there is one root node, one node with a single parent, two nodes with two parents, and the remaining all have 3 parents each.
Note: You can assume that each node is dependent only on its parents.

Assuming they form a sequential Bayes net, such as the one shown below:
$x_1 \rightarrow x_2 \rightarrow .... \rightarrow x_n$
Note: You can assume that each node is dependent only on its parents.

A. $m^n - 1$

B. $(n-1)\,m^2 + m - n$

C. $n*(m-1)$

D. $m + m^2 + 2\,m^3 + (n-4)\,m^4 - n$

Question 16 of 30

Moving to the next question prevents changes to this answer.

Close Window

MacBook Air

Remaining Time: 1 hour, 03 minutes, 52 seconds.

Question Completion Status:

Close Window

Moving to the next question prevents changes to this answer.

Question 16 of 30

Question 16

2 points Saved

Arrange the following machine learning algorithms in ascending order of their speed of model building:

3.
Naive Bayesian Classifier

2.
Perceptron

1.
Artificial Neural Network

4.
k-Nearest Nelghbors

Question 16 of 30

Moving to the next question prevents changes to this answer.

Close Window

Question Completion Status:

Close Window

Moving to the next question prevents changes to this answer.

Question 16 of 30

Question 16

2 points Save Answer

Arrange the following machine learning algorithms in ascending order of their speed of model building:

-

Naive Bayesian Classifier

-

Perceptron

-

Artificial Neural Network

-

k-Nearest Neighbors

Moving to the next question prevents changes to this answer.

Question 16 of 30

Close Window

Normalized weight of each correctly classified point = 0.0833

QUESTION 10

5 points

Save

Suppose there are a set of n discrete random variables (x1, x2, ..., xn) that can each take m distinct values. For the cases indicates on the left, find out the total number of parameters needed to specify their full joint probability distribution i.e. p(x1, x2, ..., xn)

- Assuming they are all dependent on each other
- Assuming they are completely independent
- Assuming they form a Bayes net such that there is one root node, one node with a single parent, two nodes with two parents, and the remaining all have 3 parents each. Note: You can assume that each node is dependent only on its parents.
- Assuming they form a sequential Bayes net, such as the one shown below: $x_1 \rightarrow x_2 \rightarrow .... \rightarrow x_n$

Note: You can assume that each node is dependent only on its parents.

A. $m^n - 1$

B. $m + m^2 + 2\,m^3 + (n-4)\,m^4 - n$

C. $(n-1)\,m^2 + m - n$

D. $n^*(m-1)$

QUESTION 11

Suppose you have n points x1, x2, ..., xn that are independently drawn from the following probability distribution, where A is a

Remaining Time: **56 minutes, 35 seconds.**

Question Completion Status:

Close Window

Question 16 of 30

Moving to the next question prevents changes to this answer.

5 points

Save Answer

Question 16

Suppose there are a set of n discrete random variables (x1, x2, ... xn) that can each take m distinct values. For the cases indicates on the left, find out the total number of parameters needed to specify their full joint probability distribution i.e. p(x1, x2, ...,xn)

A. Assuming they are all dependent on each other

C. Assuming they are completely independent

Assuming they form a Bayes net such that there is one root node, one node with a single parent, two nodes with two parents, and the remaining all have 3 parents each.
Note: You can assume that each node is dependent only on its parents.

Assuming they form a sequential Bayes net, such as the one shown below:
$x_1 \rightarrow x_2 \rightarrow .... \rightarrow x_n$
Note: You can assume that each node is dependent only on its parents.

A. $m^n - 1$

B. $(n-1)\,m^2 + m - n$

C. $n*(m-1)$

D. $m + m^2 + 2\,m^3 + (n-4)\,m^4 - n$

Question 16 of 30

Moving to the next question prevents changes to this answer.

Close Window

MacBook Air

Question 17

You are given the dataset below for 1 attribute (X) and its associated class:

Dataset:

X	-0.1	0.7	1.0	1.6	2.0	2.5	3.2	3.5	4.1	4.9
Class	-	+	+	-	+	+	-	-	+	+

You decide to use k-Nearest Neighbor approach on this dataset. To find the best value of k, you decide to use 10-fold cross validation and Manhattan distance. Which of the following are true?

