MLT

Section Id :	64065338402
Section Number :	5
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	21
Number of Questions to be attempted :	21
Section Marks :	100
Display Number Panel :	Yes
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 :	64065380937
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 98 Question Id : 640653566069 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 :

6406531891840. ✔ YES

6406531891841. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065380938
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 99 Question Id : 640653566070 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

A kernel k is defined as

$$k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$$
$$k([x_1, x_2]^T, [y_1, y_2]^T) = 1 + x_1 y_1 + x_1^2 y_1^2 + x_2^2 y_2^2$$

Which of the following transformation mappings corresponds to this kernel function?

Options :

6406531891842. ✖ $\phi([x_1, x_2]^T) = [1,\ x_1^2 + x_2^2]^T$

6406531891843. ✖ $\phi([x_1, x_2]^T) = [1,\ x_1 + x_1^2 + x_2^2]^T$

6406531891844. ✖ $\phi([x_1, x_2]^T) = [1,\ x_1^2,\ x_2^2]^T$

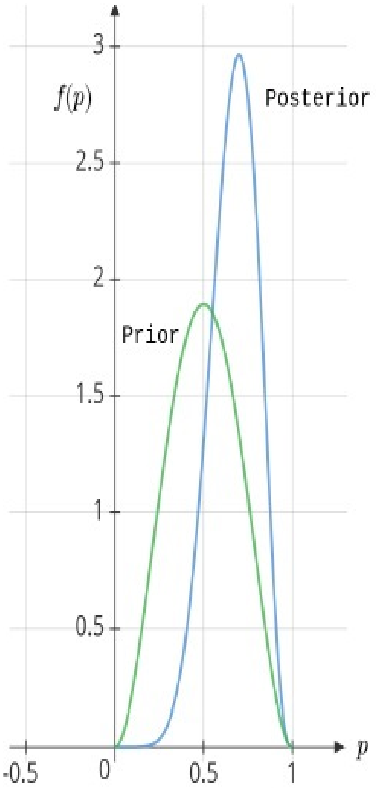
6406531891845. ✔ $\phi([x_1, x_2]^T) = [1,\ x_1,\ x_1^2,\ x_2^2]^T$

Question Number : 100 Question Id : 640653566071 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

Consider a Bayesian estimation problem for a dataset of six observations, where each observation is either one or zero. The prior (green) and posterior (blue) distributions are shown below. Recall that both are Beta distributions in this case. p denotes the parameter and $f(p)$ denotes its pdf. Also recall that the observations are sampled from a Bernoulli distribution with parameter p.



What can you say about the number of ones in the dataset? Choose the most appropriate option.

Options :

6406531891846. ✖ 1

6406531891847. ✖ 3

6406531891848. ✔ 5

Question Number : 101 Question Id : 640653566072 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

In LASSO, a higher value of the regularization parameter λ leads to more sparse weights.

Options :

6406531891849. ✔ TRUE

6406531891850. ✖ FALSE

Question Number : 102 Question Id : 640653566073 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

Consider a decision stump. The information gain for the question at the parent is **zero**. If the entropy of the left-child is equal to the entropy of the parent, what is the entropy of the right child? Assume that each child has at least one data-point in it and that the entropy of the parent is non-zero.

Options :

6406531891851. ✖ The entropy of the right-child is equal to half the entropy of the left-child.

6406531891852. ✖ The entropy of the right-child is equal to twice the entropy of the left-child.

6406531891853. ✔ The entropy of the right-child is equal to the entropy of the parent.

6406531891854. ✖ The entropy of the right-child is the sum of the entropies of the parent and the left-child.

Question Number : 103 Question Id : 640653566074 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

Consider the following training dataset for a binary classification problem in $\mathbb{R}^2$:

x_1	x_2	y
2	−1	−1
2	1	−1
−6	0	−1
6	0	1
−3	1	1
−3	−1	1

If we try to learn a perceptron model for this dataset, will the algorithm ever converge to a weight vector? Select the most appropriate answer with the information available to you.

Options :

6406531891855. ✖ Yes, it will certainly converge to a weight vector.

