# Data Mining

## (Multiple Choice Questions for Quiz)

Prepared by: P V S Maruthi Rao |

## Topics

| Topic | Number of MCQs |
|---|---|
| Exploring Data | 20 |
| Classification | 8 |

---

## Exploring Data

| Exploring Data | Question 1 |
|---|---|
| What are the primary objectives of Data Mining tasks? | |

| | | | |
|---|---|---|---|
| A | Use some variables to predict unknown or future values of other variables<br>Find human-interpretable patterns / associations that describe the data. | B | Provide distributed and parallel computing infrastructure for data processing. |
| C | Provide database queries on complex datasets. | D | All of the above |

| | |
|---|---|
| Answer | **A** |
| Remarks | |

---

| Exploring Data | Question 2 |
|---|---|
| In the data preprocessing, regression technique cannot be used for smoothing the noise in the data. | |

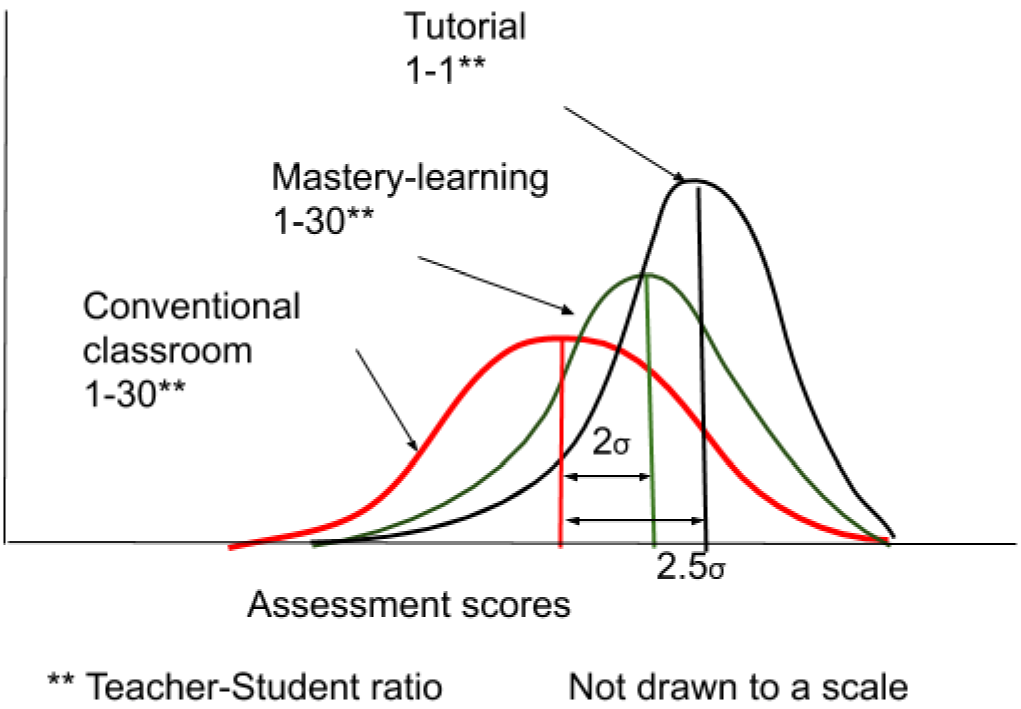
| | | | |
|---|---|---|---|
| A | True | B | False |
| C | - | D | - |

| | |
|---|---|
| Answer | **B** |
| Remarks | Regression technique can be used in smoothing the noise in the data. |

| **Exploring Data** | **Question 3** |
|---|---|
| The type of value of an attribute of a Data object can be of | |

| | | | |
|---|---|---|---|
| A | Nominal only | B | Nominal, Binary only |
| C | Nominal, Binary, Ordinal only | D | Nominal, Binary, Ordinal, Numeric |

| | |
|---|---|
| Answer | **D** |
| Remarks | |

| **Exploring Data** | **Question 4** |
|---|---|
|  <br>The above diagram is Bloom's 2-sigma model. Which of the following are correct? | |