- [ ] Using k=1, the total error for 10-fold cross validation is 40%
- [ ] Using k=1, the total error for 10-fold cross validation is 20%
- [ ] Using k=1, the total error for 10-fold cross validation is 80%
- [ ] Using k=3, the total error for 10-fold cross validation is 40%
- [ ] Using k=3, the total error for 10-fold cross validation is 80%
- [ ] Using k=3, the total error for 10-fold cross validation is 20%

⚠ Moving to the next question prevents changes to this answer.

Multiple Attempts	Not allowed. This test can only be taken once.
Force Completion	This test can be saved and resumed at any point until time has expired. The
	This test does not allow backtracking. Changes to the answer after submissic

Remaining Time: 37 minutes, 00 seconds.

Question Completion Status:

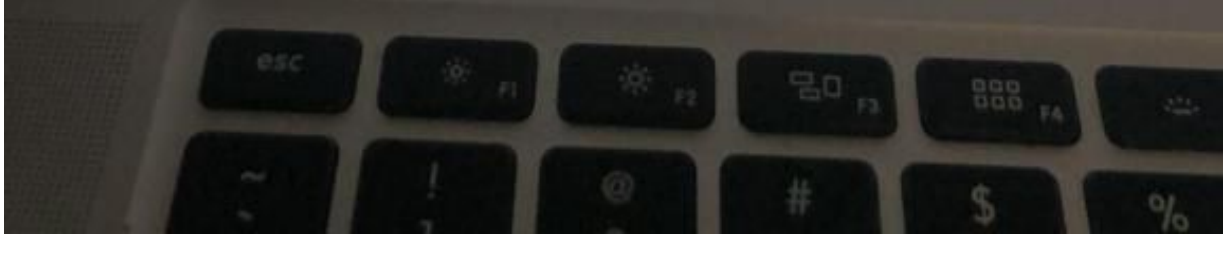
⚠ Moving to the next question prevents changes to this answer.

Question 19

A model with high variance has which of the following characteristics:

- [ ] It gives almost similar results for different datasets
- [ ] It denotes a very simple model e.g. a constant function
- [x] It is prone to overfitting
- [ ] It is highly affected by small changes in the dataset and noise

⚠ Moving to the next question prevents changes to this answer.



Remaining Time: 1 hour, 47 minutes, 09 seconds.

Question Completion Status:

⚠ Moving to the next question prevents changes to this answer.

Question 2

A model with high variance has which of the following characteristics:

- [ ] It is prone to overfitting
- [ ] It gives almost similar results for different datasets
- [x] It is highly affected by small changes in the dataset and noise
- [ ] It denotes a very simple model e.g. a constant function

⚠ Moving to the next question prevents changes to this answer.

Remaining Time: **56 minutes, 18 seconds.**

Question Completion Status:

Close Window

⚠ Moving to the next question prevents changes to this answer.

Question 19

Question 19 of 30

3 points | Save Answer

Suppose you have trained three classifiers, each of which returns either 1 or −1, and tested their accuracies to find the following:

Classifier	Accuracy
c1	0.6
c2	0.55
c3	0.45

Let C be the classifier that returns a simple majority vote of the three classifiers. Assuming the errors of the ci are independent, what is the probability that C(x) will be correct on a new test example x?

- 0.1815
- 0.5505
- 0.1215
- 0.099

⚠ Moving to the next question prevents changes to this answer.

Question 19 of 30

Close Window

MacBook Pro

Blackboard

Remaining Time: 39 minutes, 32 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 20 of 30

Question 20

3 points Saved

Suppose you have n points x1, x2, ...xn that are independently drawn from the following probability distribution, where θ is a parameter.

$$Pr(x) = \frac{\theta^x e^{-5\theta}}{x!}$$

What would be the maximum likelihood estimate for θ

$$\theta_{MLE} = \frac{\prod_{i=1}^{n} x_i}{5n}$$

$$\theta_{MLE} = \frac{\prod_{i=1}^{n} x_i}{n}$$

$$\theta_{MLE} = \frac{\sum_{i=1}^{n} x_i}{5n}$$

Question Completion Status:

⚠ Moving to the next question prevents changes to this answer.

Question 21

5 points ✓ Saved

You are given following 8 data points.
(1, 0), (3,2), (3,1), (2,0), (4, 2), (0, 1.5), (2, 3), (1, -1.5)
You decide to run k-means algorithm on them using k=2 and intial centroids as the following two points (1,0) (called cluster1) and (3,2) (called cluster 2). You will be using simple Manhattan distance for calculations.

