Week 1 : Assignment

The due date for submitting this assignment has passed.

Due on 2024-08-07, 23:59 IST.

Assignment submitted on 2024-08-07, 23:09 IST

1 point
You are working on a project to classify images of handwritten digits. You have a large dataset of images, and labels associated with them. Which learning approach would be most appropriate for this task?

Supervised Learning
Unsupervised Learning
Semi-Supervised Learning
Reinforcement Learning

Yes, the answer is correct.
Score: 1

Accepted Answers:
Supervised Learning

1 point
You have a dataset containing various attributes of customers (e.g., age, income, spending habits) but no labels or categories. You want to group the customers into distinct segments based on their similarities. Which learning approach would be most appropriate for this task?

Supervised Learning
Reinforcement Learning
Unsupervised Learning
Semi-Supervised Learning

Yes, the answer is correct.
Score: 1

Accepted Answers:
Unsupervised Learning

1 point
You are developing an AI agent to play a video game, where the agent must learn to navigate the game environment and maximize its score through trial and error. Which learning approach would be most appropriate for this task?

Supervised Learning
Unsupervised Learning
Reinforcement Learning
Semi-Supervised Learning

Yes, the answer is correct.
Score: 1

Accepted Answers:
Reinforcement Learning

1 point
Match the following tasks to the most appropriate mathematical methods used to solve them:

Tasks:

1. Predicting the probability of an event for a given sample space.
2. Solving a system of linear equations.
3. Finding the minimum or maximum value of a function subject to constraints.
4. Analyzing the statistical properties such as variance and distribution of data for a given dataset.

Mathematical Methods:

a. Linear Algebra
b. Probability and Statistics
c. Optimization

1 -> a
2 -> b
3 -> c
4 -> b
1 -> b
2 -> a
3 -> c
4 -> b
1 -> c
2 -> a
3 -> b
4 -> c

1 -> b
2 -> a
3 -> c
4 -> a

Yes, the answer is correct.
Score: 1

Accepted Answers:
1 -> b
2 -> a
3 -> c
4 -> b

1 point
Which of the following lines **best** represents the data points in the below figure?

Line A
Line B
Line C

Yes, the answer is correct.
Score: 1

Accepted Answers:
Line A

1 point
Consider an experiment of rolling 7 dice. Assume each die has 6 faces. What is the total number of possible outcomes (i.e., the size of a sample space)?

46656
117649
115649
279936

Yes, the answer is correct.
Score: 1

Accepted Answers:
279936

With continuation to the previous question what is the probability of getting atleast 2 even numbers? (Hint : Use $^{n}C_{k}$ to find out the combinations of the events that can happen and then find the probability.)

[Note : Write it till 4 decimal places. For example if the answer is 0.123456, write till 0.1234]

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Range) 0.93,0.94

2 points
2 points
In the Cookie Jar Game Show, you're a contestant and given the choice of three jars. Inside one jar is a golden cookie; inside the other two jars are regular cookies. You pick a jar, say Jar No. 1, and the host, who knows what's inside each jar, opens another jar, say Jar No. 3, which has a regular cookie. The host then asks if you want to switch to Jar No. 2. Should you switch?

Yes, I will switch
No, I won't

Yes, the answer is correct.
Score: 2

Accepted Answers:
Yes, I will switch

From NPTEL's past data, it is observed that 70% of people who do assignments on their own in the NPTEL's course "Introduction to Machine Learning (Tamil)" without copying answers from others can pass the final exam. The overall pass percentage of the course is 40%. From the data it seems that 50% of people do their assignments on their own. If a person is randomly selected, given they passed the course, how confident are you that the person would have not cheated?

[Note : Write it till 3 decimal places. For example if the answer is 0.123456, write till 0.123]

Yes, the answer is correct.
Score: 2

Accepted Answers:
(Type: Range) 0.87,0.88

The due date for submitting this assignment has passed.

Due on 2024-08-07, 23:59 IST.

Assignment submitted on 2024-08-07, 23:06 IST

1) Let w be a vector in $\mathbb{R}^2$. Which of the following is always true about the vector? [MSQ] *2 points*

- [ ] $||\mathbf{w}|| < 0$
- [x] $||\mathbf{w}|| \geq 0$
- [ ] $||\mathbf{w}||^2$ is defined as the length of the vector which measures the distance from the origin to a particular point.
- [ ] $||\mathbf{w}||$ is defined as the length of the vector which measures the distance from the origin to a particular point.

Partially Correct.
Score: 1
Accepted Answers:
$||\mathbf{w}|| \geq 0$
$||\mathbf{w}||$ *is defined as the length of the vector which measures the distance from the origin to a particular point.*

2) What is the length of the follwoing vector $\mathbf{w} = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$ starting from the origin?

Answer with 2 decimal places i.e. if the solution is 0.1234, then only type 0.12 as the answer. [NAT]

9.11

Yes, the answer is correct.
Score: 1
Accepted Answers:
(Type: Range) 9,9.2

1 point

3) Which of the following statements about the vectors are true? $\mathbf{u} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T$ and $\mathbf{v} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$ (MSQ) *1 point*

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=27&assessment=174

1 point

3) Which of the following statements about the vectors are true? $\mathbf{u} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T$ and $\mathbf{v} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$ (MSQ) *1 point*

- [ ] The vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal to each other and their respective lengths are equal to 1
- [ ] The vectors $2\mathbf{u}$ and $\mathbf{v}$ are orthonogonal to each other
- [ ] The vectors $\mathbf{u}$-$\mathbf{v}$ and $\mathbf{v}$ are orthogonal to each other but their respective lengths are not equal to 1
- [x] The vectors $\mathbf{u}$ and $\mathbf{v}$ not orthonogonal to each other

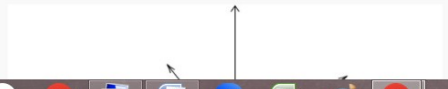
No, the answer is incorrect.
Score: 0
Accepted Answers:
The vectors $\mathbf{u}$ *and* $\mathbf{v}$ *are orthogonal to each other and their respective lengths are equal to 1*
The vectors $2\mathbf{u}$ *and* $\mathbf{v}$ *are orthonogonal to each other*

4) Given the vectors $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$, are they linearly independent? (MCQ) *1 point*

- ( ) Yes
- (•) No

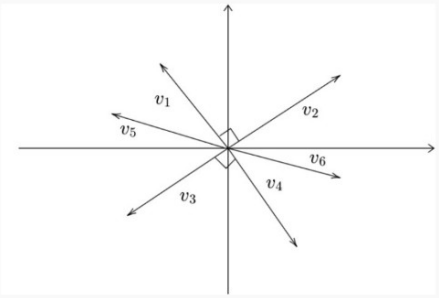
Yes, the answer is correct.
Score: 1
Accepted Answers:
No

5) Which of the following pair of vectors are orthogonal to each other? [MSQ] *2 points*



No

5) Which of the following pair of vectors are orthogonal to each other? [MSQ] **2 points**



- ☐ (v_1, v_6)
- ☐ (v_3, v_4)
- ☑ (v_1, v_2)
- ☐ (v_5, v_2)

Partially Correct.
Score: 1

Accepted Answers:
(v_3, v_4)
(v_1, v_2)

(v_3, v_4)
(v_1, v_2)

6) Which of the following sets of vectors forms a basis for $\mathbb{R}^3$?[MCQ] **2 points**

- ◉ $\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix} \right\}$
- ○ $\left\{ \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}, \begin{bmatrix}7\\8\\9\end{bmatrix} \right\}$
- ○ $\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix} \right\}$

Yes, the answer is correct.
Score: 2

Accepted Answers:
$\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix} \right\}$

7) In a simple linear regression model $\mathbf{y} = w_0 + w_1\mathbf{x} + \epsilon$, where $\mathbf{y}$ is the dependent variable, $\mathbf{x}$ is the independent variable, w_0 is the intercept, w_1 is the slope, and ϵ is some random noise present in the datset, which of the following statements is true?[MCQ] **1 point**

- ○ If w_1 is zero, the line is vertical and y does not depend on $\mathbf{x}$.
- ○ The sum of the residuals (ϵ) is always positive.
- ◉ The coefficient w_1 represents the amount by which $\mathbf{y}$ changes when $\mathbf{x}$ changes by one unit.
- ○ The error term (ϵ) is assumed to be correlated with $\mathbf{x}$

Yes, the answer is correct.
Score: 1

Accepted Answers:
The coefficient w_1 represents the amount by which $\mathbf{y}$ changes when $\mathbf{x}$ changes by one unit.