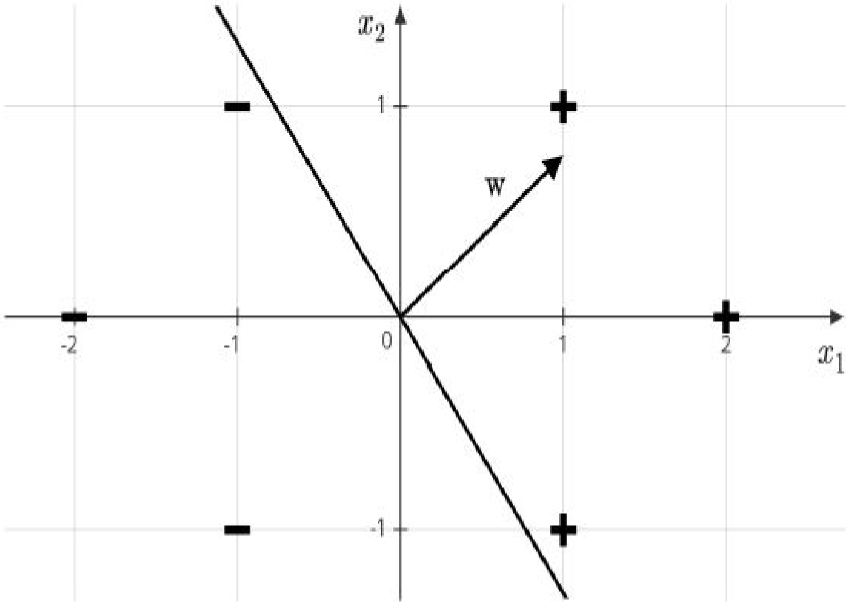
6406531891856. ✔ No, it will never converge.

Question Number : 104 Question Id : 640653566075 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

Consider a linearly separable binary classification problem. The training dataset is shown below. Is $\mathbf{w}$ the optimal weight vector corresponding to a hard-margin, linear-SVM? The symbol $+$ corresponds to label 1 and $-$ corresponds to label -1.



Options :

6406531891857. ✖ **w** shown in this diagram is the optimal weight vector for a hard-margin, linear-SVM

6406531891858. ✔ **w** shown in this diagram is **not** the optimal weight vector for a hard-margin, linear-SVM

Question Number : 105 Question Id : 640653566076 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

A hard-margin, linear-SVM has been trained on a linearly separable training dataset in a binary classification problem. The optimal weight vector is $\mathbf{w}^*$. A test-data point $\mathbf{x}_{\text{test}}$ is taken up for prediction:

$$\mathbf{w}^* = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_{\text{test}} = \begin{bmatrix} 0.5 \\ 0 \\ 0 \\ 0.1 \end{bmatrix}$$

What is the predicted label for this test point?

Options :

6406531891859. ✔ 1

6406531891860. ✖ -1

6406531891861. ✖ Since $0 < \mathbf{w}^{*^T}\mathbf{x}_{\text{test}} < 1$, this point violates the primal constraint and therefore cannot be classified.

6406531891862. ✖ Since $-1 < \mathbf{w}^{*^T}\mathbf{x}_{\text{test}} < 0$, this point violates the primal constraint and therefore cannot be classified.

Question Number : 106 Question Id : 640653566077 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

Consider a hard-margin, linear-SVM trained for a linearly separable dataset in a binary classification problem in $\mathbb{R}^7$. The training dataset has 5 points. If the optimal weight vector is some non-zero vector in $\mathbb{R}^7$, which of the following could be the optimal $\boldsymbol{\alpha}^*$, the optimal dual solution? Select the most appropriate answer.

Options :

6406531891863. ✖ $\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

6406531891864. ✖

$$\begin{bmatrix}1\\2\\0\\0\\1\\3\\0\end{bmatrix}$$

6406531891865. ✔ $\begin{bmatrix}1\\1\\0\\0\\0\end{bmatrix}$

6406531891866. ✖ $\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix}$

Sub-Section Number :	3
Sub-Section Id :	64065380939
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 107 Question Id : 640653566078 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Selectable Option : 0

Question Label : Multiple Select Question

Consider a linear regression model that was trained on dataset X of shape (d, n). Which of the following techniques could potentially decrease the loss on the training data (assuming the loss is the squared error)?