How many total iterations will the k-means algorithm take before converging?
Also, let's say we define the total error of clustering as:

$$\sum_i d(x_i, \mu_j)$$

i.e. the distance of each point (x_i) to the center of the cluster (μ_j) to which it is assigned to. You will use Manhattan distance for the calculations. What will be the final value of the total error?

Definition of convergence: If the clustering doesn't change after the n^{th} iteration, then number of iterations required is equal to n

- Not enough information is available to compute these values.
- Number of iterations = 1
  Total Error = 9.0
- Number of iterations = 2
  Total Error = 9.0
- Number of iterations = 1
  Total Error = 7.0

⚠ Moving to the next question prevents changes to this answer.

Close Window

Question 21

5 points Saved

You are given following 8 data points.
(1, 0), (3,2), (3,1), (2,0), (4, 2), (0, 1.5), (2, 3), (1, -1.5)
You decide to run k-means algorithm on them using k=2 and intial centroids as the following two points (1,0) (called cluster1) and (3,2) (called cluster 2). You will be using simple Manhattan distance for calculations.

How many total iterations will the k-means algorithm take before converging?
Also, let's say we define the total error of clustering as:

$$\sum_i d(x_i, \mu_j)$$

i.e. the distance of each point (x_i) to the center of the cluster (μ_j) to which it is assigned to. You will use Manhattan distance for the calculations. What will be the final value of the total error?

Definition of convergence: If the clustering doesn't change after the n^{th} iteration, then number of iterations required is equal to n

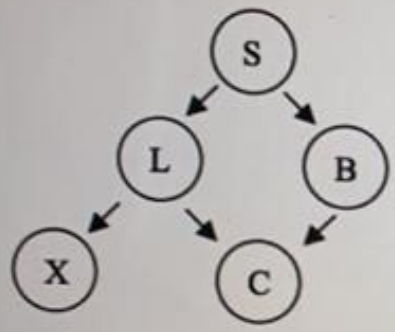
- Not enough information is available to compute these values.
- Number of iterations = 1
  Total Error = 9.0
- Number of iterations = 2
  Total Error = 9.0
- Number of iterations = 1
  Total Error = 7.0

Question Completion Status:

Question 22

3 points Save Answer

Given the following Bayesian net:



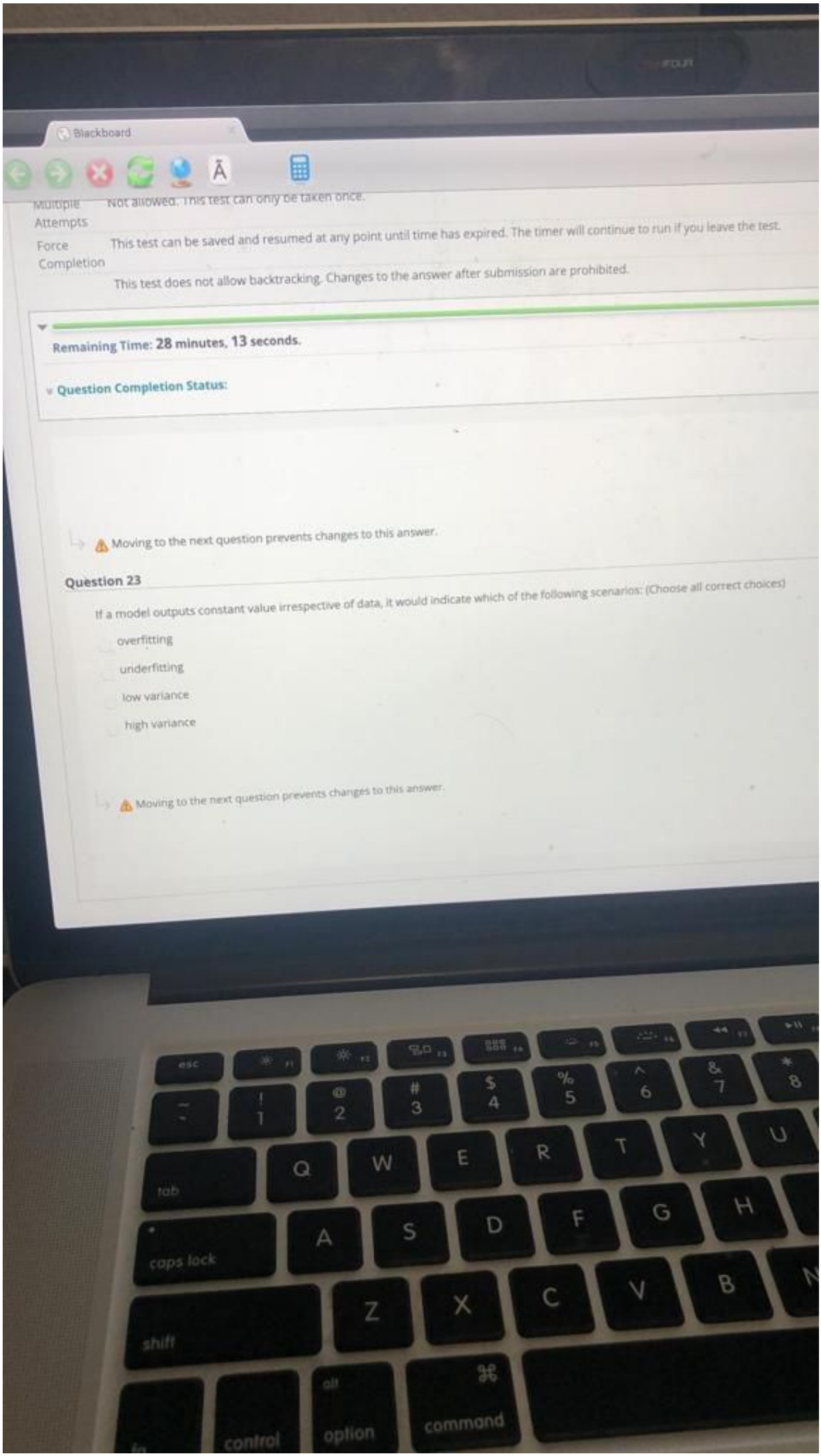
Assuming you have evidence about L and C nodes, then which one(s) of the following will NOT be conditionally independent?