Score: 1
Accepted Answers:
The coefficient w_1 represents the amount by which $\mathbf{y}$ changes when $\mathbf{x}$ changes by one unit.

8) Which of the following is the objective function for the linear regression problem with sum squared error? Here m is the total number of data points. [MSQ] *1 point*

- $\max_{\mathbf{w}} \sum_{i=1}^{m} (\mathbf{w}^T \mathbf{x}_i - y_i)^2$
- $\min_{\mathbf{w}} \sum_{i=1}^{m} (\mathbf{w}^T \mathbf{x}_i - y_i)^2$
- $\max_{\mathbf{w}} \sum_{i=1}^{m} (y_i - \mathbf{w}^T \mathbf{x}_i)^2$
- $\min_{\mathbf{w}} \sum_{i=1}^{m} (\mathbf{w}^T \mathbf{x}_i - y_i)$

Yes, the answer is correct.
Score: 1
Accepted Answers:
$\min_{\mathbf{w}} \sum_{i=1}^{m} (\mathbf{w}^T \mathbf{x}_i - y_i)^2$

Use the below table to answer the questions (9) and (10)

x_1	x_2	x_3	y
1	2	3	5
2	3	4	14
3	4	5	29
4	5	6	50
5	6	7	77

Model A : $f_1(x_1, x_2, x_3) = 2x_1 + 3x_2 + x_3 + 6$
Model B : $f_2(x_1, x_2, x_3) = x_1 + 2x_2 + x_3 - 23$
Model C : $f_3(x_1, x_2, x_3) = 6x_1 + 6x_2 + 6x_3 - 37$

9) Which of the following model will yield the least error? [Hint : Use the sum squared error for each model to calculate the error produced by the respective model.](MCQ) *1 point*

x_1	x_2	x_3	y
1	2	3	5
2	3	4	14
3	4	5	29
4	5	6	50
5	6	7	77

Model A : $f_1(x_1, x_2, x_3) = 2x_1 + 3x_2 + x_3 + 6$
Model B : $f_2(x_1, x_2, x_3) = x_1 + 2x_2 + x_3 - 23$
Model C : $f_3(x_1, x_2, x_3) = 6x_1 + 6x_2 + 6x_3 - 37$

9) Which of the following model will yield the least error? [Hint : Use the sum squared error for each model to calculate the error produced by the respective model.](MCQ) *1 point*

- Model A
- Model B
- Model C

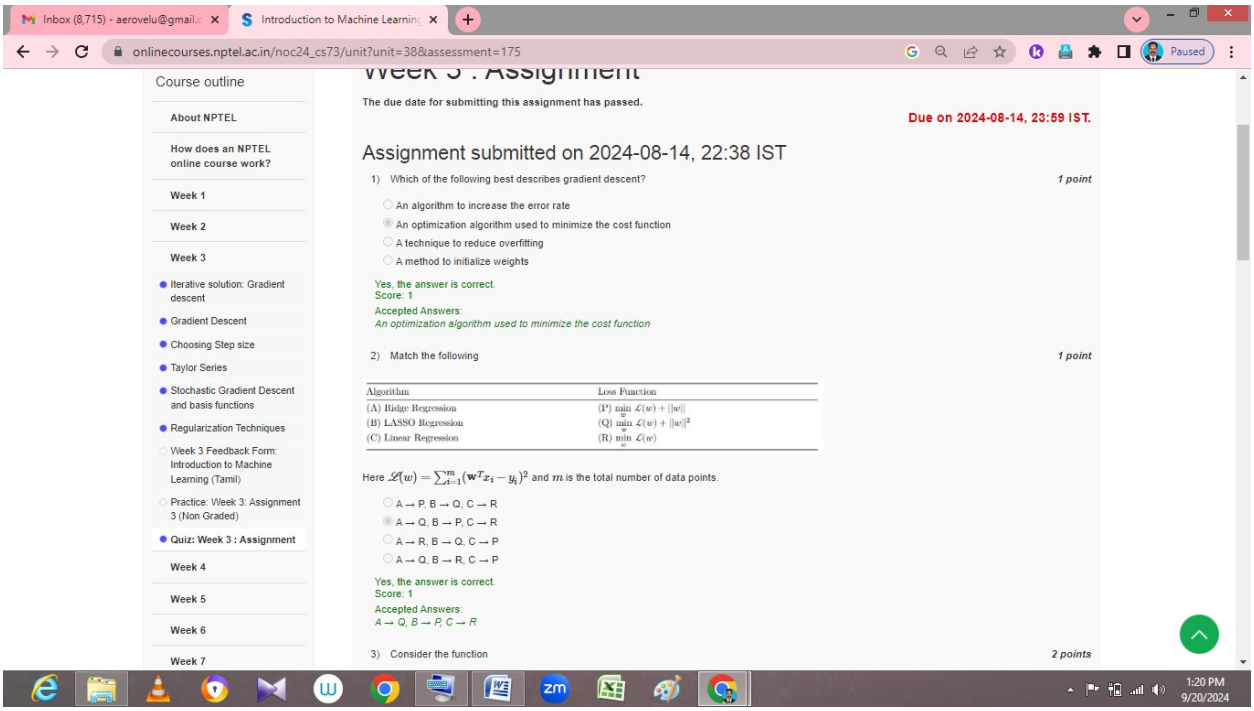
No, the answer is incorrect.
Score: 0
Accepted Answers:
Model C

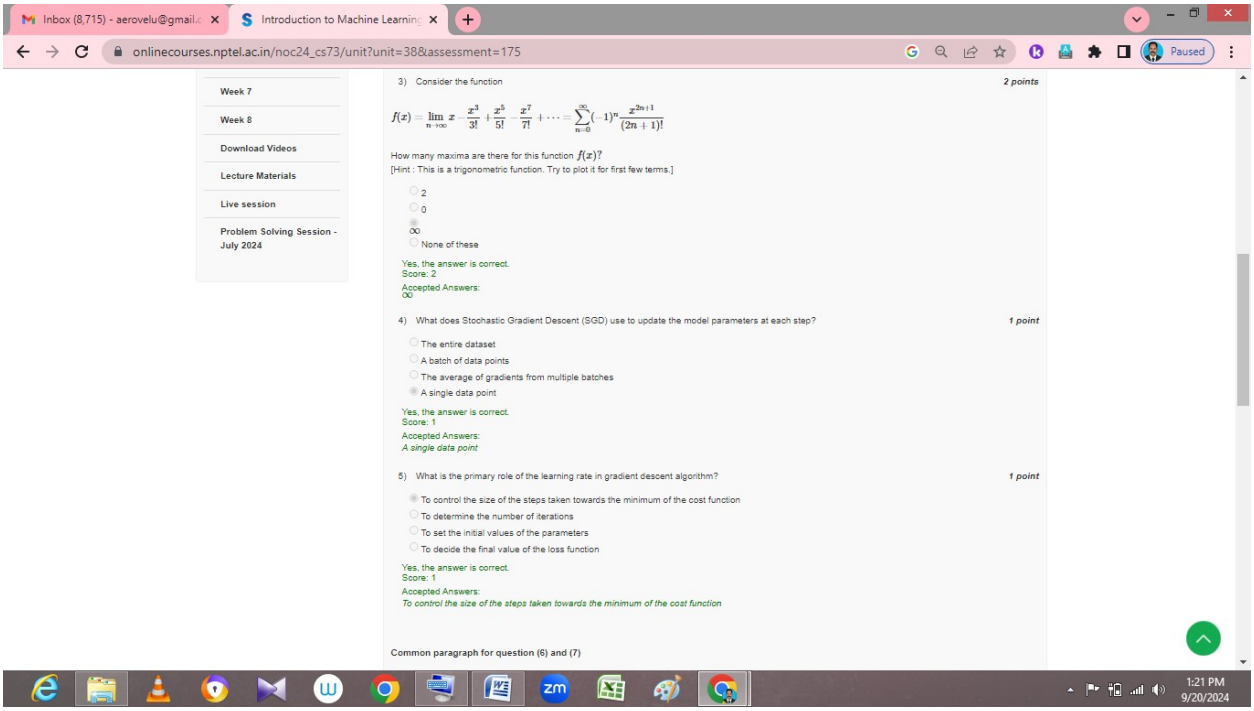
10) Compute the sum squared error of all the three models and enter the sum? Answer with 2 Decimal places. For eg. if the answer is 0.1234, type only 0.12. (NAT)