Options :

6406531891867. ✔ Adding a dummy feature in the dataset and learning the intercept w_0 as well.

6406531891868. ✖ Penalizing the model weights with L2 regularization.

6406531891869. ✖ Penalizing the model weights with L1 regularization.

6406531891870. ✔ Training the kernel regression model of degree 2.

Question Number : 108 Question Id : 640653566079 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Selectable Option : 0

Question Label : Multiple Select Question

The weight vector for two consecutive iterations of a perceptron model are given below:

$$\mathbf{w}^t = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \quad \mathbf{w}^{t+1} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

Select all possible data-points that could have been used for the update from $\mathbf{w}^t$ to $\mathbf{w}^{t+1}$. The first entry in each option is the data-point, the second entry is its true label.

Options :

6406531891871. ✔ $\begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \quad 1$

6406531891872. ✔ $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad -1$

6406531891873. ✖ $\begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}, \quad -1$

6406531891874. ✖ $\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \quad 1$

Question Number : 109 Question Id : 640653566082 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Selectable Option : 0

Question Label : Multiple Select Question

Let L be the set of loss functions:

$$L = \{\text{Hinge loss, Logistic loss, Squared loss, 0-1 loss}\}$$

All the loss functions mentioned are evaluated for a single data-point $(\mathbf{x}, y)$ and weight vector $\mathbf{w}$.
Select all true statements.

Options :

6406531891883. ✖ The hinge loss is always greater than the logistic loss.

6406531891884. ✔ If $(\mathbf{w}^T\mathbf{x})y = 1$, then the logistic loss is greater than the following losses: squared loss, hinge loss, 0-1 loss

6406531891885. ✖ The 0-1 loss is convex

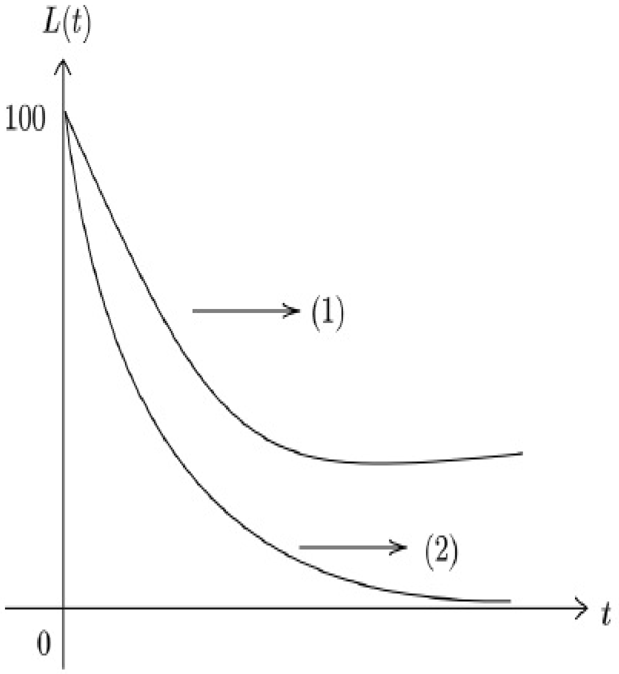
6406531891886. ✔ The squared loss has the maximum value among all the loss functions in L when $(\mathbf{w}^T\mathbf{x})y$ is a very large number.

Question Number : 110 Question Id : 640653566083 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Selectable Option : 0

Question Label : Multiple Select Question

Consider a supervised ML problem that has a training dataset and a validation dataset. A ML model is being trained on the training dataset using gradient descent and its performance is monitored on the validation dataset. The loss function $L(t)$ at time step t is plotted against t.



One of these curves corresponds to the model's loss on the training dataset and the other corresponds to the model's loss on the validation dataset. Identify the two curves. Consider a general scenario and not an extreme instance while answering this problem. Exactly two options are correct.