| | | | |
|---|---|---|---|
| A | Mastery-learning is a good choice as it shows 2$\sigma$ improvement in the assessment scores with the same teacher-student ratio. | B | Tutoring is very effective but it is not scalable (1-1 teacher-student ratio). |
| C | A and B above | D | None of the above |

| | |
|---|---|
| Answer | **C** |
| Remarks | Students' understanding on Standard deviation is tested. |

| **Exploring Data** | **Question 5** |
|---|---|
| The basic statistical measures for central tendency of data include mean, weighted mean, median, and mode. | |

|  |  |  |  |
|---|---|---|---|
| A | True | B | False |
| C |  | D |  |

|  |  |
|---|---|
| Answer | **A** |
| Remarks |  |

---

| **Exploring Data** | **Question 6** |
|---|---|
| Interquartile Range (IQR), distance between the first and third quartile, is a measure of central tendency. |  |

|  |  |  |  |
|---|---|---|---|
| A | True | B | False |
| C |  | D |  |

|  |  |
|---|---|
| Answer | **B** |
| Remarks | IQR a simple measure of spread that gives the range covered by the middle half of the data.<br>(This can also be a question - True / False) |

---

| **Exploring Data** | **Question 7** |
|---|---|
| IQR a simple measure of spread that gives the range covered by the middle half of the data. |  |

|  |  |  |  |
|---|---|---|---|
| A | True | B | False |
| C |  | D |  |

|  |  |
|---|---|
| Answer | **A** |
| Remarks | IQR a simple measure of spread that gives the range covered by the middle half of the data.<br>IQR = Q3 - Q1 |

---

| **Exploring Data** | **Question 8** |
|---|---|
| What are the following options below are true for Standard deviation ($\sigma$), a measure of the data spread? |  |

|  |  |  |  |
|---|---|---|---|
| A | σ measures spread about the mean and should be used only when the mean is chosen as the measure of center. | B | σ = 0 only when there is no spread, that is, when all observations have the same value. Otherwise σ > 0. |
| C | A and B above | D | None of the above |

| | |
|---|---|
| Answer | **C** |
| Remarks | A and B are the properties of the standard deviation. |

---

| Exploring Data | Question 9 |
|---|---|
| Data preprocessing improves the data quality and makes data mining algorithms efficient and effective. What are the data preprocessing tasks along with Data cleaning? | |

| | | | |
|---|---|---|---|
| A | Data Integration | B | Data Reduction |
| C | Data Transformation | D | All of the above. |

| | |
|---|---|
| Answer | **D** |
| Remarks | |

---

| Exploring Data | Question 10 |
|---|---|
| In general, the Dimensionality Reduction technique of Data Preprocessing is aimed at the following objectives to achieve. | |

| | | | |
|---|---|---|---|
| A | Eliminate irrelevant features and reduce noise. | B | Reduce time and space required in data mining. |
| C | Allow easier visualization | D | All of the above. |

| | |
|---|---|
| Answer | **D** |
| Remarks | |

---

| Exploring Data | Question 11 |
|---|---|
| Redundancy is another important issue that a Data scientist should deal with. An attribute (dimension / variable) may be redundant if it can be "derived" from another attribute or set of attributes.<br>Some redundancies can be detected by correlation analysis. What are the following statements are true?<br>Hint: Let A and B are two attribute vectors. | |

| | | | |
|---|---|---|---|
| A | Computing correlation coefficient (also known as Pearson's product moment coefficient) on numeric data attributes<br>(Attributes A and B are of numeric type.) | B | Correlation analysis on nominal, categorical data using $\chi^2$ tests.<br>(Attributes A and B are of nominal / discrete / categorical type.) |
| C | Variance - covariance analysis on numeric data.<br>(Attributes A and B are of numeric type.) | D | All of the above. |

| | |
|---|---|
| Answer | D |
| Remarks | |

---

| Exploring Data | Question 12 |
|---|---|

The following observations were recorded while conducting an experiment.

```
x = (-3, -2, -1, 0, 1, 2, 3)
y = ( 9,  4,  1, 0, 1, 4, 9)
```

Calculate the correlation between x and y.

| | | | |
|---|---|---|---|
| A | $y_i = x_i^2$ | B | 0 |
| C | x and y are positively correlated | D | x and y are negatively correlated |

| | |
|---|---|
| Answer | **B** |
| Remarks | |

---

| Exploring Data | Question 13 |
|---|---|

Consider the following Matrix A.

