EE2211_AY2425_Sem2_Midterm_S - Confidential

# of Questions: 24
Date and Time of Exam Creation: Thu, Mar 13, 2025 @ 16:48:29
Exam Folder: College of Design and Engineering\Electrical & Computer Engineering\AY 2024-2025 SEM2\EE2211

Question #: 1

We would like to design a 2nd order polynomial function to identify if the participants have dementia disease or not. We have collected two clinical features of 4 participants (

x_1, x_2

) that are related to this disease. We also have ground truth labels for each participant indicating whether they have dementia: a label of 1 means the participant has dementia, and -1 means they do not (see below).

$$\{x_1 = 3, x_2 = 3.5\} \rightarrow \{y = -1\}$$

$$\{x_1 = 9, x_2 = 2.3\} \rightarrow \{y = 1\}$$

$$\{x_1 = 7, x_2 = 5\} \rightarrow \{y = -1\}$$

$$\{x_1 = 2, x_2 = 7\} \rightarrow \{y = -1\}$$

1. Perform 2nd order polynomial function based on the existing training data. What is the estimated weight for the term

x_1^2

? Your Answer 1 (up to 4 decimal places)

2. If we have a new participant with the two clinical features (

x_1, x_2

) =(10, 4.5), does this participant have the disease or not? Your Answer 2
(select your answer from 1 or -1)

1. Range - Min:0.048 Max:0.049
2. Choice of: 1 | -1 - Correct Answer:1

Question #: 2

We would like to design a linear regression function to predict future cognitive performance and depressive severity in the elderly. We have collected three clinical features of 5 participants (

$$x_1, x_2, x_3$$

) that are related to future cognitive decline and depression. We also have the five-year follow-up data of cognitive scores (

$$y_1$$

) and depressive scores (

$$y_2$$

) of these 5 participants (see below).

$$\{x_1 = 71, x_2 = 3, x_3 = 15\} \rightarrow \{y_1 = 5, y_2 = 7\}$$
$$\{x_1 = 59, x_2 = 7, x_3 = 12\} \rightarrow \{y_1 = 10, y_2 = 8\}$$
$$\{x_1 = 65, x_2 = 5, x_3 = 9\} \rightarrow \{y_1 = 3, y_2 = 8\}$$
$$\{x_1 = 81, x_2 = 3, x_3 = 19\} \rightarrow \{y_1 = 7, y_2 = 3\}$$
$$\{x_1 = 55, x_2 = 6, x_3 = 11\} \rightarrow \{y_1 = 6, y_2 = 7\}$$

1. Perform least square estimation based on the existing training data. What is the mean of squared error of the estimated model for cognitive scores? Your Answer

1 (up to 4 decimal places);
What is the mean of squared error of the estimated model for depressive scores? Your Answer 2 (up to 4 decimal places).
2. If we have a new participant with the two clinical features (x_1, x_2, x_3) =(75, 5, 11). What is his/her five year follow-up cognitive score? Your Answer 3 (up to 4 decimal places);
What is his/her five year follow-up depressive score? Your Answer 4 (up to 4 decimal places).

1. Range - Min:0.33 Max:0.34
2. Range - Min:0.68 Max:0.69
3. Range - Min:4.77 Max:4.78
4. Range - Min:6.75 Max:6.76

Question #: 3

A set of equations is written as: $\mathbf{Xw} = \mathbf{Y}$, where $\mathbf{X} \in \mathbb{R}^{3\times5}$ and $\mathbf{Y} \in \mathbb{R}^{3\times4}$, How many equations are there in this set of equations?

✓A. 3
B. 4
C. 5
D. 6

Question #: 4

Which of the following is/are true when the below input is given.

$\mathbf{X}=\begin{bmatrix}4 & -5 & 0.3\\ 1 & 12 & -3\end{bmatrix}$

I. $\mathbf{X}$ is right invertible
II. Consider $\mathbf{X}$ is the input data matrix to the linear system $\mathbf{Xw} = \mathbf{Y}$, the system is an under-determined system
III. Consider $\mathbf{X}$ is the input data matrix to the linear system $\mathbf{Xw} = \mathbf{y}$, if performing ridge regression, the solution can be written as $\hat{\mathbf{w}} = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{y}$.
IV. Consider $\mathbf{X}$ is the input data matrix to the linear system $\mathbf{Xw} = \mathbf{y}$, if performing ridge regression, we need to use primal form.