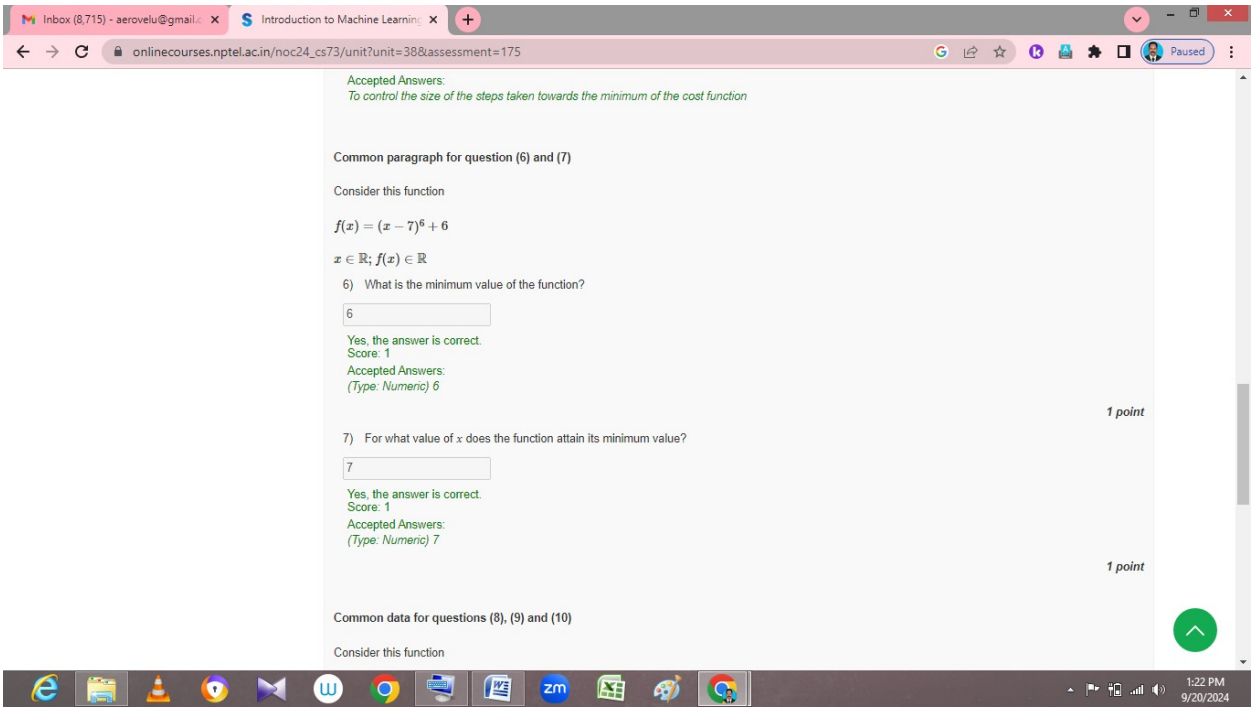
6372

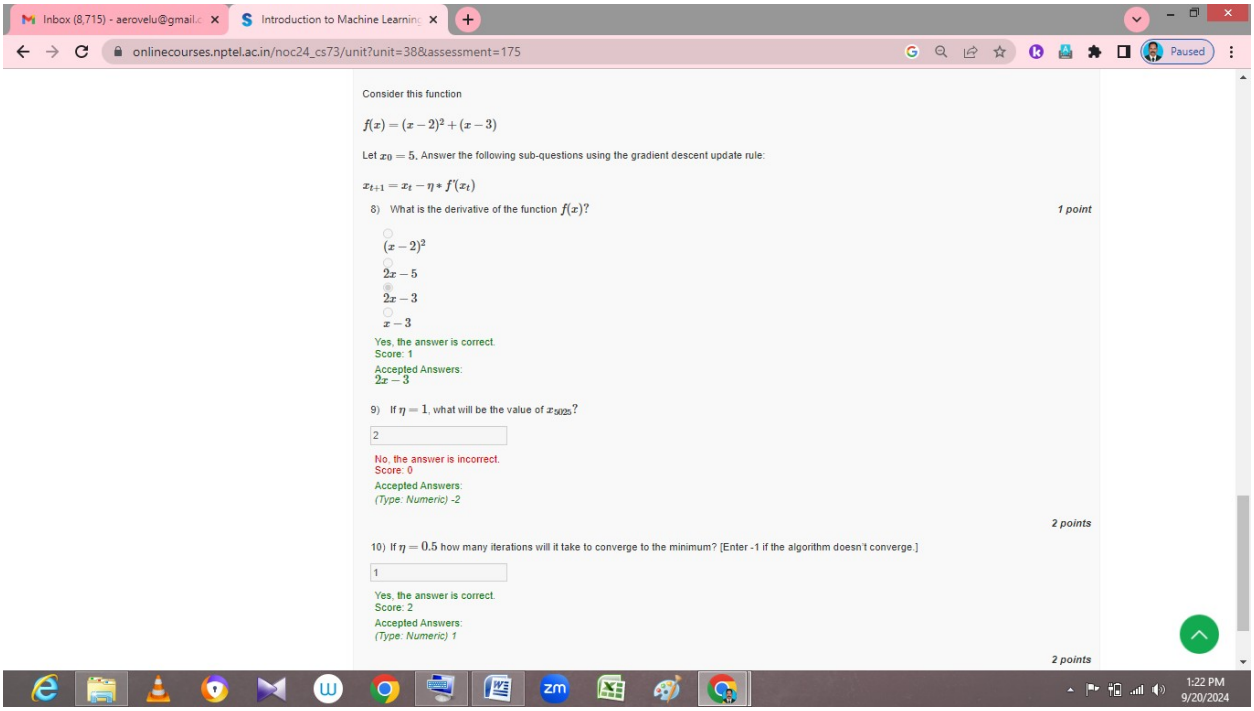
No, the answer is incorrect.
Score: 0
Accepted Answers:
(Type: Numeric) 12778

1 point


Inbox (8,715) - aerovelu@gmail.c
Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=38&assessment=175
Paused
Course outline
About NPTEL
How does an NPTEL online course work?
Week 1
Week 2
Week 3
Iterative solution: Gradient descent
Gradient Descent
Choosing Step size
Taylor Series
Stochastic Gradient Descent and basis functions
Regularization Techniques
Week 3 Feedback Form: Introduction to Machine Learning (Tamil)
Practice: Week 3: Assignment 3 (Non Graded)
Quiz: Week 3 : Assignment
Week 4
Week 5
Week 6
Week 7
Week 3 : Assignment
The due date for submitting this assignment has passed.
Due on 2024-08-14, 23:59 IST.
Assignment submitted on 2024-08-14, 22:38 IST
1) Which of the following best describes gradient descent? 1 point
An algorithm to increase the error rate
An optimization algorithm used to minimize the cost function
A technique to reduce overfitting
A method to initialize weights
Yes, the answer is correct.
Score: 1
Accepted Answers:
An optimization algorithm used to minimize the cost function
2) Match the following 1 point
Algorithm | Loss Function
(A) Ridge Regression | (P) $\min_w \mathcal{L}(w) + \|w\|$
(B) LASSO Regression | (Q) $\min_w \mathcal{L}(w) + \|w\|^2$
(C) Linear Regression | (R) $\min_w \mathcal{L}(w)$
Here $\mathcal{L}(w) = \sum_{i=1}^{m} (\mathbf{w}^T x_i - y_i)^2$ and m is the total number of data points.
A → P, B → Q, C → R
A → Q, B → P, C → R
A → R, B → Q, C → P
A → Q, B → R, C → P
Yes, the answer is correct.
Score: 1
Accepted Answers:
A → Q, B → P, C → R
3) Consider the function 2 points
1:20 PM
9/20/2024


