M1

Ques 1: Suppose you want to show a relationship between two quantitative features ‘height’ and ‘weight’ in nutri data. What will be the best and correct way to plot this relationship in Python?

1. plt.scatter(nutri.height, nutri.weight, s=12, marker='o')

   plt.xlabel('height') plt.ylabel('weight')

   plt.show()

2. plt.hist(nutri.height, nutri.weight, s=12, facecolor="cyan", edgecolor='black', linewidth=1)

   plt.xlabel('height') plt.ylabel('weight')

   plt.show()

3. plt.boxplot(nutri.height), width=width, vert=False

   plt.xlabel('height')

   plt.show()

4. plt.boxplot(nutri.height, nutri.weight, s=12, marker='o')

   plt.xlabel('height')

   plt.ylabel('weight')

   plt.show()

Answer: 1

M2

Ques 1: Suppose you want to learn how a deep learning model fits the training data. For this, you have to calculate training loss. How can you calculate the training loss?

1. The training loss is calculated by the sum of errors for each example in the training set.
2. The training loss is calculated by the sum of the multiplication of errors for each example in the validation set.
3. The training loss is calculated by the sum of errors for each example in the validation set.
4. The training loss is calculated by taking the multiplication of errors for each example in the training set.

Answer: 1

Ques 2: Which of the following are the major goals for modeling data in the field of data science?

1.

1. To accurately predict some future quantity of interest given some observed data.
2. To discover unusual or interesting patterns in the data.
3.
    1. To accurately know some past quantity of interest given some observed data.
    2. To store data in a more structured way.
4.
    1. To accurately predict some future quantity of interest given some observed data.
    2. To store data in a more structured way.
5.
    1. To accurately know some past quantity of interest given some observed data.
    2. To discover unusual or interesting patterns in the data.

Answer: 1

M3

Ques 1: The Monte Carlo simulation can be performed using any appropriate tool such as machine learning applications like TensorFlow and PyTorch for smaller sample sets, even Excel. Which of the following represents the correct basic order of the mathematical steps involved in the Monte Carlo method?

1.

1. Identify the dependent variable and the independent variables viz, input/risk, or predictor variables.
2. Run simulations for a fixed number of times.
3. Specify the probability distribution for the independent variables.

2,

1. Identify the dependent variable and the independent variables viz, input/risk, or predictor variables.
2. Specify the probability distribution for the independent variables.
3. Run simulations until you generate enough possible values for the independent variable.

3,

4. Specify the probability distribution for the independent variables.
5. Identify the dependent variable and the independent variables viz, input/risk, or predictor variables.

6. Run simulations until you generate enough possible values for the dependent variable.

4.

1. Run simulations until you generate enough possible values for the independent variable.
2. Specify the probability distribution for the independent variables.
3. Identify the dependent variable and the independent variables viz, input/risk, or predictor variables.

Answer: 2

M4

Ques 1: Which algorithm can be used to maximize the complicated (log) likelihood functions through the introduction of auxiliary variables?

1. Row-action maximum likelihood algorithm
2. Maximization algorithm
3. Newton-Raphson algorithm
4. Expectation-maximization algorithm

Answer: 4

Ques 2: Suppose you want to merge the data points into larger and larger clusters until all the points have been merged into a single cluster. Also, you need to divide the data set into smaller and smaller clusters. Which of the following approaches will be used above?

1.

1. Bottom-up approach also called agglomerative clustering.
2. Top-down approach also called divisive clustering.

2,

1. Top-down approach also called divisive clustering.
2. Bottom-up approach also called agglomerative clustering.

3.

1. Bottom-up approach also called divisive clustering.
2. Top-down approach also called agglomerative clustering.

4.

1. Top-down approach also called agglomerative clustering.
2. Bottom-up approach also called divisive clustering.

Answer: 1

M5

Ques 1: Suppose you want to predict a continuous dependent variable. Which of the following methods is best suited for it?

1. Linear and Logistic Regressions
2. Cyber Regression
3. Linear Regression
4. Logistic Regression

Answer: 3

Ques 2: What is the correct order of steps to follow to perform analysis of Variance (ANOVA)?

a. Look up F from the table

b. Calculate the Sum of Squares

c. Calculate the Degrees of Freedom (df)

d. Set up the null and alternate hypothesis and the Alpha

e. Calculate a Summary Table and Calculate the F statistic

f. Calculate the Mean Squares

g. Calculate all the means

Options:

1. g > b > d > c > f > a > e
2. g > d > b > c > e > f > a
3. b > g > d > c > f > e > a
4. g > d > b > c > f > e > a

Answer: 4

Ques 3: Suppose you have found the observed value and predicted value in the regression analysis and you have calculated the residual by the formula: residual = observed value - predicted value. Which of the following statements is true about residuals in regression analysis?

