```
code1: actions
code2: context.students.forEach(student=>{
                    if(student.marks < 55)
                        context.below_average.push(student)
                    })
```

6406532330262. ✖

```
code1: mutations
code2: state.students.forEach(student=>{
                    if(student.marks > 50)
                        state.below_average.push(student)
                    })
```

6406532330263. ✔

MLT

Section Id :	64065349266
Section Number :	7
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	17
Number of Questions to be attempted :	17
Section Marks :	100
Display Number Panel :	Yes
Section Negative Marks :	0
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	Yes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653103286
Question Shuffling Allowed :	No

Is Section Default? : null

Question Number : 155 Question Id : 640653697732 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DIPLOMA LEVEL : MACHINE LEARNING TECHNIQUES (COMPUTER BASED EXAM)"

ARE YOU SURE YOU HAVE TO WRITE EXAM FOR THIS SUBJECT?

CROSS CHECK YOUR HALL TICKET TO CONFIRM THE SUBJECTS TO BE WRITTEN.

(IF IT IS NOT THE CORRECT SUBJECT, PLS CHECK THE SECTION AT THE TOP FOR THE SUBJECTS REGISTERED BY YOU)

Options :

6406532330301. ✔ YES

6406532330302. ✖ NO

Sub-Section Number : 2

Sub-Section Id : 640653103287

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 156 Question Id : 640653697733 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the dataset $\mathcal{D} = \{(-1,1),(0,0),(1,1)\}$. What is the first principal component (i.e., the direction corresponding to the largest eigenvalue of the covariance matrix) for the above dataset?

Options :

6406532330303. ✔ $\begin{pmatrix}1\\0\end{pmatrix}$

6406532330304. ✖ $\begin{pmatrix}0\\1\end{pmatrix}$

6406532330305. ✖ $\begin{pmatrix}1\\1\end{pmatrix}$

6406532330306. ✔ $\begin{pmatrix}-1\\0\end{pmatrix}$

Sub-Section Number :	3
Sub-Section Id :	640653103288
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653697734 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (157 to 158)

Question Label : Comprehension

A team was given a dataset $X \in \mathbb{R}^{d\times n}$ where d denotes the number of features and n denotes the number of samples. They found that there are 20 samples in the dataset and each sample contains 100 features. Assume that the datapoints x_5 to x_{20} are all linear combination of linearly independent data points (x_1, x_2, x_3, x_4).

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 157 Question Id : 640653697735 Question Type : MCQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5

Question Label : Multiple Choice Question

Suppose the team applies linear PCA on the dataset and reconstructs the data points with zero error using *k* principal components (directions). For which value of *k* the reconstruction error would become zero?

Options :

6406532330307. ✖ 1

6406532330308. ✖ 2

6406532330309. ✖ 3

6406532330310. ✔ 4

6406532330311. ✖ 10

6406532330312. ✖ 100

Question Number : 158 Question Id : 640653697736 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6 Max. Selectable Options : 0

Question Label : Multiple Select Question

Suppose the team applies kernel PCA by computing the kernel matrix $K = X^T X$ on the dataset. Choose all the correct statements.

Options :

6406532330313. ✔ The principal directions given by the kernel PCA is the same as the one given by linear PCA

6406532330314. ✔

Kernel PCA takes a lesser number of computations than linear PCA to find the principal directions

The eigenvectors of the kernel matrix K are pointing in the same direction as the eigenvectors of the covariance matrix C

6406532330315. ✖

Kernel PCA takes more computations than linear PCA to find the principal directions

6406532330316. ✖

Sub-Section Number : 4

Sub-Section Id : 640653103289

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 159 Question Id : 640653697737 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider Lloyd's algorithm used for k-means clustering and choose the correct statements:

Options :

6406532330317. ✔ K-means algorithm may get stuck at local minima.

6406532330318. ✖ It guarantees finding the optimal clustering(global minima) in every run.

6406532330319. ✔ If the resources are limited and the data set is huge, it will be good to prefer K-means over K-means++.