$$A = \left[ a_{jk} \right],\ an\ n\ X\ n\ matrix$$

$$Ax = \lambda x$$

$$(A - \lambda I)\, x = 0$$

$$D(\lambda) = det\,(A - \lambda I) = \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdots & \cdots & \vdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix} = 0$$

Which of the following statements are true?

| | | | |
|---|---|---|---|
| A | **x** is known as the Eigenvector of the matrix A. | B | The ***roots*** of the characteristic polynomial of the matrix A are known as **Eigenvalues** of the matrix A |
| C | If **x** is an eigenvector of a matrix A corresponding to an eigenvalue λ, so is k**x** with any k ≠ 0 | D | All of the above. |

| | |
|---|---|
| Answer | **D** |
| Remarks | |

| Exploring Data | Question 14 |
|---|---|

The following Principal Components are calculated from the Iris dataset.

```
Importance of components:
                          Comp.1     Comp.2     Comp.3      Comp.4
Standard deviation     2.0485788 0.49053911 0.27928554 0.153379074
Proportion of Variance 0.9246162 0.05301557 0.01718514 0.005183085
Cumulative Proportion  0.9246162 0.97763178 0.99481691 1.000000000
```

Which of the following statements are true?

| | | | |
|---|---|---|---|
| A | `Comp.4` has highest loading. | B | `Comp.3` has highest loading. |
| C | `Comp.2` has highest loading. | D | `Comp.1` has highest loading. |

| | |
|---|---|
| Answer | **D** |
| Remarks | |

| Exploring Data | Question 15 |
|---|---|

Minkowski distance is a generalization of the Euclidean distance as given below.

$$d(\mathbf{x}, \mathbf{y}) = \left( \sum_{k=1}^{n} |x_k - y_k|^r \right)^{1/r}$$

Where r is a parameter, n is the number of dimensions (attributes) and $x_k$ and $y_k$ are, respectively, the k$^{th}$ attributes (components) or data objects x and y.

For which value of the parameter ‘r’, the d(x,y), Minkowski distance is also known as Supremum or Chebyshev distance.

| | | | |
|---|---|---|---|
| A | r = 1 | B | r = 2 |
| C | Limit r -> ∞ | D | None of the above |

| | |
|---|---|
| Answer | **C** |
| Remarks | |

| Exploring Data | Question 16 |
|---|---|

**p** and **q** are two data objects with binary attributes.

```
p =  1 0 0 0 0 0 0 0 0 0  and
q =  0 0 0 0 0 0 1 0 0 1
```

| The Jaccard Coefficient of the above **p** and **q** is |
|---|

| | | | |
|---|---|---|---|
| A | 0.7 | B | 0 |
| C | -1 | D | None of the above |

| | |
|---|---|
| Answer | **B** |
| Remarks | J = number of 11 matches / number of non-zero attributes<br>J = $(f_{11}) / (f_{01} + f_{10} + f_{11})$<br><br>**p** = 1 0 0 0 0 0 0 0 0 0 and<br>**q** = 0 0 0 0 0 0 1 0 0 1<br><br>$f_{01}$ = 2 the number of attributes where p was 0 and q was 1<br>$f_{10}$ = 1 the number of attributes where p was 1 and q was 0<br>$f_{00}$ = 7 the number of attributes where p was 0 and q was 0<br>$f_{11}$ = 0 the number of attributes where p was 1 and q was 1 |

---

| Exploring Data | Question 17 |
|---|---|
| Mahalanobis distance is a similarity measure that does not take into account variance-covariance between attributes. | |

| | | | |
|---|---|---|---|
| A | True | B | False |
| C | - | D | - |

| | |
|---|---|
| Answer | **B** |
| Remarks | $mahalanobis\ (x, y) = (\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y})$<br>$\mathbf{\Sigma}$ is the variance-covariance matrix |

