6406531958897. ✔ Both REST and GraphQL are in general used for fetching data from a remote storage.

6406531958898. ✖ Every API must adhere to all the REST principles.

6406531958899. ✔ In GraphQL, a write operation should be performed explicitly via a mutation, in general.

MLT

Section Id :	64065339717
Section Number :	11
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	13
Number of Questions to be attempted :	13
Section Marks :	50
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 :	64065384399
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 157 Question Id : 640653587065 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 :

6406531958900. ✔ YES

6406531958901. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065384400
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 158 Question Id : 640653587066 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Multiple Choice Question

Consider a training dataset of n points for a regression problem. Assume that the model is linear. Let $\mathbf{w}_1$ and $\mathbf{w}_2$ be the optimal weight vectors obtained from solving the following optimization problems.

$$\mathbf{w}_1 = \arg\min_{\mathbf{w}} \sum_{i=1}^{n} (\mathbf{w}^T \mathbf{x}_i - y_i)^2$$

$$\mathbf{w}_2 = \arg\min_{\mathbf{w}} \sum_{i=1}^{n} (\mathbf{w}^T \mathbf{x}_i - y_i)^3$$

Choose the most appropriate answer.

Options :

6406531958902. ✔ $\mathbf{w}_1$ will generalize better than $\mathbf{w}_2$ on the test dataset.

6406531958903. ✖ $\mathbf{w}_2$ will generalize better than $\mathbf{w}_1$ on the test dataset.

6406531958904. ✖ Both models will show identical performance on the test dataset.

Question Number : 159 Question Id : 640653587067 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4

Question Label : Multiple Choice Question

The training dataset for a binary classification problem is as follows:

$$\{\, (\mathbf{u}, 1), \quad (-\mathbf{u}, 0), \quad (2\mathbf{u}, 1), \quad (-2\mathbf{u}, 0)\,\}$$

where, $\mathbf{u} \in \mathbb{R}^d$ is a non zero constant and each element in the set given above is a data-point of the form $(\mathbf{x}_i, y_i)$. The labels lie in $\{0, 1\}$. Consider a linear classifier with weight vector $\mathbf{w}$. What condition should the weight vector satisfy for the zero-one loss to be zero on this dataset?

Options :

6406531958905. ✖ $\mathbf{w}^T\mathbf{u} < 0$

6406531958906. ✔ $\mathbf{w}^T\mathbf{u} > 0$

6406531958907. ✖ $\mathbf{w}^T\mathbf{u} = 0$

6406531958908. ✖ We can never find a $\mathbf{w}$ for which the zero-one loss becomes zero on this dataset.

Sub-Section Number :	3
Sub-Section Id :	64065384401

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 160 Question Id : 640653587068 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4 Max. Selectable Options : 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 :

6406531958909. ✔ Adding a dummy feature in the dataset and learning the intercept w_0 as well.

6406531958910. ✖ Penalizing the model weights with L2 regularization.

6406531958911. ✖ Penalizing the model weights with L1 regularization.

6406531958912. ✔ Training the kernel regression model of degree 2.

Question Number : 161 Question Id : 640653587070 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following statements are true about the decision tree algorithm?

Options :

6406531958917. ✖ Decision trees are prone to overfit if the maximum depth is set too low.

6406531958918. ✔ Decision trees are prone to underfit if the maximum depth is set too low.

6406531958919. ✔ Decision trees are sensitive to small perturbations in the dataset and can result in different tree structures.

6406531958920. ✔ Decision trees can handle both numerical and categorical features.

Sub-Section Number :	4
Sub-Section Id :	64065384402
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 162 Question Id : 640653587069 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 4.5 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following statements is/are true regarding solution of Ridge regression problem?

Options :

6406531958913. ✔ If there are multiple w solutions for minimizing mean square error, then w_R will be the one with least norm.

6406531958914. ✖ If there are multiple w solutions for minimizing mean square error, then w_R will be the one with highest norm.

6406531958915. ✔ Prior for w is $\mathrm{N}(0, \gamma^2 I)$ and $y_i|x_i \sim \mathrm{N}(w^T x_i, \sigma^2)$

6406531958916. ✖ Prior for w is $\mathrm{N}(1, \gamma^2 I)$ and $y_i|x_i \sim \mathrm{N}(0, \sigma^2)$

Sub-Section Number :	5
Sub-Section Id :	64065384403
Question Shuffling Allowed :	Yes

Is Section Default? : null

Question Number : 163 Question Id : 640653587071 Question Type : SA Calculator : None

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

Correct Marks : 4.5

Question Label : Short Answer Question

Consider kernel regression with the kernel function $(\mathbf{x}_1^T\mathbf{x}_2 + 2)^2$ applied on the following dataset.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 2 & 0 & 3 & 0 \\ 0 & 1 & 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The optimal weight vector $\mathbf{w}^*$ is given by:

$$\mathbf{w}^* = \phi(X)[0.1, 2, 3.9, 5, 6, 8]^T$$

where ϕ is transformation mapping corresponding to the given kernel. What will be the prediction for the data point $[0, 0, 1]^T$?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