6406532330320. ✖ In practice, k should be as large as possible.

Sub-Section Number : 5

Sub-Section Id : 640653103290

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 160 Question Id : 640653697738 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6

Question Label : Short Answer Question

We wish to fit a GMM with $K = 2$ for a dataset having 4 points. At the beginning of the t^{th} time step of the EM algorithm, we have $\theta^{(t)}$ as follows:

$$\pi_1 = 0.4, \pi_2 = 0.6$$
$$\mu_1 = 2, \sigma_1^2 = 1$$
$$\mu_2 = 3, \sigma_2^2 = 1$$

The density of the points given a particular mixture is given to you for all four points. f is the density of a Gaussian.

x_i	$f(x_i\|z_i = 1)$	$f(x_i\|z_i = 2)$
1	0.242	0.054
2	0.399	0.242
3	0.242	0.399
4	0.054	0.242

What is the value of λ_k^i for $i = 1$ and $k = 2$ after the E-step? Enter your answer correct to two decimal places

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.22 to 0.28

Question Number : 161 Question Id : 640653697739 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6

Question Label : Short Answer Question

Consider a linearly independent set of data points

$$X = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

and the corresponding label $y = \begin{bmatrix} 0.5 \\ 0 \\ -0.5 \end{bmatrix}$. Suppose we fit the data points using a simple linear regression model that minimizes squared error loss $L(w) = \sum(w^T x_i - y_i)^2$. Compute the value of loss at $w = w^*$, where $w^* = (XX^T)^{-1}Xy$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0

Sub-Section Number :	6
Sub-Section Id :	640653103291
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653697740 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (162 to 163)

Question Label : Comprehension

Consider the following dataset with 3 features and 5 data points:

$\mathbf{Feat_1}$	$\mathbf{Feat_2}$	$\mathbf{Feat_3}$	Y
1	2	3	10.5
2	3	1	$d1$
3	1	2	9.5
4	4	5	$d2$
5	5	4	$d3$

$(d_1, d_2, d_3 \in \mathbb{R})$

You decide to train a linear regression model on this dataset.
After training, you obtain the following weight vector

$$\mathbf{w} = \begin{bmatrix} 1.2 \\ 0.5 \\ 2.5 \end{bmatrix}$$

Now, you decide to introduce L2 regularization to your model.
You train the model again with a regularization parameter (λ)
set to 0.5. The new regularized weight vector is:

$$\mathbf{w}_{\text{regularized}} = \begin{bmatrix} 0.9 \\ 0.4 \\ 1.5 \end{bmatrix}$$

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 162 Question Id : 640653697741 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

Based on the given data, using the regularized model, predict the target variable (Y) for the following data point.

$$x_{\text{new}}: \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}$$

Options :

6406532330323. ✔ 7.9

6406532330324. ✖ 12.2

6406532330325. ✖ 11.9

6406532330326. ✖ None of these

Question Number : 163 Question Id : 640653697742 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

If you further reduce the regularization parameter (λ) to zero. What would be the new prediction for the target variable?

Options :

6406532330327. ✖ 7.9

6406532330328. ✖ 12.2

6406532330329. ✔ 11.9

6406532330330. ✖ None of these

Sub-Section Number :	7
Sub-Section Id :	640653103292
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 164 Question Id : 640653697743 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

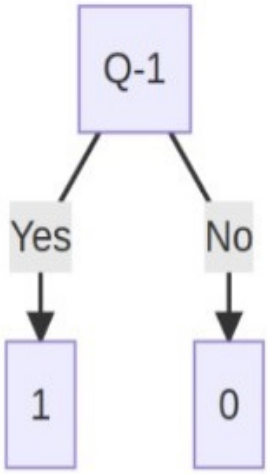
Correct Marks : 6

Question Label : Short Answer Question

Consider the following training dataset for a binary classification task:

X	y
6	1
20	0
16	0
−4	1
0	1
18	0

The following decision tree cleanly separates the two classes, such that the resulting leaves are pure.