Options :

6406531891887. ✔ (1) is the loss on the validation dataset

6406531891888. ✔ (2) is the loss on the training dataset

6406531891889. ✖ (1) is the loss on the training dataset

6406531891890. ✖ (2) is the loss on the validation dataset

Sub-Section Number :	4
Sub-Section Id :	64065380940
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 111 Question Id : 640653566080 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Let w^*, ξ^* be the optimal primal solutions, and α^*, β^* be the optimal dual solutions of the soft-margin SVM problem. If $\alpha_i^* = 0$, select the correct options.

Options :

6406531891875. ✖ The bribe paid by the i^{th} data point is equal to C.

6406531891876. ✔ The bribe paid by the i^{th} data point is equal to 0.

6406531891877. ✖ i^{th} data point lies on the supporting hyper-planes.

6406531891878. ✔ w^* classifies the i^{th} data point correctly.

Question Number : 112 Question Id : 640653566081 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

Choose the correct statements.

Options :

6406531891879. ✔ 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.

6406531891880. ✖ Weak learners have low bias and high variance.

6406531891881. ✔ In random forests, multiple decision trees (estimators) are trained simultaneously, allowing for parallel processing and faster model training.

6406531891882. ✖ In the random forest algorithm, multiple decision trees are trained on the same dataset to create an ensemble model.

Sub-Section Number :	5
Sub-Section Id :	64065380941
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 113 Question Id : 640653566084 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

Standard-PCA was applied on a centered dataset in $\mathbb{R}^6$. All the eigenvalues of the covariance matrix of this centered dataset are given below in decreasing order with $\lambda > 1$:

$$\lambda^5, \lambda^4, \lambda^3, \lambda^2, \lambda, 1$$

If the sum of the variances along the first three principal components is exactly 88.88% of the total variance, what is the value of λ? Enter the nearest integer as your answer. Note that the total variance is the sum of the variances along each principal component.

Hint: $\lambda^2 + \lambda + 1$

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2

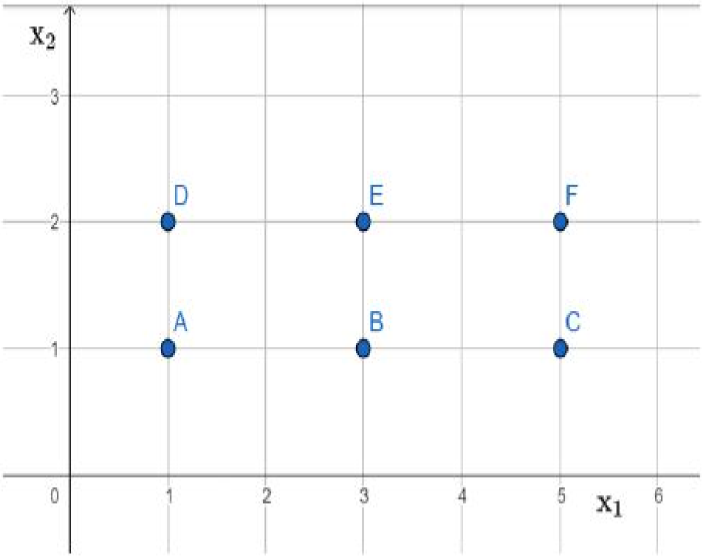
Question Number : 114 Question Id : 640653566085 Question Type : SA Calculator : None

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

Correct Marks : 5

Question Label : Short Answer Question

Consider the K-means algorithm was run on the following two-dimensional dataset with $K = 3$. The initial cluster centroids are points A, B, and D as shown in the figure.



