Indian Institute of Technology, Madras - BS in Data Science and Applications

Notations :

1.Options shown in green color and with ✔ icon are correct.

2.Options shown in red color and with ✖ icon are incorrect.

Question Paper Name :	IIT M FOUNDATION AN4 EXAM QPF4 16 JULY 2023
Subject Name :	2023 July: IIT M FOUNDATION AN4 EXAM QPF4
Creation Date :	2023-07-10 18:54:05
Duration :	240
Total Marks :	705
Display Marks:	Yes
Share Answer Key With Delivery Engine :	Yes
Actual Answer Key :	Yes
Calculator :	Scientific
Magnifying Glass Required? :	No
Ruler Required? :	No
Eraser Required? :	No
Scratch Pad Required? :	No
Rough Sketch/Notepad Required? :	No
Protractor Required? :	No
Show Watermark on Console? :	Yes
Highlighter :	No
Auto Save on Console?	Yes
Change Font Color :	No

Options :

6406531929295. ✖ hello

6406531929296. ✖ HELLO

6406531929297. ✔ OLLEH

6406531929298. ✖ olleh

MLT

Section Id :	64065339065
Section Number :	15
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 :	64065382542
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 282 Question Id : 640653577728 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 :

6406531929299. ✔ YES

6406531929300. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	64065382543
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 283 Question Id : 640653577729 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 the following kernel function:

$$k : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$k([x_1, x_2]^T, [y_1, y_2]^T) = 1 + x_1y_1 + x_2y_2 + x_1^2y_1^2x_2y_2 + x_1y_1x_2^2y_2^2$$

Find the appropriate transformation mapping ϕ for the given kernel.

Options :

6406531929301. ✖ $\phi([x_1, x_2]^T) = [1, \ x_1x_2, \ x_1^2x_2^2]$

6406531929302. ✖ $\phi([x_1, x_2]^T) = [1, \ x_1, \ x_2, \ x_1^2x_2^2]$

6406531929303. ✔ $\phi([x_1, x_2]^T) = [1, \ x_1, \ x_2, \ x_1^2x_2, \ x_1x_2^2]$

6406531929304. ✖ $\phi([x_1, x_2]^T) = [1, \ x_1, \ x_2, \ x_1x_2, \ x_1^2x_2^2]$

Question Number : 284 Question Id : 640653577730 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 the following Hard K-means clustering problem with four points: $P(1,1), Q(2,1), R(4,3)$, and $S(5,4)$. Consider the number of clusters to be $k = 2$ and the initial centroids to be $C_1 = (0,0)$ and $C_2 = (4,4)$. After how many iterations will the algorithm terminate and what will be the final centroids? use euclidean distance metric to calculate distance.

Options :

6406531929305. ✖ $2, (1,1), (4,4)$

6406531929306. ✖ $3, (1,1), (4,4)$

6406531929307. ✔ $2, (1.5,1), (4.5,3.5)$

6406531929308. ✖ $4, (1.5,1), (4.5,3.5)$

Sub-Section Number : 3

Sub-Section Id : 64065382544

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 285 Question Id : 640653577731 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

Select all true statements regarding standard PCA applied on a centered dataset.

Options :

6406531929309. ✔ The variance of the dataset along the first principal component is maximum.

6406531929310. ✖ The variance of the dataset along the first principal component is minimum.

6406531929311. ✔ The first principal component is the unit norm eigenvector corresponding to the largest eigenvalue of the covariance matrix.

6406531929312. ✖ The first principal component is the unit norm eigenvector corresponding to the smallest eigenvalue of the covariance matrix.

Question Number : 286 Question Id : 640653577732 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 dataset of 2000 points all of which lie in $\mathbb{R}^{30}$. If all these points lie in four dimensional low subspace then after applying the PCA algorithm which of the following statements is/are true?

Options :

6406531929313. ✖ The covariance matrix corresponding to this dataset has 30 non-zero eigenvalues.

6406531929314. ✔ The covariance matrix corresponding to this dataset has only four non-zero eigenvalues.

6406531929315. ✔ The residues will become zero after four rounds

6406531929316. ✔ By utilizing four principal component vectors and their corresponding coefficients, we have the ability to reconstruct the complete dataset using representatives.

Sub-Section Number : 4

Sub-Section Id : 64065382545

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 287 Question Id : 640653577733 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

Which of the following statements about kernel PCA is/are always true?

Options :

6406531929317. ✖ The number of principal components corresponding to nonzero eigenvalues obtained using kernel PCA on d dimensional dataset cannot be more than d.

6406531929318. ✔ For every transformation mapping $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$, there exists a valid kernel function.

6406531929319. ✔ For two valid kernels $k_1, k_2 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, $ak_1 + bk_2$ is a valid kernel function. Here both a and b are positive real numbers.

6406531929320. ✖ The output of a valid kernel function can never be a negative real number.