Q-1 is of form $x < p$. How many possible integer value can p take?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

10

Sub-Section Number : 8

Sub-Section Id : 640653103293

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 165 Question Id : 640653697744 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5

Question Label : Short Answer Question

Suppose you have a four-class classification problem where class label $y \in 0, 1, 2, 3$ and each training example x_i has binary features $f_1, f_2, f_3 \in 0, 1$. How many parameters do we need to know to classify an example using Naive Bayes classifier?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

15

Sub-Section Number : 9

Sub-Section Id : 640653103294

Question Shuffling Allowed : Yes

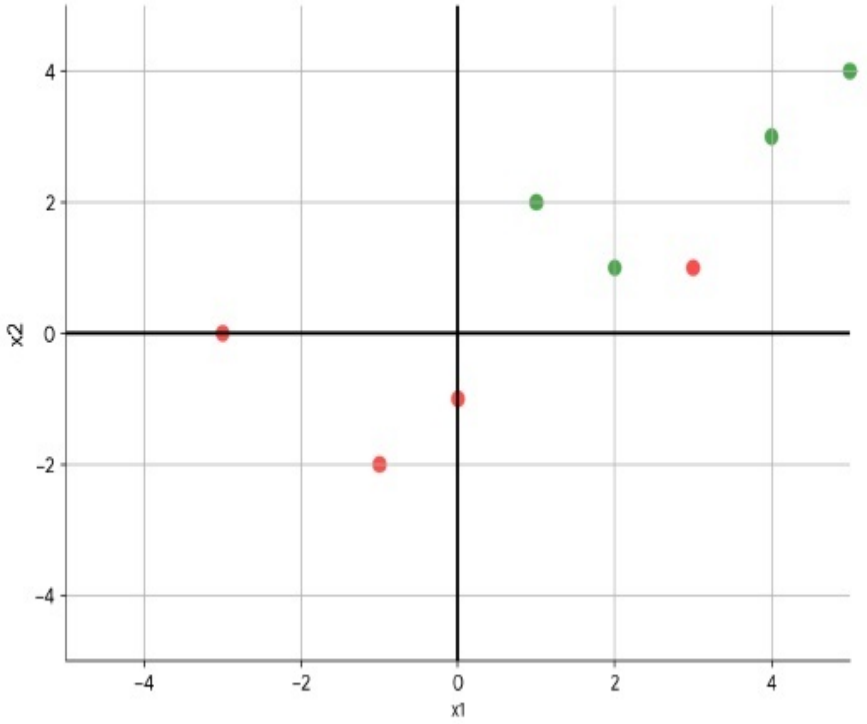
Is Section Default? : null

Question Number : 166 Question Id : 640653697745 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5

Question Label : Multiple Choice Question

The figure below displays the samples from dataset D, where each sample $x_i \in \mathbb{R}^2$. The green point belongs to **class1** and red point belongs to **class2**



Suppose we run the perceptron learning algorithm by initializing the weight vector to zero. Does the algorithm converge (with zero error) in a finite number of iterations?

Options :

6406532330333. ✖ Yes, it will converge.

6406532330334. ✔ No, it will never converge.

6406532330335. ✖ Insufficient data

Sub-Section Number :	10
Sub-Section Id :	640653103295
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 167 Question Id : 640653697746 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6

Question Label : Short Answer Question

Consider a logistic regression model that has been trained for a binary classification problem on a dataset in $\mathbb{R}^2$. The final weight vector learned by the model is $\mathbf{w} = \begin{bmatrix} 3/2 \\ 5/12 \end{bmatrix}$. Given a test data point as input to the model, it returns 1 as the predicted label if the probability output by the model is greater than 0.75 and 0 otherwise.