100

Sub-Section Number : 6

Sub-Section Id : 64065384404

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 164 Question Id : 640653587072 Question Type : SA Calculator : None

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

Correct Marks : 4

Question Label : Short Answer Question

Consider a ridge regression model with the loss $L(\mathbf{w}) = ||X^T\mathbf{w} - \mathbf{y}||^2 + \lambda||\mathbf{w}||^2$ is trained on a given dataset with $\lambda = 0.1, 0, 1, 10, 100$. Which of the following value of λ is more likely to underfit the model?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

100

Question Number : 165 Question Id : 640653587073 Question Type : SA Calculator : None

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

Correct Marks : 4

Question Label : Short Answer Question

Consider the following data set:

$X = [8, 6, 10]$

Assuming a ridge penalty $\lambda = 100$, what will be the value of $\frac{\hat{w}_{ridge}}{\hat{w}_{MLE}}$?

Here $\hat{w}_{ridge}$ and $\hat{w}_{MLE}$ are the Ridge and MLE estimates of the weight vectors, respectively. Assume that the label vector y of shape $(3, 1)$ is known. 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 :

0.65 to 0.70

Sub-Section Number :	7
Sub-Section Id :	64065384405
Question Shuffling Allowed :	Yes

Is Section Default? : null

Question Number : 166 Question Id : 640653587074 Question Type : SA Calculator : None

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

Correct Marks : 2.5

Question Label : Short Answer Question

A binary classification dataset contains only one feature and the data points given the label follow the Gaussian distributions whose means and variances are already estimated as:

$$x|(y=0) \sim \mathrm{N}(0,1)$$
$$x|(y=1) \sim \mathrm{N}(2,2)$$

What will be the prediction for the point $x = 1$? Assume that $\hat{p}$, an estimate for $P(y=1)$, is 0.5.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0

Sub-Section Number : 8

Sub-Section Id : 64065384406

Question Shuffling Allowed : No

Is Section Default? : null

Question Id : 640653587075 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 : (167 to 168)

Question Label : Comprehension

Consider a binary classification problem and a decision tree that is being trained to classify the points. In one of the internal nodes in this tree, 75% of the data-points belong to one of the two classes and the rest belong to the other class. You are not given the information about which class is more numerous in this node.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 167 Question Id : 640653587076 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1.5

Question Label : Multiple Choice Question

Do you have enough information to find the entropy of this node?

Options :

6406531958925. ✔ Yes

6406531958926. ✖ No

Question Number : 168 Question Id : 640653587077 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

If the answer to the previous questions is "Yes", find the entropy of the node. Use $\log_2$ and enter your answer correct to three decimal places.

If the answer to the previous question is "No", enter -1 as your answer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.79 to 0.83

Sub-Section Number : 9

Sub-Section Id : 64065384407

Question Shuffling Allowed : No

Is Section Default? : null

Question Id : 640653587078 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 : (169 to 171)

Question Label : Comprehension

Consider a probability distribution over (X, y) where features are one-dimensional and $y \in \{+1, -1\}$. Let $X|(y=1)$ follow a uniform distribution over $[0, 4]$ and $X|(y=-1)$ follows a uniform distribution over $[2, 4]$.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 169 Question Id : 640653587079 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

If $p = P(y = 1)$ is estimated to be 0.4, what will be the prediction for the point $x = 3$ using the Bayes classifier? Enter 1 or -1.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

-1

Question Number : 170 Question Id : 640653587080 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

Let $x = 2$ and let $\hat{p}$ be the estimate for $p = P(y = 1)$. Find conditions on $\hat{p}$ such that the Bayes classifier predicts 1 for this x. Consider that the tie-breaker is predicted in class 1.

Options :

6406531958929. ✖ $\hat{p} \leq \frac{1}{4}$

6406531958930. ✖ $\hat{p} \geq \frac{1}{4}$

6406531958931. ✖ $\hat{p} \leq \frac{2}{3}$

6406531958932. ✔ $\hat{p} \geq \frac{2}{3}$

Question Number : 171 Question Id : 640653587081 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

If $p = P(y = 1)$ is estimated to be 0.5 using MLE on a given training dataset, what will be the training error of the Bayes classifier for this problem?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.5

Sub-Section Number :	10
Sub-Section Id :	64065384408
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653587082 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 : (172 to 173)

Question Label : Comprehension

Consider a naive Bayes model is trained on the following data matrix X of shape (d, n) and corresponding label vector y:

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad y = [1 \quad 0 \quad 1 \quad 0]^T$$

Assume that $\hat{p}$ and $\hat{p}_j^{y_i}$ are estimates for $P(y = 1)$ and $P(f_j = 1 | y = y_i)$, respectively. Here, f_i; $i = 1, 2, 3$ is the i^{th} feature. These parameters are estimated using MLE. Do not apply any smoothing on the dataset.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 172 Question Id : 640653587083 Question Type : SA Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

Calculate the value of $\hat{p}_2^0$.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0.5

Question Number : 173 Question Id : 640653587084 Question Type : SA Calculator : None

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

Correct Marks : 2

Question Label : Short Answer Question

Calculate the value of $\hat{p}_2^1$.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

0

MLP

Section Id :	64065339718
Section Number :	12
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	24