1. There is no such rule for residuals.
2. Mean of residuals is always less than zero.
3. Mean of residuals is always greater than zero.
4. Mean of residuals is always zero.

Answer: 4

Ques 4: We are using a student survey data set survey.csv from the book's GitHub site, which contains measurements such as height, weight, sex, etc., from a survey conducted among n=100 university students. Suppose we wish to investigate the relation between the shoe size (explanatory variable) and the height (response variable) of a person. First, we load the data and draw a scatter plot of the points (height versus shoe size). We analyze the data

through the simple linear regression model Yi=β0+β1xi+ciY_i = \beta_0 + \beta_1 x_i + c_iYi=β0+β1xi+ci, i=1,…,ni = 1, \ldots, ni=1,…,n in statsmodels. This is performed via the OLS method. A summary of the results is obtained with the method summary. In the summary, what do these outputs indicate? R square 2. AIC 3.std error

Options:

1. (a) Coefficient of determination R
   (b) Minus two times the log-likelihood Plus two times the number of model parameters

   (c) Mean of the estimators of the regression line.

2. (a) Coefficient of determination R*R
   (b) Minus two times the log-likelihood Plus two times the number of model parameters

   (c) Standard deviation of the estimators of the regression line.

3. (a) Coefficient of determination R*R

   (b) Minus five times the log-likelihood Plus five times the number of model parameters

   (c) Standard deviation of the estimators of the regression line.

4. (a) Coefficient of determination R*R
   (b) Minus two times the log-likelihood Plus two times the number of model parameters

   (c) Mean of the estimators of the regression line.

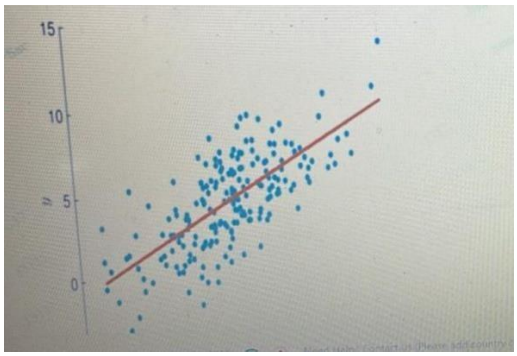
Answer: 2

Ques 5: Which of the following methods is used for predictive modeling in machine learning?

1. Classification
2. Clustering
3. Regression
4. Transfer learning

Answer: 3

Ques 6: The given image depicts the data from a simple linear regression model. What is the relationship depicted in this image?



1. The relationship between one independent variable and two dependent variables using a straight line.
2. The relationship between one independent variable and one dependent variable using a straight line.
3. The relationship between two dependent variables and one dependent variable using a straight line.
4. The relationship between two independent variables and two dependent variables using a straight line.

Answer: 2

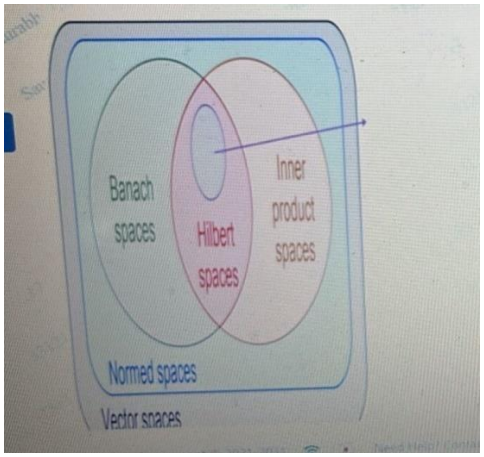
Ques 7: The following picture depicts data taken from a model. Which of the following is the correct model?

1. Bayesian linear regression
2. Polynomial regression
3. Simple linear regression model
4. Multiple linear regression models

Answer: 3

Question 1

In the given image, the circular region in the Hilbert space indicates?



Select an option:

    - Metric spaces
    - Topological spaces
    - Vector spaces
    - Reproducing kernel Hilbert spaces
- Ans: **Answer: Reproducing kernel Hilbert spaces**

Question 2

Suppose you have two vectors x1 and x2 and you have found out the dot product of two vectors x1 and y1 in the form K(x1,x2)=x1.X2. Which kernel defines the dot product?

- **Answer: Linear kernel**
- Select an option:
    - Polynomial kernel
    - Linear kernel
    - Polynomial kernel
    - Gaussian kernel

Question 3

Ridge regression is a type of?

- **Answer: Linear regression**
- Select an option:
    - Linear regression
    - Logistic regression
    - Lasso regression
    - Polynomial regression

M8

Question 1

Following are the random steps to use bagging for a regression tree. In this, we compare the decision tree estimator with the corresponding bagged estimators. We use the R 2 metric (coefficient of determination) for comparison.

a. splitting to train/test set

b. training

c. out of bag loss estimation

d. creating regression problem.

c. testing.

f. bagging prediction

g bagging construction

h. creating regression problem

- **Answer: a>b>e>h>g>f>c**
- Select an option:
    1. h>a>e>g>c>f>b
    2. a>b>e>h>g>f>c
    3. a>b>c>d>e>g>h>f
    4. h>a>b>e>g>f>c

Question 2

Which of the ensemble learning methods is used to reduce variance within a noisy data set?

- **Answer: Bagging**
- Select an option:
    1. Boosting
    2. Bagging
    3. Stacking
    4. Blending

Question 3

Suppose you have calculated correctly predicted labels and also the total number of predicted and actual labels. What can one derive from the ratio of correctly predicted labels and the total number of predicted and actual labels?

- **Answer: Accuracy**
- Select an option:
    - Hamming box
    - Precision
    - Recall
    - Accuracy

Question 4

Below is a list of commonly used metrics. For simplicity, we call an object whose actual class is_____object. Match the columns.

- **Answer: a-4, c-3, b-2, d-1**
- Select an option:
    1. a-4, c-2, b-3, d-1
    2. a-3, c-4, b-2, d-1
    3. a-1, b-2, c-3, d-4
    4. a-4, c-3, b-2, d-1

Question 5

Python provides a lot of tools for implementing Classification and Regression. The most popular open-source Python data science library is scikit-learn. What is the correct order of supervised machine learning?

- **Answer: e>b>a>c>d**
- Select an option:
    1. a>c>b>e>d
    2. e>a>c>b>d
    3. a>b>c>e>d
    4. e>b>a>c>d

Question 6

Which algorithms are used to predict the output of the categorical data?

- **Answer: Classification algorithms**
- Select an option:
    - Random forest algorithm
    - KNN algorithm
    - Linear regression
    - Classification algorithms

Question 7

Which of the following represents the correct features of multilabel and hierarchical classification?

- **Answer: In multilabel classification, the classes do not have to be mutually exclusive and in hierarchical classification, the relations between classes/labels are modeled via a tree or a directed acyclic graph.**
- Select an option:
    1. In multilabel classification, the classes do not have to be mutually exclusive and in hierarchical classification, the relations between classes/labels are modeled via a tree or a directed cyclic graph
    2. In multilabel classification, the classes have to be mutually exclusive and in hierarchical classification, the relations between classes/labels are modeled via a tree or a directed acyclic graph
    3. In multilabel classification, the classes have to be mutually exclusive and in hierarchical classification, the relation between classes/labels are modeled via a tree or a directed cyclic graph
    4. In multilabel classification, the classes have to be mutually exclusive and in hierarchical classification, the relations between classes/labels are modeled via a tree or a directed acyclic graph

M9

Question 1

What is the last layer of deep learning?

- **Answer: Output layer**
- Select an option:
    1. Output layer
    2. Hidden layer
    3. Pooling layer
    4. Input layer

Question 2

Suppose that you are working on the neural network you come across a continuous function. These functions can be represented by a neural network which has three layers. What is the function of hidden layers of the neural network?

- **Answer: The hidden layer performs non-linear transformation of the input entered into the networks**
- Select an option:
    1. The hidden layer coalesces and concretely produces the end result
    2. The hidden layer is responsible for the leaks in the server system
    3. The hidden layer performs non-linear transformation of the input entered into the networks
    4. The hidden layer brings the initial input data into the system for further processing by subsequent layers of artificial neurons

Question 3

Which of the following open-source machine learning libraries of Python can easily exploit any graphics processing unit (GPU) for accelerated computation?

- **Answer: PyTorch**
- Select an option:
    1. Pandas
    2. Numpy
    3. Scipy
    4. Pytorch

Question 4

Deep learning is a field of software engineering that has accumulated a massive amount of study over the years. There are several neural network designs that have been developed for use with diverse data kinds. Feedforward neural networks are so named because all information flows in a forward manner only. Which of the following are correct practical uses of feedforward neural networks?

- **Answer: a, b, c, & d**
- Select an option:
    1. only d

2. a, b, c & d
3. a, b & c
4. a & d

M10

Question 1

What technique is used by linear algebra in dimensionality reduction?

- **Answer: Singular value decomposition**
- Select an option:
  - Singular value decomposition
  - Factorization analysis
  - Rationalizing analysis
  - Principle component analysis

Question 2

Suppose you are given a set of vectors (v1,v2,v3..vk). Which of the following statements justifies that the given set of vector are linearly independent?

- **Answer: $\alpha 1v1+\ldots+\alpha nvn= 0$, then it must hold that $\alpha i=0$ for all i=1,…,n**
- Select an option:
  - $\alpha 1v1 +...+ \alpha nvn=1$, then it must hold $\alpha i =1$ for all i=1,…,n
  - $\alpha 1v1+\ldots+ \alpha nvn= 0$, then it must hold that $\alpha i=0$ for all i=1,…,n
  - $\alpha 1v1+\ldots+ \alpha nvn= 0$, then it must hold that $\alpha i=0$ for all i=2,…,n
  - $\alpha 1v1+\ldots+ \alpha nvn= 0$, then it must hold that $\alpha i=1$ for all i=1,…,n

Question 3

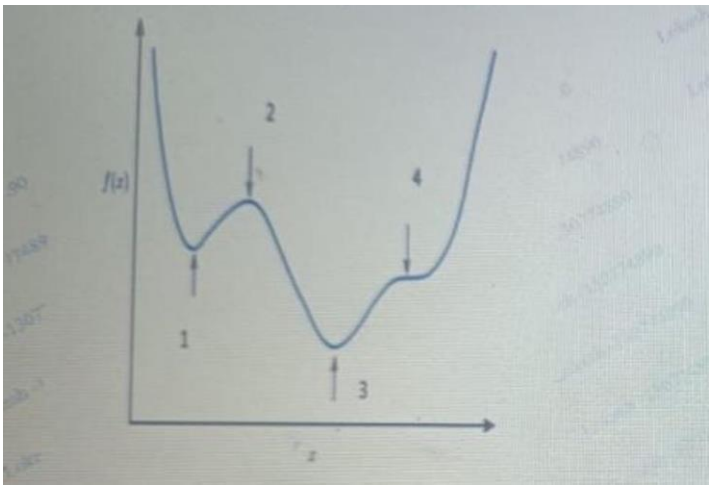
You are given a matrix A. What is the trace of the matrix?

- **Answer: 7** (This requires the matrix to be provided to compute the actual trace. Assuming the question is theoretical.)
- Select an option:
  - 0
  - 8
  - 3
  - 7

M11

Question 1

In the given image, what do point 2 and point 3 on the graph indicate?



- **Answer: local minimum global maximum**
- Select an option:
    1. Saddle point, Local maximum
    2. Local minimum, Local maximum
    3. Local maximum, Global maximum
    4. Local minimum, Saddle point

Question 2

Suppose you are working on a maximization problem. You want to convert this problem into a minimization problem. How will you convert this problem into a minimization problem?

- **Answer: Via the equivalence maxx f(x)= - minx - f(x)**
- Select an option:
    1. Via the equivalence maxx f(x)= - minx - f(x)
    2. Via the equivalence maxx f(x)= minx – f(x)
    3. Via the equivalence maxx f(x)= minx + f(x)
    4. Via the equivalence maxx f(x) + minx + f(x)

M12

Question 1

In Mohali, it's snowing one third of the days. Given that it is snowing, there will be heavy traffic with probability 1/12, and given that it is not snowing there will be heavy traffic with probability 1/14. If it's snowing and there is heavy traffic, a hotel worker arrives late for work with probability 1/12. On the other hand, the probability of being late is reduced to 1/18 if it is not snowing and there is no heavy traffic. In other situations (snowing and no traffic, not snowing and traffic), the probability of being late is 0.25. He picks a random day.

1. **What is the probability that it's not snowing and there is heavy traffic and he is not late?**
2. **Given that he arrived late at work, what is the probability that it snowed that day?**

- **Answer: 1--1/8 2—11/48**
- Select an option:

1. 1/8
   6/11
2. 1/8

   11/48
3. 1/12

   6/11
4. 1/12

   5/11

Question 1

Which of the following are correct examples of independent events?

- **Answer: A, b, c & d**
- Select an option:
  1. D & e
  2. A, b, c & e
  3. A, b & c
  4. A, b, c & d

M13

Question 1

Suppose you are given an array a=[2,3,5,7,11,13,17,19,23]. How can you print [19,23] in your output window in Python?

- **Answer: Slice notation of a[-2:]**
- Select an option:
  1. Slice notation of a[-2:]
  2. Slice notation of a[8:9]
  3. Slice notation of a[7:8]
  4. Slice notation of a[:-2]

Question 2

Which Python module/package provides an ecosystem for data analysis and which one provides libraries for HTTP requests and interfacing with the web?

- **Answer: Pandas and requests**
- Select an option:
    1. Pandas and requests
    2. Pandas and scipy
    3. PyTorch and scipy
    4. Numpy and matplotlib

Ques 3. Suppose you are given this code in python:

```
x=[1,4,5]

def change_list (y):

y.append (10)

y[0]=0

y = [2,3,5]

return sum(y)

print(change_list (x))

print(x)
```

What will be it's correct output?

1. 19
   [3,2,3,4,10]
2. 19
   [0,2,3,10]
3. 10
   [0,4,5,10]
4. 10
   [0,2,3,4,10]

Ans: 10

[0,4,5,10]