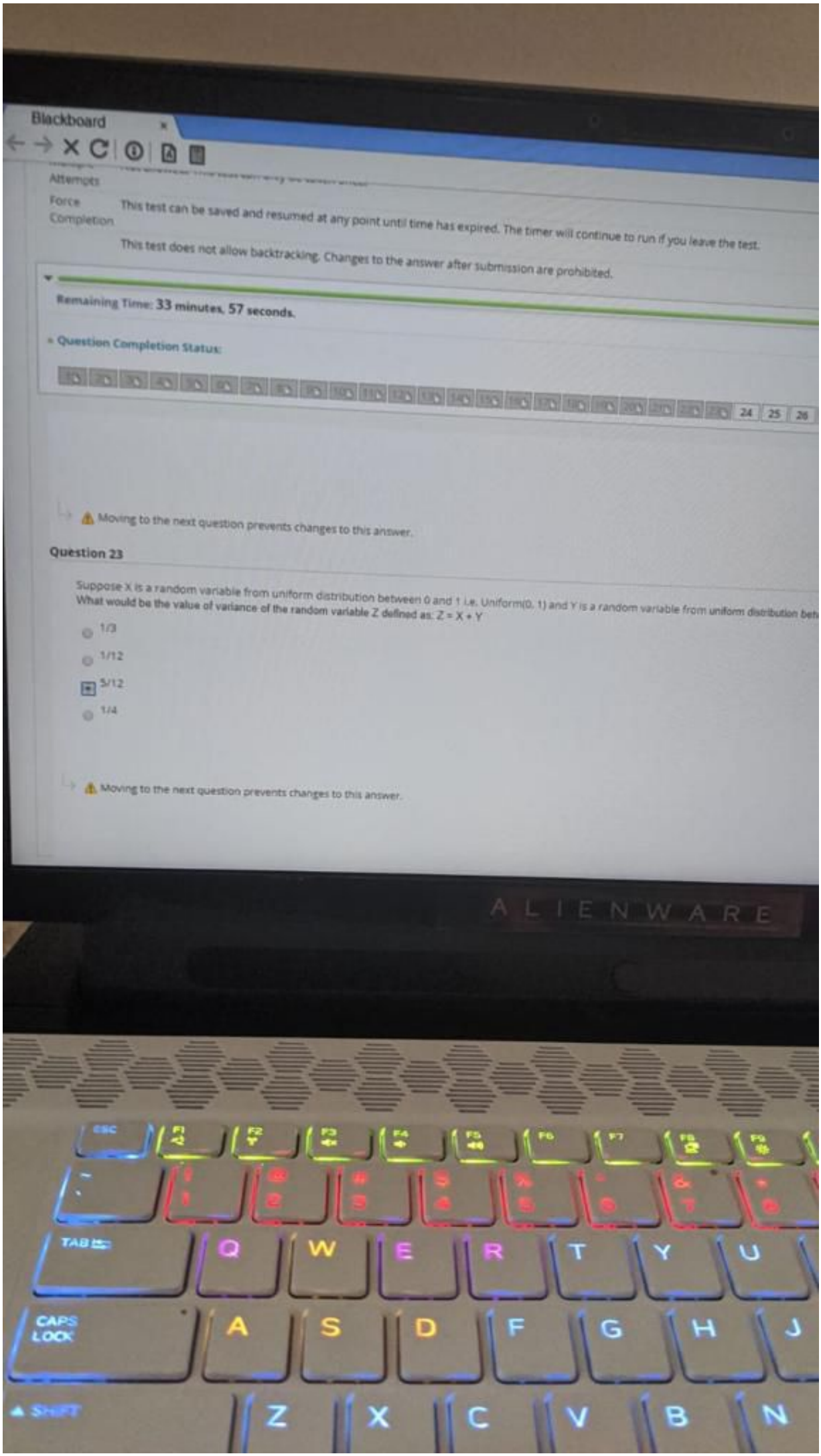
- [ ] L and B
- [x] S and C
- [x] S and B
- [x] S and X

Moving to the next question prevents changes to this answer.

Question 22 of 30

Close Window


Blackboard
Multiple Attempts
Not allowed. This test can only be taken once.
Force Completion
This test can be saved and resumed at any point until time has expired. The timer will continue to run if you leave the test.
This test does not allow backtracking. Changes to the answer after submission are prohibited.
Remaining Time: 28 minutes, 13 seconds.
Question Completion Status:
Moving to the next question prevents changes to this answer.
Question 23
If a model outputs constant value irrespective of data, it would indicate which of the following scenarios: (Choose all correct choices)
overfitting
underfitting
low variance
high variance
Moving to the next question prevents changes to this answer.
esc
tab
caps lock
shift
control
option
command


Blackboard
Attempts
Force Completion
This test can be saved and resumed at any point until time has expired. The timer will continue to run if you leave the test.
This test does not allow backtracking. Changes to the answer after submission are prohibited.
Remaining Time: 33 minutes, 57 seconds.
Question Completion Status:
24
25
26
Moving to the next question prevents changes to this answer.
Question 23
Suppose X is a random variable from uniform distribution between 0 and 1 i.e. Uniform(0, 1) and Y is a random variable from uniform distribution bet
What would be the value of variance of the random variable Z defined as: Z = X + Y
1/3
1/12
5/12
1/4
Moving to the next question prevents changes to this answer.
ALIENWARE
ESC
TAB
CAPS LOCK
SHIFT
Q
W
E
R
T
Y
U
A
S
D
F
G
H
J
Z
X
C
V
B
N

Blackboard

Remaining Time: 22 minutes, 29 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 24

Suppose you have trained three classifiers, each of which returns either 1 or −1, and tested their accuracies to find the following:

Classifier	Accuracy
c1	0.6
c2	0.55
c3	0.45

Let C be the classifier that returns a simple majority vote of the three classifiers. Assuming the errors of the ci are independent, what is the probability

0.5505

0.1815

0.1215

0.099

Moving to the next question prevents changes to this answer.

Suppose you have trained three classifiers, each of which returns either 1 or −1, and tested their accuracies to find the following:

Classifier	Accuracy
c1	0.6
c2	0.55
c3	0.45

Let C be the classifier that returns a simple majority vote of the three classifiers. Assuming the errors of the ci are independent, what is the probability that C(x) will be correct on a new test example x?

- 0.1215
- 0.099
- 0.5505
- 0.1815

ck Save and Submit to save and submit. Click Save All Answers to save all answers.

Question Completion Status:

Close Window

Moving to the next question prevents changes to this answer.

Question 25 of 30

Question 25

5 points Save Answer

Professor Ann spends 60% of her time in her office. The rest of her time is spent elsewhere. When Ann is in her office, half the time her light is off (when she is trying to hide from students and get research done). When she is not in her office, she leaves her light on only 5% of the time. 80% of the time she is in her office, Ann is logged onto the computer. Because she sometimes logs onto the computer from home, 10% of the time she is not in her office, she is still logged onto the computer.
Suppose a student checks Ann's login status and sees that she is logged on. What effect does this have on the student's belief that her light is on?

- 0.465
- 0.5
- 0.32
- 0.92

Moving to the next question prevents changes to this answer.

Question 25 of 30

Close Window

Blackboard

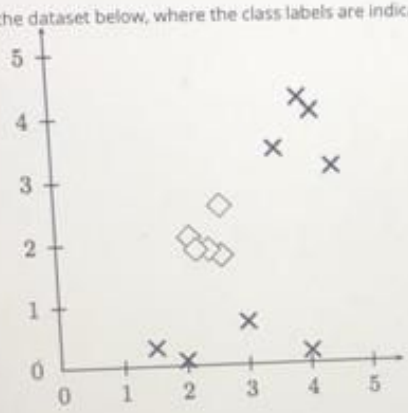
Remaining Time: 12 minutes, 08 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

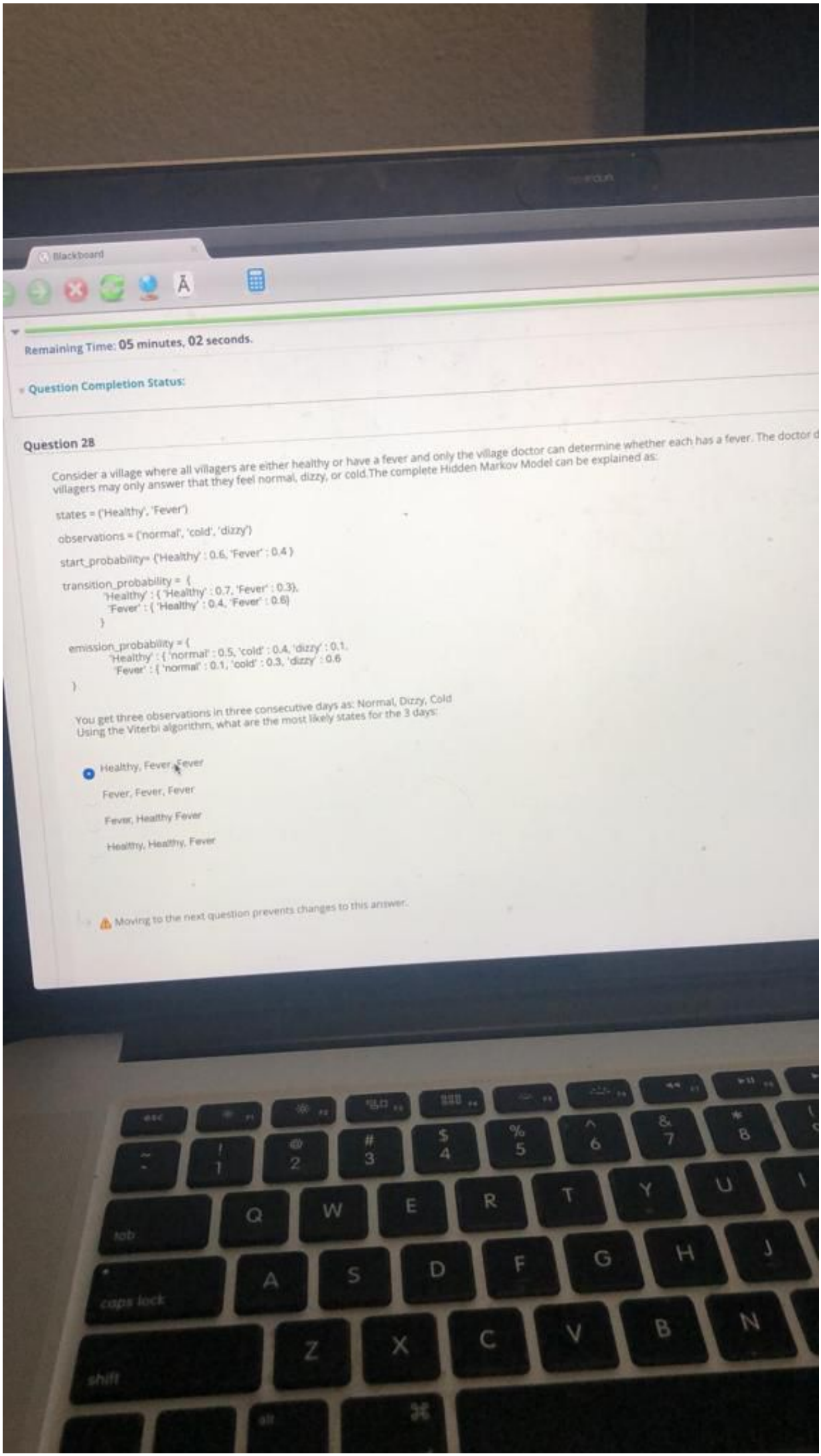
Question 26

Given the dataset below, where the class labels are indicated by the two different shapes. Indicate which of the following classifiers would



- [x] SVM with quadratic kernel
- [x] 3-Nearest Neighbor
- [ ] Adaboost with decision tree stumps
- [x] Decision tree of maximum depth 2
- [ ] Logistic Regression

Moving to the next question prevents changes to this answer.



Blackboard

Remaining Time: 05 minutes, 02 seconds.

Question Completion Status:

Question 28

Consider a village where all villagers are either healthy or have a fever and only the village doctor can determine whether each has a fever. The doctor d
villagers may only answer that they feel normal, dizzy, or cold.The complete Hidden Markov Model can be explained as:

```
states = ('Healthy', 'Fever')

observations = ('normal', 'cold', 'dizzy')

start_probability= ('Healthy' : 0.6, 'Fever' : 0.4 )

transition_probability = {
        'Healthy' : { 'Healthy' : 0.7, 'Fever' : 0.3},
        'Fever' : { 'Healthy' : 0.4, 'Fever' : 0.6}
    }

emission_probability = {
        'Healthy' : { 'normal' : 0.5, 'cold' : 0.4, 'dizzy' : 0.1},
        'Fever' : { 'normal' : 0.1, 'cold' : 0.3, 'dizzy' : 0.6
}
```

You get three observations in three consecutive days as: Normal, Dizzy, Cold
Using the Viterbi algorithm, what are the most likely states for the 3 days:

- Healthy, Fever, Fever
- Fever, Fever, Fever
- Fever, Healthy Fever
- Healthy, Healthy, Fever

Moving to the next question prevents changes to this answer.

Blackboard

Remaining Time: 04 minutes, 22 seconds.

Question Completion Status:

Moving to the next question prevents changes to this answer.