---

| Exploring Data | Question 18 |
|---|---|
| Let d(**x, y**) be the distance between two points **x** and **y**. Which of the properties of the distance d(**x, y**) hold good if the measure is a **metric**?<br>Hint: Consider Euclidean and Manhattan distance. **z** is also another point in the same space. | |

| | | | |
|---|---|---|---|
| A | d(**x, y**) ≥ 0 for all **x** and **y**.<br>d(**x, y**) = 0 if **x** = **y**. | B | d(**x, y**) = d(**y, x**) for all **x** and **y**. |
| C | d(**x, y**) ≤ d(**x, z**) + d(**z, y**) | D | All of the above |

| | |
|---|---|
| Answer | **D** |
| Remarks | |

---

| Exploring Data | Question 19 |
|---|---|
| What is the best available distance measure to compute the similarities if magnitude between two data objects (points with numeric attributes) is important. | |

| | | | |
|---|---|---|---|
| A | Euclidean distance | B | Cosine Similarity |
| C | Jaccard Coefficient | D | None of the above |

| | |
|---|---|
| Answer | **A** |
| Remarks | |

---

| Exploring Data | Question 20 |
|---|---|
| In an experiment, a student recorded the observations. Surprisingly, each measurement was distinct.<br>What is the **mode** (the central tendency) of the data the student recorded? | |

| | | | |
|---|---|---|---|
| A | 0 | B | No mode |
| C | Multimodal | D | None of the above |

| | |
|---|---|
| Answer | **B** |
| Remarks | |

---

# Classification of Data

… work in progress ...

| Classification of Data | Question 1 |
|---|---|

Let X and Y be a pair of random variables.

a) The joint probability P(X = x, Y = y), refers to the probability that variable X will take on the value x and variable Y take on the value y.

b) A conditional probability is the probability that a random variable will take on a particular value given that the outcome of another random variable is known.
For example, P(Y = y | X = x) refers to the probability that the variable Y will take on the value y, given that the variable X is observed to have the value x.
Conditional probability is denoted by
P(X | Y) = P (X,Y) / P(Y) and
P(Y | X) = P (X,Y) / P(X)

The joint and conditional probabilities for X and Y are related in the following way:
P(X, Y) = P(Y | X) * P(X) = P(X | Y) * P(Y)

| | | | |
|---|---|---|---|
| A | False | B | True |

| | | | |
|---|---|---|---|
| C | | D | |

| | |
|---|---|
| Answer | **B** |
| Remarks | |

---

| **Classification of Data** | **Question 2** |
|---|---|

Let X and Y be a pair of random variables.

c) The joint probability P(X = x, Y = y), refers to the probability that variable X will take on the value x and variable Y take on the value y.

d) A conditional probability is the probability that a random variable will take on a particular value given that the outcome of another random variable is known.
For example, P(Y = y | X = x) refers to the probability that the variable Y will take on the value y, given that the variable X is observed to have the value x.
Conditional probability is denoted by
P(X | Y) = P (X,Y) / P(Y) and
P(Y | X) = P (X,Y) / P(X)

The joint and conditional probabilities for X and Y are related in the following way:
P(X, Y) = P(Y | X) * P(X) = P(X | Y) * P(Y)

Rearranging the above equation, we obtain the Bayes' Theorem.

$$P(Y \mid X) = \frac{P(X \mid Y) * P(Y)}{P(X)}$$

**P(X) can be calculated using Total Probability Theorem as below and is constant for all classes. Is it correct?**

Let X = {$X_1$, $X_2$ … $X_k$}, a set of mutually exclusive and exhaustive outcomes of the random variable X.

$$P(X) = \sum_{i=1}^{k} P(X, Y_i) = \sum_{i=1}^{k} P(X \mid Y_i) P(Y_i)$$

| | | | |
|---|---|---|---|
| A | Correct | B | Incorrect |
| C | | D | |

| | |
|---|---|
| Answer | **A** |
| Remarks | |

---

| **Classification of Data** | **Question 3** |
|---|---|

Let X and Y be a pair of random variables. Usually X denotes the attribute set and Y denotes the class variable.

e) The joint probability P(X = x, Y = y), refers to the probability that variable X will take on the value x and variable Y take on the value y.

f) A conditional probability is the probability that a random variable will take on a

particular value given that the outcome of another random variable is known.
For example, P(Y = y | X = x) refers to the probability that the variable Y will take on the value y, given that the variable X is observed to have the value x.
Conditional probability is denoted by
P(X | Y) = P (X,Y) / P(Y) and
P(Y | X) = P (X,Y) / P(X)

The joint and conditional probabilities for X and Y are related in the following way:
P(X, Y) = P(Y | X) * P(X) = P(X | Y) * P(Y)

Rearranging the above equation, we obtain the Bayes' Theorem.

$$P(Y \mid X) = \frac{P(X \mid Y) * P(Y)}{P(X)}$$

Which is correct of the above Bayes' Theorem?

| | | | |
|---|---|---|---|
| A | The conditional probability, P(Y \| X) is known as **posterior probability**. | B | The conditional probability, P(X \| Y) is known as **class-conditional probability**. |
| C | P (Y) is known as **prior probability** of Y | D | All of the above |

| | |
|---|---|
| Answer | **D** |
| Remarks | |

---

| **Classification of Data** | **Question 4** |
|---|---|

Let X and Y be a pair of random variables. Usually X denotes the attribute set and Y denotes the class variable.

g) The joint probability P(X = x, Y = y), refers to the probability that variable X will take on the value x and variable Y take on the value y.

h) A conditional probability is the probability that a random variable will take on a particular value given that the outcome of another random variable is known.
For example, P(Y = y | X = x) refers to the probability that the variable Y will take on the value y, given that the variable X is observed to have the value x.
Conditional probability is denoted by
P(X | Y) = P (X,Y) / P(Y) and
P(Y | X) = P (X,Y) / P(X)

The joint and conditional probabilities for X and Y are related in the following way:
P(X, Y) = P(Y | X) * P(X) = P(X | Y) * P(Y)

Rearranging the above equation, we obtain the Bayes' Theorem.

$$P(Y \mid X) = \frac{P(X \mid Y) * P(Y)}{P(X)}$$

Is it true of a naïve Bayes classifier below?
A naïve Bayes classifier estimates the class-conditional probability, P(X | Y), by assuming that the attributes are conditionally independent, given by the class label y.

$$P(X \mid Y = y) = \prod_{i=1}^{d} P(X_i \mid Y = y)$$

Where each attribute set X = {$X_1$, $X_2$ … $X_d$} consists of d attributes.

| | | | |
|---|---|---|---|
| A | True | B | False |
| C | | D | |

| | |
|---|---|
| Answer | **A** |
| Remarks | |

---

| Classification of Data | Question 5 |
|---|---|

An online computer store uses naïve Bayes classifier to estimate the probability of the registered store user buying a computer or not.

Let X is a set of attributes of the registered user.
X = {id, age, income, student, credit_rating}

Let Y is the class labels to assign
Y = buys_computer = {yes, no}

There exists a training dataset, D as below. (For brevity, the dataset D is omitted for this question).

| id | age | income | student | credi_rating | **buys_computer** |
|---|---|---|---|---|---|
| ... | ... | ... | ... | ... | ... |

A new user was registered and the tuple is as below.

| id | age | income | student | credi_rating | **buys_computer** |
|---|---|---|---|---|---|
| 99 | youth | medium | yes | fair | ? |

The above tuple implies the following attribute values.
X = {age = youth, income = medium, student = yes, credit_rating = fair}
Y = buys_computer = { yes, no}

The naïve Bayes classifier has to estimate the probability (predict) of the new user buying a computer or not. i.e., assigning a class label to the tuple.

The following **posterior probabilities,** P(Y | X) were computed on the new user.
P(buys_computer = yes | X) = 0.028
P(buys_computer = no | X) = 0.048

Hint: A naïve Bayes classifier is to estimate the posterior probabilities P(Y = y | X).

$$P(Y = y \mid X) = \frac{P(Y) \prod_{i=1}^{d} P(X_i \mid Y = y)}{P(X)}$$

| | | | |
|---|---|---|---|
| A | The naïve Bayes classifier predicts the new user buys_computer = no | B | The naïve Bayes classifier predicts the new user buys_computer = yes |
| C | The naïve Bayes classifier is inconclusive. | D | None of the above. |

| | |
|---|---|
| Answer | **A** |
| Remarks | The posterior probability of buys_computer = no is higher. |

---

| **Classification of Data** | **Question 6** |
|---|---|

An online computer store uses naïve Bayes classifier to estimate the probability of the registered store user buying a computer or not.

Let X is a set of attributes of the registered user.
X = {id, age, income, student, credit_rating}

Let Y is the class labels to assign
Y = buys_computer = {yes, no}

There exists a training dataset, D as below. (For brevity, the dataset D is omitted for this question).

| id | age | income | student | credi_rating | **buys_computer** |
|---|---|---|---|---|---|
| ... | ... | ... | ... | ... | ... |

A new user was registered and the tuple is as below.

| id | age | income | student | credi_rating | **buys_computer** |
|---|---|---|---|---|---|
| 99 | youth | medium | yes | fair | ? |

The above tuple implies the following attribute values.
X = {age = youth, income = medium, student = yes, credit_rating = fair}
Y = buys_computer = { yes, no}

The naïve Bayes classifier has to estimate the probability (predict) of the new user buying a computer or not. i.e., assigning a class label to the tuple.

The following **posterior probabilities,** P(Y | X) were computed on the new user.
P(buys_computer = yes | X) = 0
P(buys_computer = no  | X) = 0

Hint: A naïve Bayes classifier is to estimate the posterior probabilities P(Y = y | X).

$$P(Y = y \mid X) = \frac{P(Y) \prod_{i=1}^{d} P(X_i \mid Y = y)}{P(X)}$$

| | | | |
|---|---|---|---|
| A | The naïve Bayes classifier predicts the new user buys_computer = no | B | The naïve Bayes classifier predicts the new user buys_computer = yes |
| C | The naïve Bayes classifier is | D | None of the above. |

| | | | |
|---|---|---|---|
| | inconclusive. | | |

| | |
|---|---|
| Answer | **C** |
| Remarks | This is a case for M-estimate of conditional probability<br>Laplace correction or Laplace estimation |

| Classification of Data | Question 7 |
|---|---|

Below is the generic algorithm to generate a decision tree. It results a decision tree with nodes split on a certain criterion.

```
Node Generate_decision_tree (Dj, attribute_list) {
    Create a Node N
    if tuple in D are all of the same class, C, then
        return N as a leaf node labelled with the class C;
    if attribute_list is empty then
       // majority voting
       return N as a leaf node labelled with the majority class in D;
    apply Attribute_selection_method (D, attribute_list) to find the best
    splitting_criterion;
    label node N with splitting_criterion;
    if splitting_attribute is discrete-valued and
       multiway splits allowed then  // not restricted to binary trees
    // remove splitting_attribute
    attribute_list <- attribute_list -  splitting_attribute;
    for each outcome j of splitting_criterion
    // partition the tuples and grow subtrees for each partition
     let Dj be the set of data tuples in D satisfying outcome j; // a
    partition
       if Dj is empty then
          attach a leaf labeled with the majority class in D to node N;
       else attach the node returned by Generate_decision_tree (Dj,
    attribute_list) to node N;
   endfor
    return N;
}
```

What are popular **attribute selection methods** for the above decision tree generation algorithm?

| | | | |
|---|---|---|---|
| A | Information Gain | B | Gain Ratio |
| C | Gini Index | D | All of the above. |

| | |
|---|---|
| Answer | **D** |
| Remarks | Decision Tree algorithm, and attribute selection methods. |

| Classification of Data | Question 8 |
|---|---|

IF-THEN rules can be extracted directly from the training data using a sequential covering algorithm. What are the popular sequential covering algorithms?

| | | | |
|---|---|---|---|
| A | AQ | B | CN2 |
| C | RIPPER | D | All of the above |

| | |
|---|---|
| Answer | **D** |
| Remarks | |

---

| **Classification of Data** | **Question 9** |
|---|---|
| | |

| | | | |
|---|---|---|---|
| A | | B | |
| C | | D | |

| | |
|---|---|
| Answer | |
| Remarks | |

---

| **Classification of Data** | **Question 10** |
|---|---|
| | |

| | | | |
|---|---|---|---|
| A | | B | |
| C | | D | |

| | |
|---|---|
| Answer | |
| Remarks | |

---