Inbox (8,715) - aerovelu@gmail.c
Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=38&assessment=175
Paused
Week 7
Week 8
Download Videos
Lecture Materials
Live session
Problem Solving Session - July 2024
3) Consider the function 2 points
$f(x) = \lim_{n\to\infty} x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$
How many maxima are there for this function $f(x)$?
[Hint : This is a trigonometric function. Try to plot it for first few terms.]
2
0
∞
None of these
Yes, the answer is correct.
Score: 2
Accepted Answers:
∞
4) What does Stochastic Gradient Descent (SGD) use to update the model parameters at each step? 1 point
The entire dataset
A batch of data points
The average of gradients from multiple batches
A single data point
Yes, the answer is correct.
Score: 1
Accepted Answers:
A single data point
5) What is the primary role of the learning rate in gradient descent algorithm? 1 point
To control the size of the steps taken towards the minimum of the cost function
To determine the number of iterations
To set the initial values of the parameters
To decide the final value of the loss function
Yes, the answer is correct.
Score: 1
Accepted Answers:
To control the size of the steps taken towards the minimum of the cost function
Common paragraph for question (6) and (7)
1:21 PM
9/20/2024


Inbox (8,715) - aerovelu@gmail.c
Introduction to Machine Learnin
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=38&assessment=175
Paused
Accepted Answers:
To control the size of the steps taken towards the minimum of the cost function
Common paragraph for question (6) and (7)
Consider this function
$f(x) = (x-7)^6 + 6$
$x \in \mathbb{R}; f(x) \in \mathbb{R}$
6) What is the minimum value of the function?
6
Yes, the answer is correct.
Score: 1
Accepted Answers:
(Type: Numeric) 6
1 point
7) For what value of x does the function attain its minimum value?
7
Yes, the answer is correct.
Score: 1
Accepted Answers:
(Type: Numeric) 7
1 point
Common data for questions (8), (9) and (10)
Consider this function
1:22 PM
9/20/2024


Inbox (8,715) - aerovelu@gmail.c
Introduction to Machine Learnin
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=38&assessment=175
Paused
Consider this function
$f(x) = (x-2)^2 + (x-3)$
Let $x_0 = 5$, Answer the following sub-questions using the gradient descent update rule:
$x_{t+1} = x_t - \eta * f'(x_t)$
8) What is the derivative of the function $f(x)$?
1 point
$(x-2)^2$
$2x-5$
$2x-3$
$x-3$
Yes, the answer is correct.
Score: 1
Accepted Answers:
$2x-3$
9) If $\eta = 1$, what will be the value of x_{5025}?
2
No, the answer is incorrect.
Score: 0
Accepted Answers:
(Type: Numeric) -2
2 points
10) If $\eta = 0.5$ how many iterations will it take to converge to the minimum? [Enter -1 if the algorithm doesn't converge.]
1
Yes, the answer is correct.
Score: 2
Accepted Answers:
(Type: Numeric) 1
2 points
1:22 PM
9/20/2024

Week 4 : Assignment

The due date for submitting this assignment has passed.

Due on 2024-08-21, 23:59 IST.

As per our records you have not submitted this assignment.

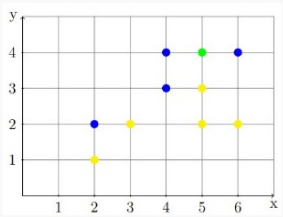
1) In the context of machine learning, which of the following statements best describes the concept of feature independence in a Naive Bayes classifier? ***1 point***

- Two features X_1 and X_2 are independent if $P(X_1|X_2) = P(X_2)$.
- Two features X_1 and X_2 are independent if $P(X_1 \cap X_2) = P(X_1) + P(X_2)$.
- Two features X_1 and X_2 are independent if $P(X_1|X_2) = P(X_1)$.
- Two features X_1 and X_2 are independent if $P(X_1 \cup X_2) = P(X_1) \cdot P(X_2)$

No, the answer is incorrect.
Score: 0

Accepted Answers:
Two features X_1 and X_2 are independent if $P(X_1|X_2) = P(X_1)$.

2) Consider the following data points ***2 points***



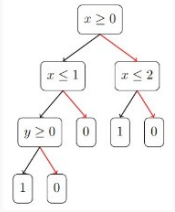
The green point denotes the test data point. Suppose we use KNN algorithm with k=5.
The test point would turn into

- Yellow
- Blue
- Either Blue or Yellow

No, the answer is incorrect.
Score: 0

Accepted Answers:
Blue

3) Consider the following decision tree for a binary classification problem. The dataset lies in $\mathbb{R}$. Here the black edges denote "Yes" and red edges denote "No" ***0 points***



Given a coordinate (x, y) which is classified with the label 0, what are the possible value/values of (x, y)?

- 2, 1
- 1, −1
- 1, 1
- 2, 5

No, the answer is incorrect.
Score: 0

Accepted Answers:
1, −1

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=46&assessment=179

Score: 0

Accepted Answers:
1, −1

4) Which of the following statements about the K-Nearest Neighbors (KNN) algorithm is TRUE? **1 point**

- The KNN algorithm requires a model training phase where it learns weights for features.
- The KNN algorithm's performance is independent of the value of k.
- The KNN algorithm's decision boundaries are always linear.
- The KNN algorithm's classification decision is based on the majority class of the k nearest neighbors.

No, the answer is incorrect.
Score: 0

Accepted Answers:
The KNN algorithm's classification decision is based on the majority class of the k nearest neighbors.

5) In a machine learning experiment, you are given a dataset that is divided into two parts: 80 % of the data is used for training, and the remaining is used for validation. During the training phase, the model achieved an accuracy of 0.8. It is known that 640 data points were classified correctly in the training set. Find the number of data points in the validation set.

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Numeric) 200

2 points

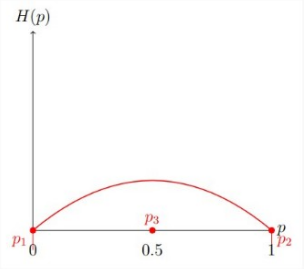
6) Given the graph where the $x-$ **axis** represents probability p and the $y-$ **axis** represents the corresponding entropy $H(p)$ At which point is the measure of impurity at its maximum? **1 point**

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=46&assessment=179

Accepted Answers:
(Type: Numeric) 200

2 points

6) Given the graph where the $x-$ **axis** represents probability p and the $y-$ **axis** represents the corresponding entropy $H(p)$ At which point is the measure of impurity at its maximum? **1 point**



- p_1
- p_3
- p_2

No, the answer is incorrect.
Score: 0

Accepted Answers:
p_3

7) In the context of Naive Bayes classifiers, consider a dataset with features $X_1, X_2, \ldots, X_n$ and class labels Y. Which of the following statements best describes the concept of class conditional independence? **1 point**

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learnin | onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=46&assessment=179

7) In the context of Naive Bayes classifiers, consider a dataset with features $X_1, X_2, \ldots, X_n$ and class labels Y. Which of the following statements best describes the concept of class conditional independence? **1 point**

- The features $X_1, X_2, \ldots, X_n$ are assumed to be independent of each other given the class label Y.
- The features $X_1, X_2, \ldots, X_n$ are assumed to be independent of the class label Y.
- The class label Y is assumed to be independent of the features $X_1, X_2, \ldots, X_n$.
- The features $X_1, X_2, \ldots, X_n$ are assumed to be dependent on each other given the class label Y.