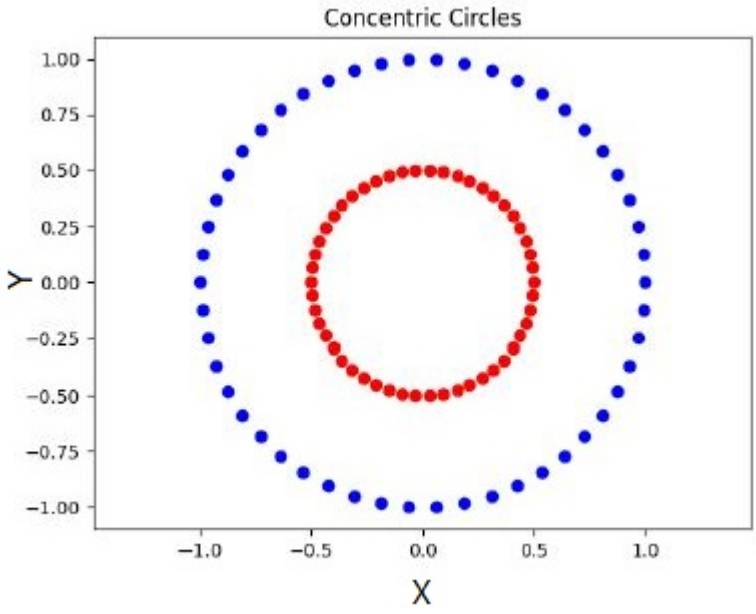
Question Number : 288 Question Id : 640653577735 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 5 Max. Selectable Options : 0

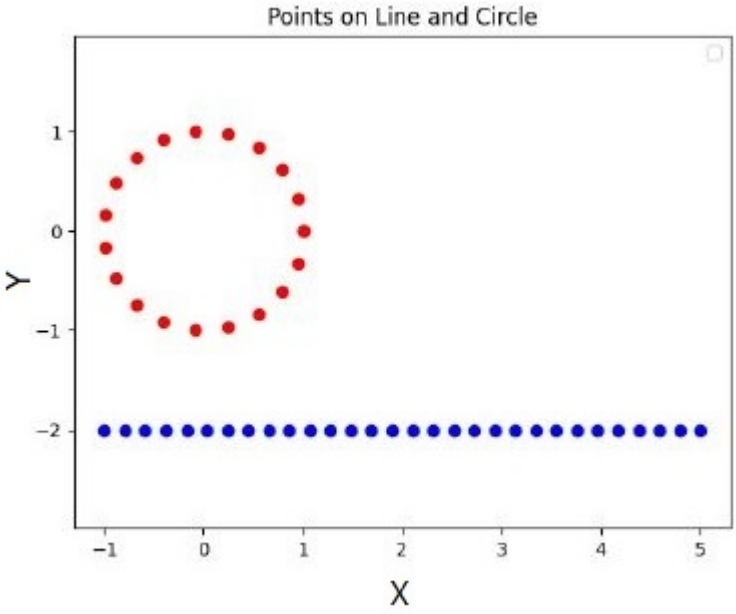
Question Label : Multiple Select Question

Which of the following clusters cannot be obtained as final clusters using Lloyd's algorithm? Assume that Lloyd's algorithm was run on these datasets with $k = 2$.

Options :