✓A. I and II
B. II and III
C. I, II and IV
D. II, III and IV

Question #: 5

The table below shows the values of feature $\mathbf{x}$ (each a 2-dimensional vector) and their corresponding target values y:

x	y
[3, 2]	4.0
[4, 5]	5.0
[5.3, 1]	3.5
[8, 7]	9.0

We assume $\mathbf{Xw} = y$. Based on the least squares regression framework, please select the correct option below:
I. The estimated parameter vector $\mathbf{w}^*$ can be found by minimizing the sum of squared errors between the predicted output $y\prime$ and the target output y over all training samples.
II. This problem is an example of learning a vector function.
III. When using least squares regression, there will be two unknown parameters to estimate, and this is an over-determined system.
IV. The squared error summed across all training samples is often called cost function.

A. I and II
✓B. I and IV
C. I, III and IV
D. II and IV
E. None of the above

Question #: 6

This question relates to the understanding of linear systems and partial derivatives. Which of the following statements is/are correct?

I. In an over-determined linear system, the number of equations is greater than the number of unknown parameters.

II.If $f : \mathbb{R}^3 \to \mathbb{R}^h$ is a vector-valued function, then the Jacobian of f with respect to $\mathbf{x}$ is an $h \times 3$ matrix.

III.For a function $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, the partial derivative of f with respect to $\mathbf{x}$ is the constant matrix A, regardless of $\mathbf{b}$.

IV.Consider the linear system $\mathbf{Xw} = \mathbf{y}$, where $\mathbf{X} \in \mathbb{R}^{m \times d}$, $\mathbf{w} \in \mathbb{R}^{d \times 1}$, and $\mathbf{y} \in \mathbb{R}^{m \times 1}$. If $\mathbf{m < d}$, the system is over-determined.

A. I, II
✓B. I, II and III
C. II, III and IV
D. II, III

Question #: 7

Given $f(a) = -2a^2 + 6$, where $a \in [-1, 1]$, therefore $\arg\max_a f(a) = 0$.

✓A. True
B. False

Question #: 8

Suppose we have 7 training samples, we would like to perform multi-class classification to predict **y** (3 classes) using input with 5 features, using the optimal linear regression model with an offset term. In this case, the estimated parameter matrix **W** is of dimension 3x6.

A. True
✓B. False

Question #: 9

Consider that we have the inputs below and want to perform 2nd order polynomial regression, there will be 9 unknown parameters to be estimated, and we need to do the right inverse.

$$\{x_1 = 3, x_2 = 0.9, x_3 = -11\} \rightarrow \{y = 3\}$$
$$\{x_1 = 23, x_2 = 1, x_3 = -6\} \rightarrow \{y = 4\}$$

A. True
✓B. False

Question #: 10

The system has no exact solution but an approximated solution is available using the left inverse.

$$\begin{bmatrix} 5 & -6 \\ 2 & 1 \\ 3 & 7 \\ -6 & 11 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 3 \\ -5.3 \\ 9.2 \\ 1 \end{bmatrix}$$

✓A. True
B. False

Question #: 11

A 4-input 2nd order polynomial regression system has 15 parameters.

✓A. True
B. False

Question #: 12

Which of the following properties must a function

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

satisfy to be considered linear function?

A.

$$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) \text{ only.}$$

B.

$f(a\mathbf{x}) = af(\mathbf{x})$ and $f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x})f(\mathbf{y})$ for any scalar a.

✓C.

$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ and $f(a\mathbf{x}) = af(\mathbf{x})$ for any scalar a.

D.

$$f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x})f(\mathbf{y}) \text{ only.}$$

Question #: 13

A financial institution wants to develop a machine learning model to analyze customer transactions and detect fraudulent activity. The dataset includes customer demographic information, transaction history, and behavioral patterns. Consider the following scenarios and select the correct answers:

The institution has a labeled dataset with past transactions, where each transaction is labeled as "Fraudulent" or "Legitimate." Which type of learning should be used?

A) Supervised Learning
B) Unsupervised Learning
C) Reinforcement Learning

Your answer: 1

The institution wants to group customers into different spending behavior categories without predefined labels. Which type of learning is best suited?

A) Supervised Learning
B) Unsupervised Learning
C) Reinforcement Learning

Your answer: 2

The dataset records customer risk ratings as "Low Risk," "Medium Risk," and "High Risk." What type of data does this represent?

A) Nominal Data
B) Ordinal Data
C) Interval Data
D) Ratio Data

Your answer: 3

The institution finds that fraudulent transactions occur more frequently in certain regions. They decide to encode geographical regions (e.g., "North," "South," "East," "West") as a vector. What is the most appropriate encoding technique?