No, the answer is incorrect.
Score: 0

Accepted Answers:
The features $X_1, X_2, \ldots, X_n$ are assumed to be independent of each other given the class label Y.

8) Given a dataset containing emails, where a label of 1 indicates spam and 0 indicates non-spam, and there are two labels, 1 and 0. Calculate the probability

$$P(y = 1) \cdot P(x|y = 1)$$

for the feature vector $x = [1, 1, 0]$.

Given the following probabilities:

$$\hat{P} = 0.6, \ \hat{P}_1^1 = 0.3, \ \hat{P}_2^1 = 0.4, \ \hat{P}_3^1 = 0.7$$

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Range) 0.02,0.03

1 point

9) You are using the K-Nearest Neighbors (KNN) algorithm to classify a new data point. Given the following training data points and their **2 points**

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learnin | onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=46&assessment=179

Accepted Answers:
(Type: Range) 0.02,0.03

1 point

9) You are using the K-Nearest Neighbors (KNN) algorithm to classify a new data point. Given the following training data points and their corresponding class labels: **2 points**

Point	Coordinates (x, y)	Class Label
A	(1, 1)	0
B	(2, 2)	0
C	(3, 3)	1
D	(6, 5)	1
E	(7, 7)	1

Using $k = 3$, classify the new data point P with coordinates (4, 4). Assume Euclidean distance is used to measure the distance between points.

- Class 0
- Class 1
- Unable to determine
- Depends on the value of k

No, the answer is incorrect.
Score: 0

Accepted Answers:
Class 1

10) In the context of decision trees, which of the following statements best describes the concept of information gain? **1 point**

- Information gain measures the total number of samples in each node after the split.
- Information gain is the difference between the number of positive and negative samples in a node.
- Information gain is the sum of the entropies of the child nodes after the split.
- Information gain is the difference between the entropy of the parent node and the weighted average entropy of the child nodes.

No, the answer is incorrect.
Score: 0

Accepted Answers:
Information gain is the difference between the entropy of the parent node and the weighted average entropy of the child nodes.

1) The Perceptron algorithm can inherently handle multi-class classification problems without any modifications. *1 point*

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
False

2) Which of the following statements about logistic regression is true? *1 point*

- Logistic regression can only be used for regression problems.
- Logistic regression can be used for both binary and multi-class classification problems.
- Logistic regression does not use a linear decision boundary.
- Logistic regression assumes that the input features are normally distributed.

Yes, the answer is correct.
Score: 1
Accepted Answers:
Logistic regression can be used for both binary and multi-class classification problems.

3) Given the following set of points in a 2D space: *2 points*

- Class A : $(2, 4), (3, 5), (4, 7)$
- Class B : $(1, 1), (2, 1), (3, 2)$

Which of the following statements is true?

- The points are linearly separable and can be separated by the line $y = x + 1$.
- The points are not linearly separable.
- The points are linearly separable and can be separated by the line $y = 2x + 2$.
- Adding the point (4,2) which belongs to class A makes the data points non linearly separable.

No, the answer is incorrect.
Score: 0
Accepted Answers:
The points are linearly separable and can be separated by the line $y = x + 1$.

4) A logistic regression model has weights $w = [0.3, -0.6]$. Given an input vector $x = [2, 3]$: What is the predicted class label $\hat{y}$ using the decision rule:

The points are linearly separable and can be separated by the line $y = x + 1$.

4) A logistic regression model has weights $w = [0.3, -0.6]$. Given an input vector $x = [2, 3]$: What is the predicted class label $\hat{y}$ using the decision rule:

$$\hat{y} = \begin{cases} 1 & \text{if } P(y=1) \geq 0.5 \\ 0 & \text{if } P(y=1) < 0.5 \end{cases}$$

0

Yes, the answer is correct.
Score: 2
Accepted Answers:
(Type: Numeric) 0

2 points

5) Consider a Perceptron with initial weights $\mathbf{w} = [0.2, 0.5, 0.3]$. The Perceptron receives an input vector $\mathbf{x} = [1, -2, 1]$ and makes a prediction of $\hat{y} = 0$, but the true label is $y = 1$. Update the weight vector and enter the sum of the elements in weight vector.

5.5

No, the answer is incorrect.
Score: 0
Accepted Answers:
(Type: Numeric) 1

1 point

6) Which of the following statements about linear separability is correct in case of binary classification? *1 point*

- Linear separability means that a dataset can be perfectly divided into three or more classes using a linear boundary.
- Linear separability means that a dataset can be perfectly divided into two classes using a curved boundary.
- Linear separability means that a dataset can be perfectly divided into two classes using a linear boundary.
- Linear separability means that a dataset cannot be divided using a linear boundary.

Yes, the answer is correct.
Score: 1
Accepted Answers:
Linear separability means that a dataset can be perfectly divided into two classes using a linear boundary.

7) Given the following joint probabilities for a binary classification problem: *2 points*

- $P(x = 0, y = 0) = 0.2$
- $P(x = 1, y = 0) = 0.3$
- $P(x = 0, y = 1) = 0.1$

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learning | onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=60&assessment=180

Yes, the answer is correct.
Score: 1
Accepted Answers:
Linear separability means that a dataset can be perfectly divided into two classes using a linear boundary.

7) Given the following joint probabilities for a binary classification problem: **2 points**

- $P(x=0, y=0) = 0.2$
- $P(x=1, y=0) = 0.3$
- $P(x=0, y=1) = 0.1$
- $P(x=1, y=1) = 0.4$

Compute the conditional probability $P(y=1|x=1)$.

- 0.6
- 0.4
- 0.91
- 0.571

No, the answer is incorrect.
Score: 0
Accepted Answers:
0.571

8) In the context of maximum likelihood estimation (MLE) in machine learning, which of the following statements is true? **1 point**

- MLE aims to find the model parameters that minimize the sum of squared errors.
- The likelihood function is defined as the product of the probabilities of the observed data points given the model parameters.
- MLE is only applicable to linear regression models.
- The likelihood function and the posterior distribution are the same in Bayesian statistics.

Yes, the answer is correct.
Score: 1
Accepted Answers:
The likelihood function is defined as the product of the probabilities of the observed data points given the model parameters.

1:26 PM 9/20/2024

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learning | onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=69&assessment=181

online course work?

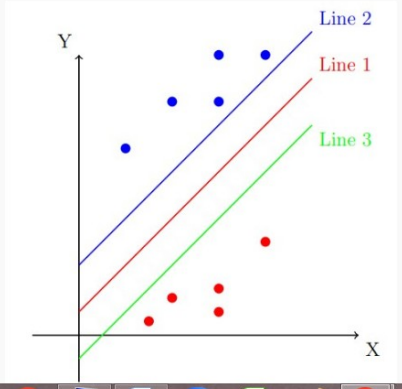
- Week 1
- Week 2
- Week 3
- Week 4
- Week 5
- Week 6
  - Support Vector Machine
  - Optimizing weights
  - Handling Outliers
  - Dual Formulation
  - Kernel formulation
  - Week 6 Feedback Form: Introduction to Machine Learning (Tamil)
  - Practice: Week 6: Assignment 6 (Non Graded)
  - Quiz: Week 6 : Assignment
- Week 7
- Week 8
- Download Videos
- Lecture Materials
- Live session

1) What is the major goal of a Support Vector Machine (SVM)? **1 point**

- To minimize the distance between all data points in the dataset.
- To reduce the dimensionality of the dataset.
- To increase the number of support vectors in the model.
- To maximize the margin between different classes in the data.

Yes, the answer is correct.
Score: 1
Accepted Answers:
To maximize the margin between different classes in the data.

2) Consider the following XY plane with linearly separable points. Three possible decision boundaries are shown. **1 point**



1:27 PM 9/20/2024

Live session

Problem Solving Session - July 2024

Which line is more likely to be chosen by a good SVM model?

- Line 1
- Line 2
- Line 3

Yes, the answer is correct.
Score: 1
Accepted Answers:
Line 1

3) Consider these three weight vectors: ***2 points***

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 2 \\ 8 \\ 5 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix}$$

All these vectors are linearly separable for a given dataset. If you are running a Support Vector Machine (SVM) algorithm, which weight vector would you choose?