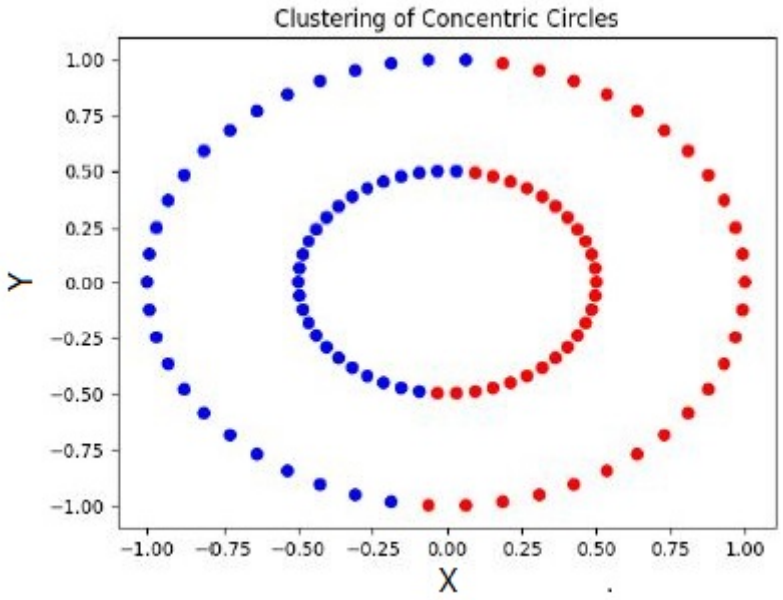


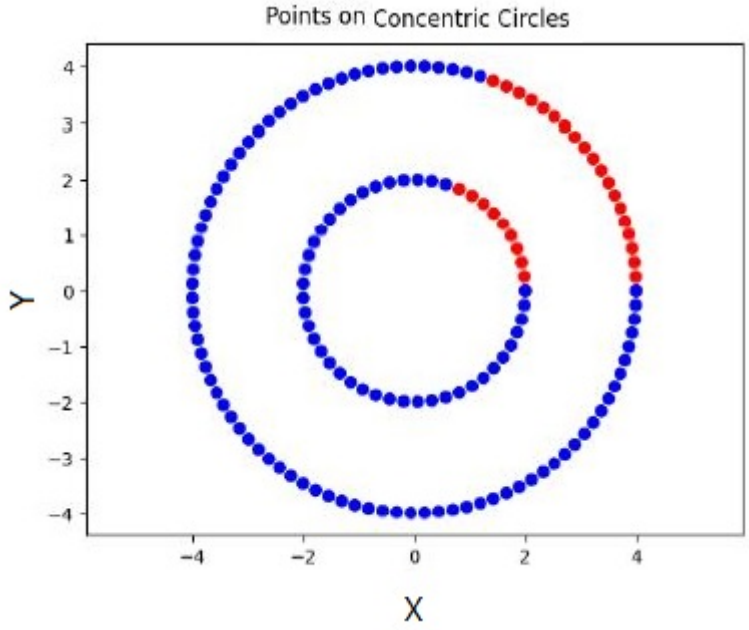
6406531929325. ✔



6406531929326. ✔

6406531929327. ✖





6406531929328.

Sub-Section Number : 5

Sub-Section Id : 64065382546

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 289 Question Id : 640653577734 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3.5 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following statements is/are true?

Options :

6406531929321. ✔ $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, k(x_1, x_2) = (x_1^T x_2)^3$ is a valid kernel.

6406531929322. ✖ $k : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, k(x_1, x_2) = -(x_1^T x_2)^3$ is a valid kernel.

6406531929323. ✔ nonzero Eigenvalues of XX^T and X^TX are the same.

6406531929324. ✖ Eigenvectors of XX^T are same as eigenvectors of X^TX.

Sub-Section Number :	6
Sub-Section Id :	64065382547
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 290 Question Id : 640653577736 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 dataset in $\mathbb{R}$:

$$\{-3, \quad -1, \quad 0, \quad 1, \quad 4, \quad 299, \quad 300, \quad 301\}$$

Lloyd's algorithm ($K = 2$) is run on this dataset with the points -10 and 310 as the initial cluster centers. Let a, b be the final cluster centers on convergence, what is the value of the product ab?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

60

Question Number : 291 Question Id : 640653577737 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 dataset that has 100 data-points, each of which is either a zero or one. The Bernoulli distribution is used to model this data. If the MLE estimate for the parameter p of the Bernoulli distribution is 0.3, how many zeros does the dataset have?

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

70

Question Number : 292 Question Id : 640653577739 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 dataset with 60 points in which all points are either 0 or 1. We use a Bernoulli distribution with the parameter p to model this problem. The prior and posterior distributions for the parameter p are Beta(10,5) and Beta(30,45) respectively. How many data points have the value 0 in this dataset? Note that $p = P(x = 1)$ as usual.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

40

Sub-Section Number : 7

Sub-Section Id : 64065382548

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 293 Question Id : 640653577738 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 a GMM with two components:

$$\pi_1 = 0.3, \quad \pi_2 = 0.7$$

$$\mu_1 = 0, \quad \sigma_1^2 = 1$$

$$\mu_2 = 1, \quad \sigma_2^2 = 2$$

What is the probability that the point $x = 1.2$ comes from the first component? Use the following table for your reference. Here, $\mathcal{N}(x, \mu, \sigma^2)$ is the density of the Gaussian with mean μ and variance σ^2 evaluated at x.

μ	σ^2	$\mathcal{N}(1.2, \mu, \sigma^2)$
0	1	0.194
0	2	0.197
1	1	0.391
1	2	0.279

Enter your answer correct to three decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

0.15 to 0.35

Sub-Section Number :	8
Sub-Section Id :	64065382549
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653577740 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 : (294 to 295)

Question Label : Comprehension

Standard PCA has been performed on a centered dataset in $\mathbb{R}^3$.
The first two principal components are given below:

$$\mathbf{w}_1 = \frac{1}{\sqrt{5}} \cdot \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{6}} \cdot \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

Consider the following data-point in the dataset:

$[1 \quad 2 \quad 1]^T$. (a, b) is the representation of this point in the coordinate system formed by the two principal components given above. The first and second coordinates correspond to PC-1 and PC-2 respectively.

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 294 Question Id : 640653577741 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

What is the value of *a*? Enter your answer correct to three decimal places.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

Possible Answers :

1.25 to 1.5

Question Number : 295 Question Id : 640653577742 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

What is the value of *b*? 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 :

0.25 to 0.55