A) One-Hot Encoding
B) Linear Scaling
C) Z-score standardization

Your answer: 4

1. Choice of: A) | B) | C) - Correct Answer:A)
2. Choice of: A) | B) | C) - Correct Answer:B)
3. Choice of: A) | B) | C) | D) - Correct Answer:B)
4. Choice of: A) | B) | C) - Correct Answer:A)

Question #: 14

Unsupervised learning is one type of deductive reasoning.

A. True
✓B. False

Question #: 15

For a random experiment, an event may never contain the same number of outcomes as the sample space does.

A. True
✓B. False

Question #: 16

The Product rule, Pr(X=x, Y=y) =Pr(Y=y|X=x) * Pr(X=x), holds only when X and Y are independent random variables.

A. True
✓B. False

Question #: 17

The elements in a vector are unordered. However, the elements in a matrix are ordered.

A. True
✓B. False

Question #: 18

Randomized Controlled Trial (RCT) can be used for extracting features for classification.

A. True
✓B. False

Question #: 19

We are given the average daily temperature of Singapore in the past 20 years, and we are asked to learn a machine learning model that predicts the average temperature next Friday. This task, by nature, is a clustering task.

A. True

✓B. False

Question #: 20

Artificial intelligence relies on using unsupervised learning for making predictions.

A. True
✓B. False

Question #: 21

Please select the correct option.

i) Data wrangling serves as the solution to Simpson's paradox.

ii) Data wrangling is one step within the process of data cleaning.

iii) Nearest neighbour classifier works for binary classification (i.e., 2-class classification) only, but not for multi-class classification.

iv) Given two different sets of data samples, it is not possible that these two sets end up having the same regression line, the same mean, and the same variance.

A. i)
B. ii)
C. iii)
D. iv)
E. ii) and iv)
F. i), ii) and iv)
G. ii) and iii)
✓H. None of the others is correct.

Question #: 22

Please select the correct option.

i) Data imputation applies to supervised learning only, but not to unsupervised learning.

ii) Data visualization applies to supervised learning only, but not to unsupervised learning.

iii) Binary coding applies to reinforcement learning only, but not to supervised learning.

iv) Feature extraction applies to supervised learning only, but not to unsupervised learning.

A. i)
B. ii)
C. iii)
D. iv)
E. i), ii), and iv)
F. i) and iv)
G. i) and ii)
H. ii), iii), and iv)
✓I. None of others is correct.

Question #: 23

A self-driving car relies on a machine-learning system to detect road signs and classify them into four categories: "Stop," "Speed Limit," "Yield," and "No Entry."

We have collected a dataset of images, containing 500 images for each category (a total of 2,000 images). We then ask an expert to review all 2,000 images. The expert eventually identifies 200 poor-quality images and suggests that we remove these images from the dataset. We follow the suggestion and end up with 1,800 images.

Which of the following statement(s) is/are correct?

i) For each image, the machine-learning system extracts the shape, color, and edge patterns of the signs, as well as the text in the image. This step can be considered unsupervised learning.

ii) Given a new sign image, the system predicts one of the four categories. This can be considered a regression task.

iii) The category of the sign (i.e., "Stop," "Speed Limit," "Yield," and "No Entry") belongs to interval data.

iv) The expert inspection process can be considered an unsupervised-learning step.

A. i)
B. ii)

C. iii)
D. iv)
E. i) and ii)
F. i), ii), and iii)
G. ii) and iii)
H. i), ii), iii), and iv)
✓I. None of the others is correct.

Question #: 24

A company has 600 employees divided into two departments:

Department A: 350 employees
Department B: 250 employees

Each department has two teams:

Department A consists of:
Team A-1: 200 employees
Team A-2: 150 employees

Department B consists of:
Team B-1: 120 employees
Team B-2: 130 employees

At the end of the year, the company conducts a performance review where each employee either gets promoted ('p') or not ('n').
Promotion Results:

In Department A: 210 employees get promoted, among whom 120 are from Team A-1.
In Department B: 140 employees get promoted, among whom 60 are from Team B-1.

1. What is the probability that a randomly selected employee gets promoted, i.e., P(X='p')?

Your Answer 1 (Write your answer to three decimal places, between 0 and 1.)

2. Given that an employee was not promoted, what is the probability that they are from Team A-2, i.e., P(T= 'A-2' | X='n')?

Your Answer 2 (Write your answer to three decimal places, between 0 and 1.)

3. Which team has the highest promotion rate?

Your Answer 3

1. Range - Min:0.29 Max:0.31
2. Range - Min:0.23 Max:0.25
3. Choice of: Team A-1 | Team A-2 | Team B-1 | Team B-2 - Correct Answer:Team B-2