How many cluster centroids in the final clusters will be the same as the initial cluster centroids? Use Euclidean distance to calculate the distance.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2

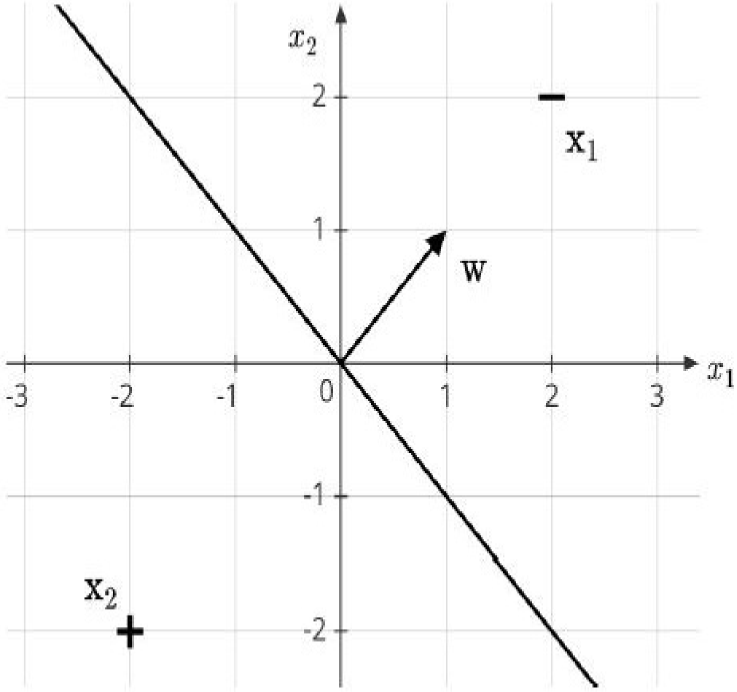
Question Number : 115 Question Id : 640653566086 Question Type : SA Calculator : None

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

Correct Marks : 5

Question Label : Short Answer Question

Consider a logistic regression model that has been trained for a binary classification problem on a dataset in $\mathbb{R}^2$. Consider two data-points from the training set: $(\mathbf{x}_1, y_1)$ and $(\mathbf{x}_2, y_2)$. The model's weight vector and these two data-points (along with their labels) are drawn in the diagram given below. The symbol $+$ corresponds to label 1 and $-$ corresponds to label 0.



σ_1 and σ_2 are the probabilities output by the model for the two data-points. If $|\sigma_1 - y_1|$ and $|\sigma_2 - y_2|$ are treated as the errors in prediction for the two data-points, compute the following quantity:

$$\frac{|\sigma_1 - y_1|}{|\sigma_2 - y_2|}$$

Note that the probability output by a logistic regression model is $P(y = 1 \mid \mathbf{x})$. The figure is **drawn to scale** and you can use it to get the required values. Enter the nearest positive integer as your answer

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

1

Question Number : 116 Question Id : 640653566087 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

Consider the following architecture for a neural network:

Layer	Neurons
Input	5
Hidden Layer-1	30
Hidden layer-2	10
Output layer	1

How many weights does this network have? Assume that there is no bias associated with any neuron.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

460

Sub-Section Number :	6
Sub-Section Id :	64065380942
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653566088 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 : (117 to 119)

Question Label : Comprehension

The Gaussian Naive Bayes algorithm was run on the following dataset:

feature 1 (f_1)	feature 2 (f_2)	Label
1.5	1.6	1
2.2	2.4	1
2.9	1.5	0
1.7	0.8	1

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 117 Question Id : 640653566089 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Short Answer Question

What will be the value of $\hat{p}$?

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 :

0.74 to 0.76

Question Number : 118 Question Id : 640653566090 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What will be the value of $\hat{\mu}_0$?

Options :

6406531891896. ✖ 2.9

6406531891897. ✖ 2.2

6406531891898. ✖ (1.8,1.6)

6406531891899. ✔ (2.9,1.5)

Question Number : 119 Question Id : 640653566091 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What will be the value of $\hat{\mu}_1$?