What is the predicted label for the test data point $\mathbf{x} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$? Note that the probability output by a logistic regression model is $P(y = 1 \mid \mathbf{x})$.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0

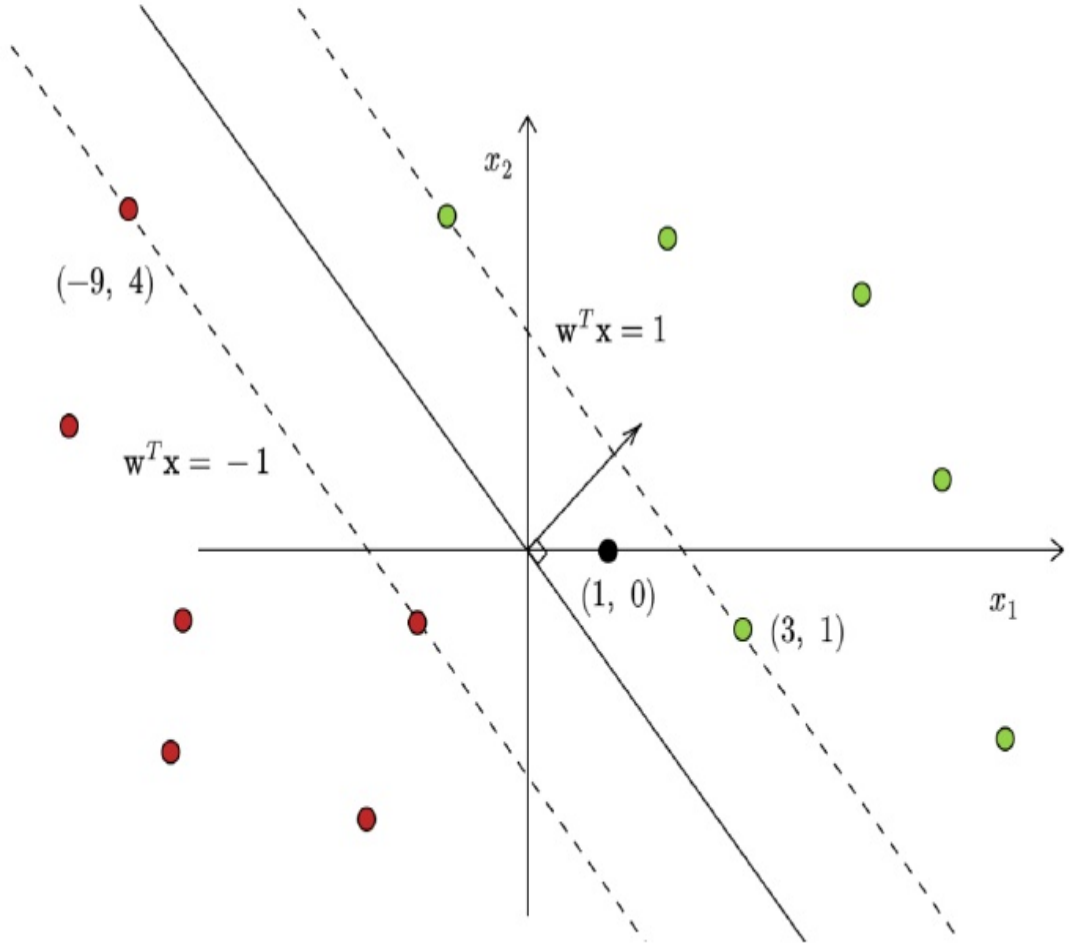
Sub-Section Number :	11
Sub-Section Id :	640653103296
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653697747 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (168 to 169)

Question Label : Comprehension

Consider a hard-margin SVM trained on a dataset in $\mathbb{R}^2$ for a binary classification task. Red and green points belong to the training dataset. Red points belong to class -1 and green points belong to class +1. The black-point is a test data-point. The dotted lines are the supporting hyperplanes for the SVM. Note: We don't have access to the test data-point during training; it is given to us after the model has been learned on the training dataset.



Based on the above data, answer the given subquestions.

Sub questions

Question Number : 168 Question Id : 640653697748 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6 Max. Selectable Options : 0

Question Label : Multiple Select Question

What is the equation of the decision boundary? Select all options that are correct.