- $\mathbf{w}_1$
- $\mathbf{w}_2$
- $\mathbf{w}_3$

Yes, the answer is correct.
Score: 2
Accepted Answers:
$\mathbf{w}_3$

4) What is the objective function used to optimize a Hard Margin Support Vector Machine (SVM) model? ***1 point***

- $\min_{\mathbf{w}} \frac{2}{||\mathbf{w}||^2} \text{ s.t} (\mathbf{w}^T x_i) y_i \geq 1 \; \forall i$

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=69&assessment=181
Paused

4) What is the objective function used to optimize a Hard Margin Support Vector Machine (SVM) model? ***1 point***

- $\min_{\mathbf{w}} \frac{2}{||\mathbf{w}||^2} \text{ s.t} (\mathbf{w}^T x_i) y_i \geq 1 \; \forall i$
- $\min_{\mathbf{w}} \frac{||\mathbf{w}||^2}{2} \text{ s.t} (\mathbf{w}^T x_i) y_i \geq 1 \; \forall i$
- $\max_{\mathbf{w}} \frac{2}{||\mathbf{w}||^4} \text{ s.t} (\mathbf{w}^T x_i) y_i \geq 1 \; \forall i$
- $\max_{\mathbf{w}} \frac{||\mathbf{w}||^2}{2} \text{ s.t} (\mathbf{w}^T x_i) y_i \geq 1 \; \forall i$

Yes, the answer is correct.
Score: 1
Accepted Answers:
$\min_{\mathbf{w}} \frac{||\mathbf{w}||^2}{2} \text{ s.t} (\mathbf{w}^T x_i) y_i \geq 1 \; \forall i$

5) For the below primal and dual problems to be equivalent what kind of functions should f and g should be? ***2 points***

$$\min_w \left[\max_{\lambda \geq 0} f(x) + \lambda g(w) \right] \equiv \max_{\lambda \geq 0} \left[\min_w f(x) + \lambda g(w) \right]$$

- f should be convex function and g can be any function.
- f can be any function but g should be a convex function
- f and g must be convex functions
- f and g can be any functions.

Yes, the answer is correct.
Score: 2
Accepted Answers:
f and g must be convex functions

6) Consider the following 1-dimensional dataset: ***1 point***

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learnin | onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=69&assessment=181

6) Consider the following 1-dimensional dataset: *1 point*

x	y
-3	-1
-1	1
1	1
2	-1

(Note: x is the feature and is the output)

A Hard-margin SVM, with no kernels, can correctly classify the above given data

- True
- False

Yes, the answer is correct.
Score: 1
Accepted Answers:
False

7) In the lecture, we discussed the formulation for a soft margin Support Vector Machine (SVM): *2 points*

$$\min_{\mathbf{w}} \frac{1}{2}||\mathbf{w}^2||^2 + \lambda \sum_{i=1}^{n} e_i$$

where there are n data points. What does the term e_i represent?

- The cost for misclassification.
- The margin of separation between the classes.
- The regularization parameter controlling the trade-off between margin width and classification error.
- The weight vector that defines the decision boundary.

Yes, the answer is correct.
Score: 2
Accepted Answers:
The cost for misclassification.

1:29 PM 9/20/2024

Inbox (8,715) - aerovelu@gmail.c | Introduction to Machine Learnin | onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=76&assessment=182

Assignment submitted on 2024-09-11, 21:43 IST

1) You have 2 binary Classifiers I and II. Classifier I has a test accuracy of 0% and Classifier II has test accuracy of 50%. Suppose you are allowed to invert/negate (if the classifier predicts class 0, take the output as class 1 and vice-versa) the outputs of the classifier. Which classifier will perform better? *1 point*

- Classifier II
- Both are bad
- Can not be determined
- Classifier I

Yes, the answer is correct.
Score: 1
Accepted Answers:
Classifier I

2) Why is creating diverse weak learners important in ensemble techniques like bagging and boosting? *1 point*

- Diversity helps in reducing the number of weak learners required for effective ensembling
- Creating diverse learners makes the ensemble technique computationally efficient
- Diverse learners are less prone to overfitting and can capture different patterns in the data
- None of the above

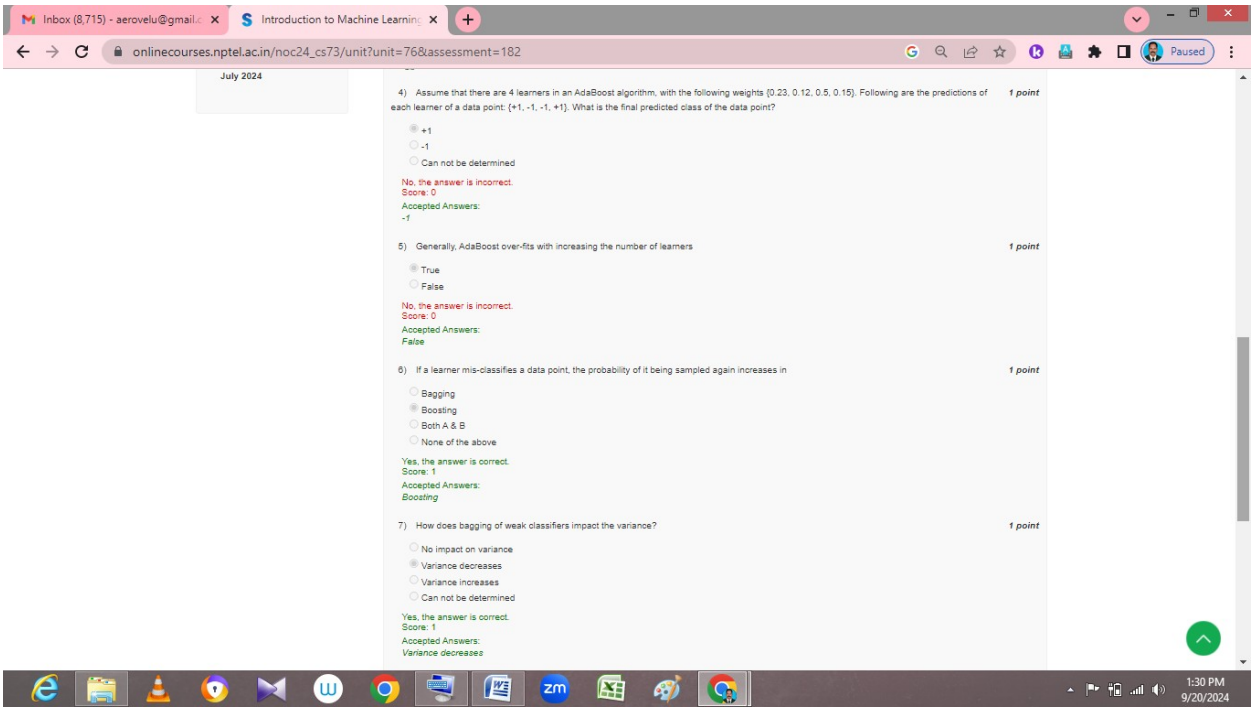
Yes, the answer is correct.
Score: 1
Accepted Answers:
Diverse learners are less prone to overfitting and can capture different patterns in the data

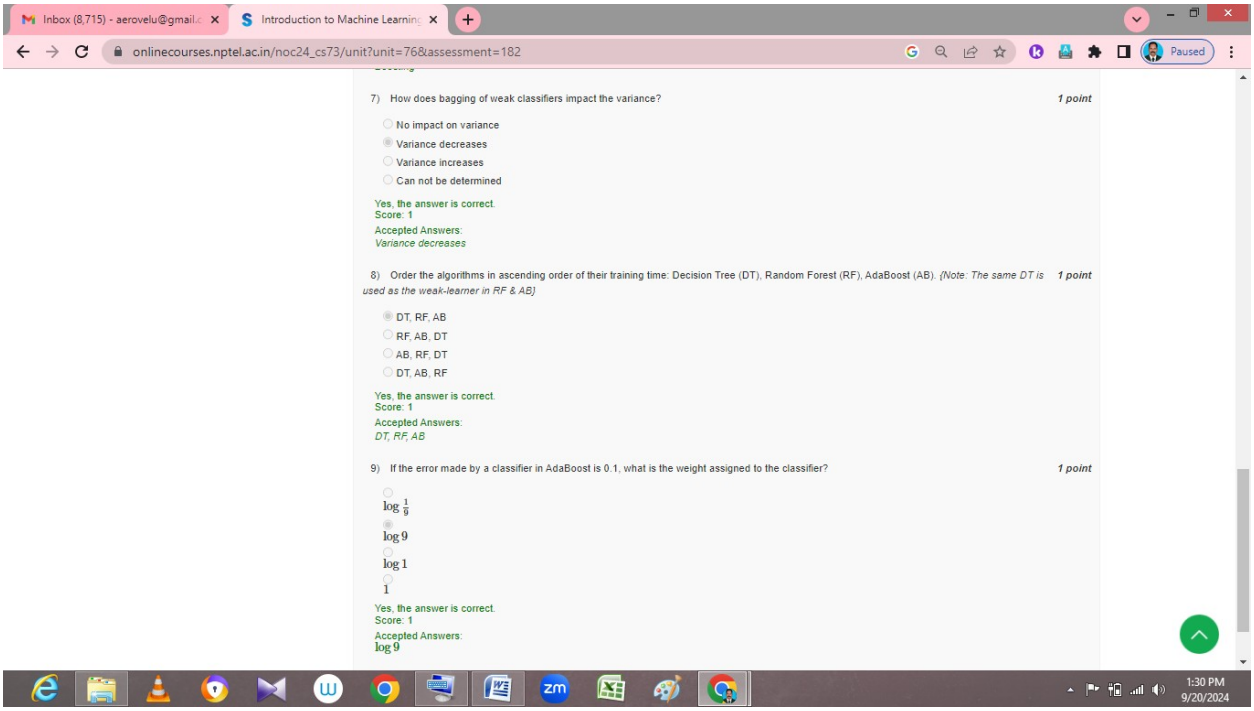
3) In a bag, there are 2 red flags, 3 blue flags, and 1 green flags. If you randomly draw 3 flags from the bag with replacement, what is the probability of them all being red? *1 point*

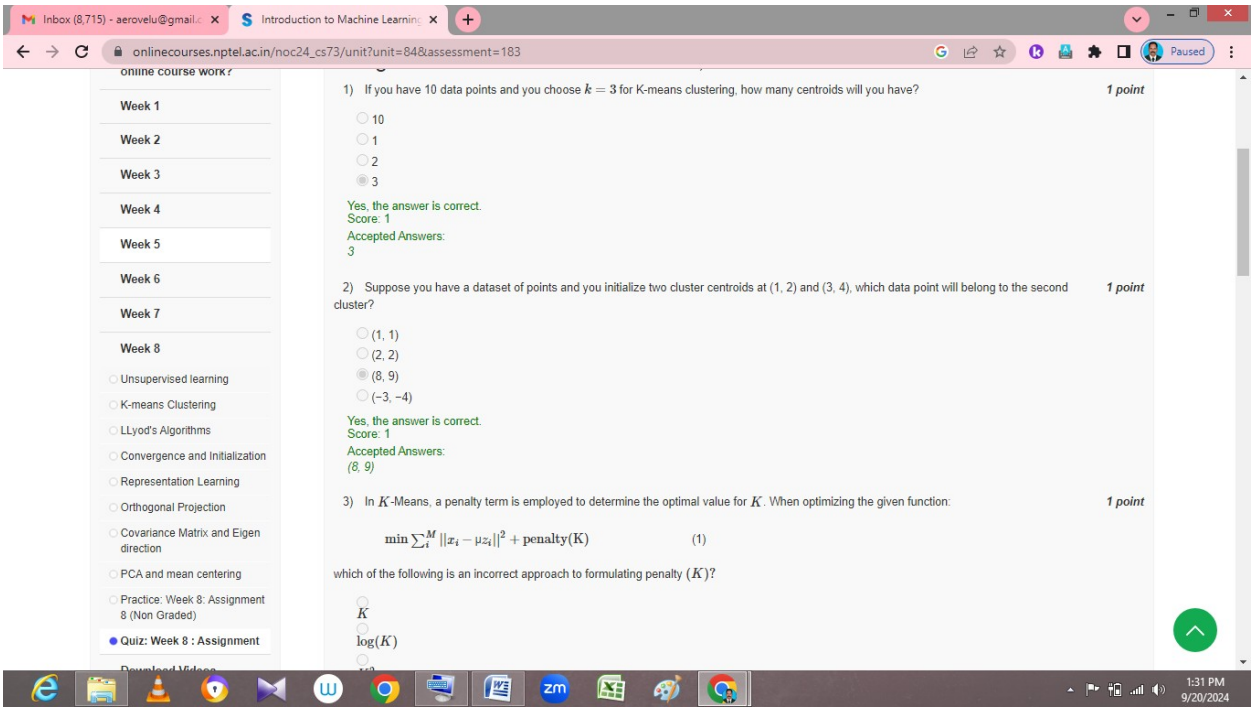
- $\frac{1}{15}$
- $\frac{2}{15}$
- $\frac{1}{30}$
- $\frac{0}{15}$

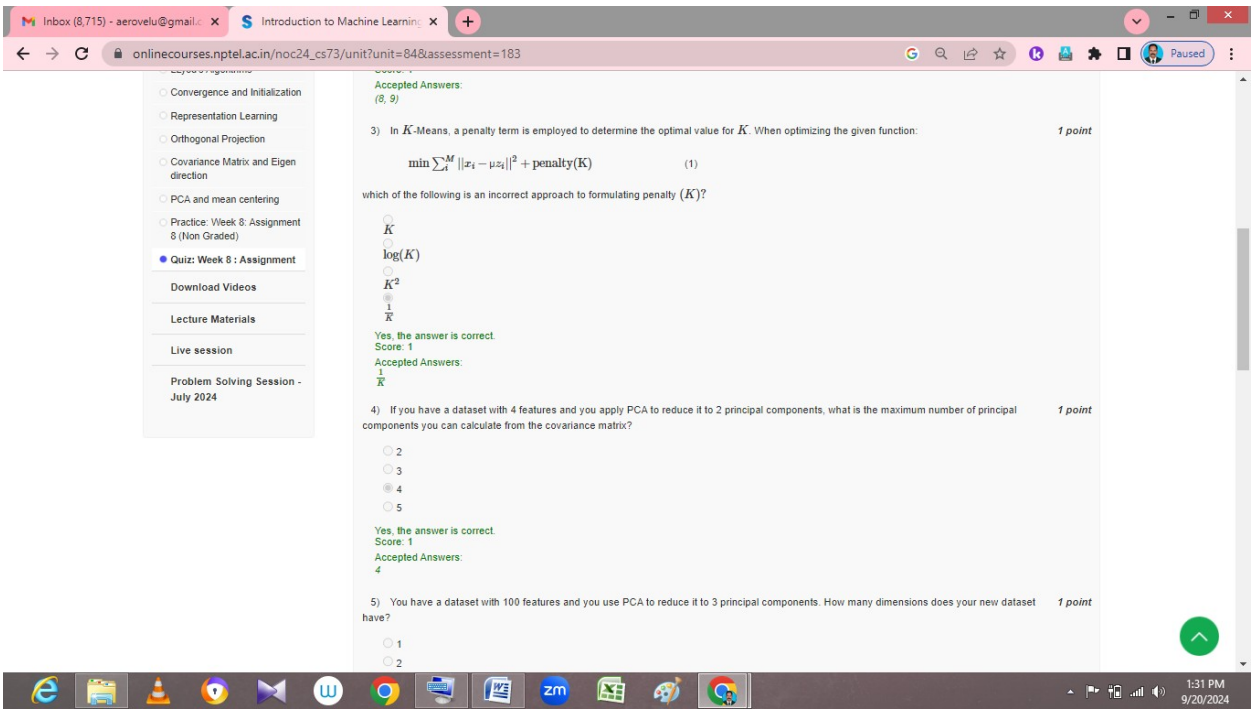
Yes, the answer is correct.
Score: 1
Accepted Answers:
$\frac{0}{15}$

1:30 PM 9/20/2024


Inbox (8,715) - aerovelu@gmail.c
Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=76&assessment=182
Paused
July 2024
4) Assume that there are 4 learners in an AdaBoost algorithm, with the following weights {0.23, 0.12, 0.5, 0.15}. Following are the predictions of each learner of a data point: {+1, -1, -1, +1}. What is the final predicted class of the data point? 1 point
+1
-1
Can not be determined
No, the answer is incorrect.
Score: 0
Accepted Answers:
-1
5) Generally, AdaBoost over-fits with increasing the number of learners 1 point
True
False
No, the answer is incorrect.
Score: 0
Accepted Answers:
False
6) If a learner mis-classifies a data point, the probability of it being sampled again increases in 1 point
Bagging
Boosting
Both A & B
None of the above
Yes, the answer is correct.
Score: 1
Accepted Answers:
Boosting
7) How does bagging of weak classifiers impact the variance? 1 point
No impact on variance
Variance decreases
Variance increases
Can not be determined
Yes, the answer is correct.
Score: 1
Accepted Answers:
Variance decreases
1:30 PM
9/20/2024


Inbox (8,715) - aerovelu@gmail.c
Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=76&assessment=182
Paused
7) How does bagging of weak classifiers impact the variance? 1 point
No impact on variance
Variance decreases
Variance increases
Can not be determined
Yes, the answer is correct.
Score: 1
Accepted Answers:
Variance decreases
8) Order the algorithms in ascending order of their training time: Decision Tree (DT), Random Forest (RF), AdaBoost (AB). {Note: The same DT is used as the weak-learner in RF & AB} 1 point
DT, RF, AB
RF, AB, DT
AB, RF, DT
DT, AB, RF
Yes, the answer is correct.
Score: 1
Accepted Answers:
DT, RF, AB
9) If the error made by a classifier in AdaBoost is 0.1, what is the weight assigned to the classifier? 1 point
$\log \frac{1}{9}$
$\log 9$
$\log 1$
1
Yes, the answer is correct.
Score: 1
Accepted Answers:
$\log 9$
1:30 PM
9/20/2024


Inbox (8,715) - aerovelu@gmail.c
Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=84&assessment=183
Paused
online course work?
Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Unsupervised learning
K-means Clustering
LLyod's Algorithms
Convergence and Initialization
Representation Learning
Orthogonal Projection
Covariance Matrix and Eigen direction
PCA and mean centering
Practice: Week 8: Assignment 8 (Non Graded)
Quiz: Week 8 : Assignment
1) If you have 10 data points and you choose $k = 3$ for K-means clustering, how many centroids will you have? 1 point
10
1
2
3
Yes, the answer is correct.
Score: 1
Accepted Answers:
3
2) Suppose you have a dataset of points and you initialize two cluster centroids at (1, 2) and (3, 4), which data point will belong to the second cluster? 1 point
(1, 1)
(2, 2)
(8, 9)
(−3, −4)
Yes, the answer is correct.
Score: 1
Accepted Answers:
(8, 9)
3) In K-Means, a penalty term is employed to determine the optimal value for K. When optimizing the given function: 1 point
$\min \sum_i^M ||x_i - \mu_{z_i}||^2 + \text{penalty(K)}$ (1)
which of the following is an incorrect approach to formulating penalty (K)?
K
$\log(K)$
1:31 PM
9/20/2024


Inbox (8,715) - aerovelu@gmail.c
Introduction to Machine Learning
onlinecourses.nptel.ac.in/noc24_cs73/unit?unit=84&assessment=183
Paused
Convergence and Initialization
Representation Learning
Orthogonal Projection
Covariance Matrix and Eigen direction
PCA and mean centering
Practice: Week 8: Assignment 8 (Non Graded)
Quiz: Week 8 : Assignment
Download Videos
Lecture Materials
Live session
Problem Solving Session - July 2024
Accepted Answers:
(8, 9)
3) In K-Means, a penalty term is employed to determine the optimal value for K. When optimizing the given function: 1 point
$\min \sum_i^M ||x_i - \mu_{z_i}||^2 + \text{penalty(K)}$ (1)
which of the following is an incorrect approach to formulating penalty (K)?
K
$\log(K)$
K^2
$\frac{1}{K}$
Yes, the answer is correct.
Score: 1
Accepted Answers:
$\frac{1}{K}$
4) If you have a dataset with 4 features and you apply PCA to reduce it to 2 principal components, what is the maximum number of principal components you can calculate from the covariance matrix? 1 point
2
3
4
5
Yes, the answer is correct.
Score: 1
Accepted Answers:
4
5) You have a dataset with 100 features and you use PCA to reduce it to 3 principal components. How many dimensions does your new dataset have? 1 point
1
2
1:31 PM
9/20/2024

Yes, the answer is correct.
Score: 1
Accepted Answers:
4

5) You have a dataset with 100 features and you use PCA to reduce it to 3 principal components. How many dimensions does your new dataset have? *1 point*

- 1
- 2
- 3
- 100

Yes, the answer is correct.
Score: 1
Accepted Answers:
3

6) Find the correct sequence of statements that reproduce the PCA algorithm *1 point*

1. Compute the eigen values and vectors
2. Project the datapoints in the new feature space
3. Normalize the data points
4. Compute the covariance matrix
5. Select the eigen values correspondingly

- 3, 4, 1, 5, 2
- 3, 2, 1, 4, 5
- 3, 4, 2, 5, 1
- None of them

Yes, the answer is correct.
Score: 1
Accepted Answers:
3, 4, 1, 5, 2

7) Given a line of points denoted by $u = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, and a point $v = \begin{bmatrix} 9 \\ 2 \end{bmatrix}$. What point on the line has the minimum projection error of v on u? *1 point*

- (1.5, 4.5)
- (1.3, 3.9)

7) Given a line of points denoted by $u = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, and a point $v = \begin{bmatrix} 9 \\ 2 \end{bmatrix}$. What point on the line has the minimum projection error of v on u? *1 point*

- (1.5, 4.5)
- (1.3, 3.9)
- (1.1, 3.3)
- (0, 0)

Yes, the answer is correct.
Score: 1
Accepted Answers:
(1.5, 4.5)

Common data for Questions 8 and 9

Following is an incomplete pseudo-code of the Lloyd's algorithm for k-means clustering.

Algorithm 1 Lloyd's algorithm

```
1: Randomly initialize the cluster centers
2: Randomly initialize each data point to a cluster center
3: while convergence do
4:     Reassign each point to the nearest cluster center
5:     Update each cluster center by the Y of its corresponding points
6: end while
```

8) What condition must be checked to ensure the algorithm converges in step 3? *1 point*

- Whether the number of cluster centers starts increasing
- Whether any point has been reassigned to a different center
- Whether the total error is above a threshold
- None of the above

Yes, the answer is correct.
Score: 1
Accepted Answers:
Whether any point has been reassigned to a different center

9) What is Y in step 5? *1 point*

Following is an incomplete pseudo-code of the Lloyd's algorithm for k-means clustering.

Algorithm 1 Lloyd's algorithm

```
1: Randomly initialize the cluster centers
2: Randomly initialize each data point to a cluster center
3: while convergence do
4:     Reassign each point to the nearest cluster center
5:     Update each cluster center by the Y of its corresponding points
6: end while
```

8) What condition must be checked to ensure the algorithm converges in step 3? *1 point*

- Whether the number of cluster centers starts increasing
- Whether any point has been reassigned to a different center
- Whether the total error is above a threshold
- None of the above

Yes, the answer is correct.
Score: 1

Accepted Answers:
Whether any point has been reassigned to a different center

9) What is Y in step 5? *1 point*

- Maximium
- Minimum
- Mean
- Median

Yes, the answer is correct.
Score: 1

Accepted Answers:
Mean