Options :

6406531891900. ✖ 2.9

6406531891901. ✖ 1.8

6406531891902. ✔ (1.8,1.6)

6406531891903. ✖ (2.9,1.5)

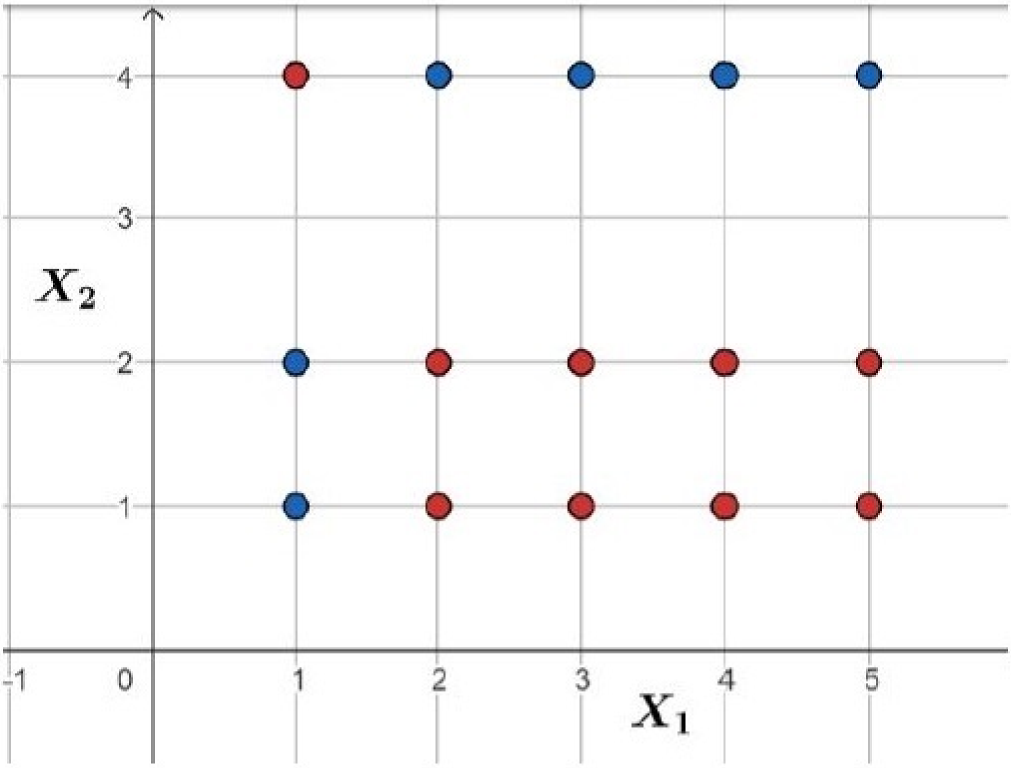
Sub-Section Number :	7
Sub-Section Id :	64065380943
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653566092 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 : (120 to 122)

Question Label : Comprehension

Consider the following two-dimensional dataset with two classes: $+1$ for blue points and -1 for red points. An AdaBoost algorithm was run on this dataset using decision stumps as weak learners.



When training the new weak learner $h_t(x)$ (decision stump at t^{th} iteration), we choose the split that minimizes the weighted misclassification error with respect to current weights D_t i.e. choose h_t that minimizes

$$\sum_{i=1}^{n} D_t(i)\mathbb{1}(h_t(x_i) \neq y_i).$$

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 120 Question Id : 640653566093 Question Type : SA Calculator : None

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

Correct Marks : 3

Question Label : Short Answer Question

What will be the misclassification error incurred by the first decision stump?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.2

Question Number : 121 Question Id : 640653566094 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Selectable Option : 0

Question Label : Multiple Select Question

To train the second decision stump, which pair of points will be assigned equal weights to create the training dataset?

Options :

6406531891905. ✔ $[2, 2]^T, [2, 4]^T$

6406531891906. ✖ $[2, 2]^T, [1, 4]^T$

6406531891907. ✔ $[1, 1]^T, [1, 4]^T$

6406531891908. ✔ $[3, 1]^T, [4, 1]^T$

Question Number : 122 Question Id : 640653566095 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Short Answer Question

What weight will be assigned to the point [1, 1]*T* when training the second decision stump, assuming the weights are not normalized to add up to one? Please enter your answer rounded to two decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.11 to 0.15

AppDev1

Section Id :	64065338403
Section Number :	6
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	31
Number of Questions to be attempted :	31
Section Marks :	100
Display Number Panel :	Yes
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 :	64065380944
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 123 Question Id : 640653566096 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction