**MCQs**

1. For calculation of principal components which of the thing is used to multiple the original data or scaled data?
   1. Eigen Vectors Matrix
   2. Eigen Values Matrix
   3. Identity Matrix
   4. Transpose of the Covariance matrix

Ans: a

1. What is not true about the calculated principal component columns?
   1. All the columns are orthogonal to each other
   2. The variances of the columns are in the decreasing order with variance of the first column being highest
   3. The columns as features can be used in supervised learning algorithm replacing the original data of features
   4. All the columns are highly correlated with each other.

Ans: d

1. A good ML project that deserves to be funded can be that which
   1. Uses best machine learning algorithm
   2. Is complicated
   3. Serves the purpose of optimizing the profit or loss
   4. Uses ensembled learning

Ans: C

1. Competitive advantage refers
   1. to factors that allow a company to produce goods or services better or more cheaply than its rivals
   2. to the company reducing prices of its goods
   3. to the company increasing prices of its goods
   4. to employees competing with each other

Ans: a

1. The nominal scale of the data is the scale in which
   1. Numbers with floating point are there
   2. Categories are there with no superiority or inferiority with each other
   3. Categories are there with some hierarchy of superiority or inferiority
   4. Numbers with integers are there

Ans: b

1. The ordinal scale of the data is the scale in which
   1. Numbers with floating point are there
   2. Categories are there with no superiority or inferiority with each other
   3. Categories are there with some hierarchy of superiority or inferiority
   4. Numbers with integers are there

Ans: c

1. What is not acceptable for any machine learning algorithm to be implemented?
   1. Variables (columns) in the data with different scales
   2. Variables containing categorical values
   3. Data with some missing values
   4. No numerical column in the data

Ans: c

1. Skewness gives
   1. Measure of asymmetry in the data
   2. Measure of flatness in the data
   3. Measure of variation in the data
   4. Measure of uncertainty in the data

Ans: a

1. Kurtosis gives
   1. Measure of asymmetry in the data
   2. Measure of flatness in the data
   3. Measure of variation in the data
   4. Measure of uncertainty in the data

Ans: b

1. The normalize option in pandas.crosstab function
   1. Calculates the proportions
   2. Calculates the percentages
   3. Calculates the totals
   4. Calculates the grouped aggregate values

Ans: a

1. The .groupby method in pandas object
   1. Calculates the proportions
   2. Calculates the percentages
   3. Calculates the totals
   4. Calculates the grouped aggregate values

Ans: d

1. What are the different types of random variables?
   1. Discrete and Positive
   2. Discrete and Negative
   3. Discrete and Continuous
   4. Continuous and positive

Ans: c

1. In a column of amount values there are many values between the range 1000 to 10,000. Very few values are there which are bigger than 200,000.
   1. You will call values bigger than 200,000 as inliers
   2. You will call values between range 1000 to 10,000 as inliers
   3. You will call values bigger than 200,000 as outliers
   4. You will call values between range 1000 to 10,000 as outliers

Options:

1. I and IV are correct
2. Only III is correct
3. Only III is correct
4. II and III are correct

Ans: d

1. Which is the measure used in the formula of Pearson’s correlation?
   1. Variance
   2. Mean
   3. Skewness
   4. Covariance

Ans: d

1. Which of the plot depicts percentages more correctly?
   1. Pie Chart
   2. Histogram
   3. Boxplot
   4. Density Plot

Ans: a

1. Class Intervals and their frequencies are plotted in which of the following graphs?
   1. Pie Chart
   2. Histogram
   3. Boxplot
   4. Density Plot

Ans: b

1. In package dplyr, which two functions can make the data longer?
   1. spread and pivot\_longer
   2. gather and pivot\_wider
   3. gather and pivot\_longer
   4. spread and pivot\_wider

Ans: C

1. In package dplyr, which two functions can make the data wider?
   1. spread and pivot\_longer
   2. gather and pivot\_wider
   3. gather and pivot\_longer
   4. spread and pivot\_wider

Ans: d

1. Which of the following can be best example of strong negative correlation?
   1. Correlation coefficient being -0.9
   2. Correlation coefficient being 0.9
   3. Correlation coefficient being 0
   4. Correlation coefficient being 1

Ans: a

1. Suppose that P(A) = v , P(B) = u and given that events A and B are independent events. What can be P(A U B)?
   1. v\*u
   2. u + v – u\*v
   3. u + v
   4. u / (u+v)

Ans: b

1. Suppose that P(A) = v , P(B) = u and given that events A and B are independent events. What can be P(A and B)?
   1. v\*u
   2. u + v – u\*v
   3. u + v
   4. u / (u+v)

Ans: a

1. What can be said to be a sample space?
   1. Set of all the impossible events
   2. Set of all the possible events
   3. Set of all the independent events
   4. Set of all the mutually exclusive events

Ans: b

1. Given is the joint probability distribution between A and B:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | B1 | B2 | B3 | Total |
| A1 | 0.066474 | 0.060694 | 0.130058 | 0.257225 |
| A2 | 0.031792 | 0.066474 | 0.225434 | 0.323699 |
| A3 | 0.104046 | 0.031792 | 0.283237 | 0.419075 |
| Total | 0.202312 | 0.15896 | 0.638728 | 1 |

What is the value of P(B1)?

* + 1. 0.15896
    2. 0.419075
    3. 0.202312
    4. 0.225434

Ans: C

1. Given is the joint probability distribution between A and B:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | B1 | B2 | B3 | Total |
| A1 | 0.066474 | 0.060694 | 0.130058 | 0.257225 |
| A2 | 0.031792 | 0.066474 | 0.225434 | 0.323699 |
| A3 | 0.104046 | 0.031792 | 0.283237 | 0.419075 |
| Total | 0.202312 | 0.15896 | 0.638728 | 1 |

What is value of P(A1 and B3)?

1. 0.130058
2. 0.066474
3. 0.031792
4. 0.257225

Ans: a

1. Given is the joint probability distribution between A and B:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | B1 | B2 | B3 | Total |
| A1 | 0.066474 | 0.060694 | 0.130058 | 0.257225 |
| A2 | 0.031792 | 0.066474 | 0.225434 | 0.323699 |
| A3 | 0.104046 | 0.031792 | 0.283237 | 0.419075 |
| Total | 0.202312 | 0.15896 | 0.638728 | 1 |

What can be the value of P(A1 | B2)?

* 1. 0.060694 / 0.257225
  2. 0.060694 / 0.15896
  3. 0.15896
  4. 0.060694

Ans: b

1. Given is the joint probability distribution between A and B:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | B1 | B2 | B3 | Total |
| A1 | 0.066474 | 0.060694 | 0.130058 | 0.257225 |
| A2 | 0.031792 | 0.066474 | 0.225434 | 0.323699 |
| A3 | 0.104046 | 0.031792 | 0.283237 | 0.419075 |
| Total | 0.202312 | 0.15896 | 0.638728 | 1 |

What can be the value of P(B2 | A1)?

* 1. 0.060694 / 0.257225
  2. 0.060694 / 0.15896
  3. 0.15896
  4. 0.060694

Ans: a

1. For any Multi-Class classification problem with three possible outcomes. What can be number of columns of the output of function predict\_proba( )?
   1. 1
   2. 2
   3. 3
   4. 4

Ans: c

1. The common thing in Hypothesis testing and Machine Learning is that
   1. We do predictions in both
   2. We do not do predictions in both
   3. **We conclude on population whether a model or a test**
   4. We don’t draw any conclusion in both
2. You want to find out the causes of satisfaction of customers. In other words, you want to build a model for predicting whether a particular customer might be satisfied or not. Which type of algorithm should you plan to use?
   1. Association Rules
   2. Regression Type Supervised learning
   3. **Classification Type Supervised Learning**
   4. Any Unsupervised Learning
3. Which of the algorithms can be categorized as Prescriptive Analytics algorithm?
   1. **Linear Programming Problem**
   2. Cluster Analysis
   3. Linear Regression
   4. Association Rules
4. Which can be the example of supervised learning algorithm?
   1. Linear Programming Problem
   2. Cluster Analysis
   3. **Linear Regression**
   4. Association Rules
5. Which is the example of unsupervised learning algorithm?
   1. Linear Programming Problem
   2. T-test and ANOVA
   3. Linear Regression
   4. **Association Rules**
6. Any Machine Learning projects starts with
   1. Python coding
   2. Software Development
   3. **Defining the scope**
   4. Decision Making
7. What is the purpose of evaluating any machine learning model?
   1. **To convince ourselves that this model will be able to predict good**
   2. As a formality
   3. Decision Making
   4. Python coding
8. You have a task of running a code which may take long time to execute because of its complexity and big volume of the data. This might not be a repetitive task. By what means can you prefer to get it done as soon as possible?
   1. Purchase high end hardware
   2. Contact System Administrator for help
   3. **Avail any service of cloud computing platform**
   4. Outsource it to a company which has high end hardware infrastructure
9. In the machine learning problem of cash consumption in ATM centres, latitude and longitude of location of ATM centre can be said to be
   1. **Independent Variables or Features**
   2. Dependent Variable or Response Variable
   3. Metrics
   4. Something taken just as information
10. In the machine learning problem of cash consumption in ATM centres, the amount of cash consumption of the ATM centre can be said to be
    1. Independent Variables or Features
    2. **Dependent Variable or Response Variable**
    3. Metrics
    4. Something taken just as information