Question 29

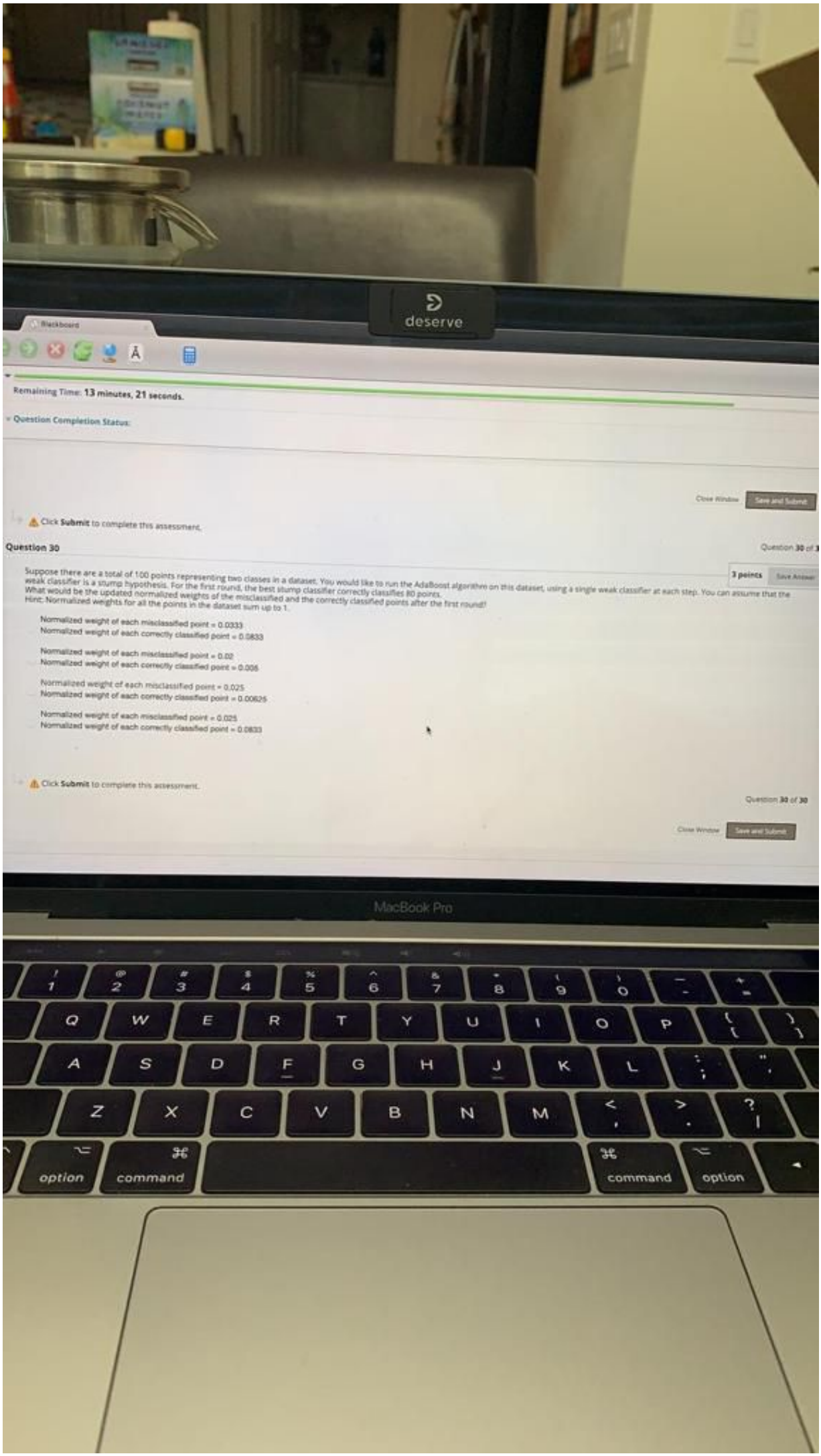
You are given following 8 data points.
(1, 0), (3,2), (3,1), (2,0), (4, 2), (0, 1.5), (2, 3), (1, -1.5)
You decide to run k-means algorithm on them using k=2 and intial centroids as the following two points (1,0) (called cluster1) and (3,2) (called cluster 2). You will be using simple Manh
calculations.

How many total iterations will the k-means algorithm take before converging?
Also, let's say we define the total error of clustering as:
$\sum_i d(x_i, \mu_j)$
i.e. the distance of each point (x_i) to the center of the cluster (μ_j) to which it is assigned to. You will use Manhattan distance for the calculations. What will be the final value of the tota

Definition of convergence: If the clustering doesn't change after the n^{th} iteration, then number of iterations required is equal to n

- Number of iterations = 1
  Total Error = 9.0
- Not enough information is available to compute these values.
- Number of iterations = 1
  Total Error = 7.0
- Number of iterations = 2
  Total Error = 9.0

Moving to the next question prevents changes to this answer.


deserve
Remaining Time: 13 minutes, 21 seconds.
Question Completion Status:
Close Window
Save and Submit
Click Submit to complete this assessment.
Question 30
Question 30 of 30
3 points
Save Answer
Suppose there are a total of 100 points representing two classes in a dataset. You would like to run the AdaBoost algorithm on this dataset, using a single weak classifier at each step. You can assume that the weak classifier is a stump hypothesis. For the first round, the best stump classifier correctly classifies 80 points.
What would be the updated normalized weights of the misclassified and the correctly classified points after the first round?
Hint: Normalized weights for all the points in the dataset sum up to 1.
Normalized weight of each misclassified point = 0.0333
Normalized weight of each correctly classified point = 0.0833
Normalized weight of each misclassified point = 0.02
Normalized weight of each correctly classified point = 0.006
Normalized weight of each misclassified point = 0.025
Normalized weight of each correctly classified point = 0.00625
Normalized weight of each misclassified point = 0.025
Normalized weight of each correctly classified point = 0.0833
Click Submit to complete this assessment.
Question 30 of 30
Close Window
Save and Submit
MacBook Pro
option
command
command
option