Options :

6406532330337. ✔ $5x_1 + 6x_2 = 0$

6406532330338. ✔ $\frac{5}{21}x_1 + \frac{2}{7}x_2 = 0$

6406532330339. ✔ $\frac{5}{2}x_1 + 3x_2 = 0$

6406532330340. ✖ $\frac{27}{2}x_1 + 21x_2 = 0$

Question Number : 169 Question Id : 640653697749 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5

Question Label : Short Answer Question

What is the width of the separation between the two supporting hyperplanes? Enter your answer correct to two decimal places. (Hint: Calculate width using formulae $\frac{2}{||w||}$)

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

5.32 to 5.44

Sub-Section Number : 12

Sub-Section Id : 640653103297

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 170 Question Id : 640653697750 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider a soft-margin Support Vector Machine (SVM) for a binary classification problem with a dataset in a two-dimensional space. The optimization problem for the soft-margin SVM is formulated as:

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

subject to the constraints:

$$y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \text{ for all } i$$

Where C is a positive constant.

Which of the following statements about the soft-margin SVM is **correct**?

Options :

6406532330342. ✔ When $C = 0$, the optimal value of the objective function of the soft-margin problem is 0.

6406532330343. ✔ For a dataset with n data-points, there are $2n$ constraints for soft-margin SVM.

6406532330344. ✖ A smaller value of C allows for a larger margin, potentially leading to less misclassifications on the training data.

6406532330345. ✖ For a dataset with n data-points, there are n constraints for soft-margin SVM.

Question Number : 171 Question Id : 640653697751 Question Type : MSQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following statements are **correct**?

Options :

6406532330346. ✔ Underfitting models have high bias and low variance.

6406532330347. ✔ Overfitting models have low bias and high variance.

6406532330348. ✖ Generally, weak learners in the random forest tend to underfit.

6406532330349. ✔ If the performance of each estimator in the bagging algorithm is almost identical, the benefit of using bagging to combine them may be minimal or insignificant.

6406532330350. ✔ In random forests, multiple decision trees (estimators) are trained simultaneously, allowing for parallel processing and faster model training.

Sub-Section Number :	13
Sub-Section Id :	640653103298
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 172 Question Id : 640653697752 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Max. Selectable Options : 0

Question Label : Multiple Select Question

Suppose we have trained four different models using the same training set from the dataset D and recorded the training error. The testing error for each model was also recorded using a separate test set. The recorded values are summarized in the table below

Model	Training error	Test error
1	0.2	1.8
2	1.0	1.1
3	0.5	0.7
4	5.9	6.3

Based on the above information, which of the following statement(s) is/are correct?

Options :

6406532330351. ✖ Model 4 tends to overfit.

6406532330352. ✔ Model 4 tends to underfit.

6406532330353. ✖ Model 1 tends to underfits.

6406532330354. ✔ Model 1 tends to overfits.

Question Number : 173 Question Id : 640653697753 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider a binary classification problem. Suppose that we have 10 samples and each sample belongs to a positive (+1) or a negative class (-1). Suppose we define the squared error loss as follows

$$L(w) = \sum (h(x_i)y_i - y_i)^2$$

For which of the following $h(x)$ the loss function values can never be greater than 10?

Options :

6406532330355. ✖ $h(x_i) = w^T x_i$

6406532330356. ✔ $h(x_i) = \frac{1}{1+exp(-w^T x_i)}$

6406532330357.

✔ $h(x_i) = sign(w^T x_i)$

Sub-Section Number : 14

Sub-Section Id : 640653103299

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 174 Question Id : 640653697754 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 6 Max. Selectable Options : 0

Question Label : Multiple Select Question

Choose all the correct statements about neural networks

Options :

6406532330358. ✖ It can not be used for both regression and classification problems

6406532330359. ✔ It can have more than two hidden layers

6406532330360. ✔ The activation functions have to be non-linear to separate not linearly separable data points

6406532330361. ✔ Each neuron in the neural network may or may not have bias associated with it

MLP

Section Id : 64065349267

Section Number : 8

Section type : Online

Mandatory or Optional : Mandatory

Number of Questions : 36

Number of Questions to be attempted : 36