# Final Exam

**Due** May 9 at 3:45pm **Points** 50 **Questions** 42
**Available** May 9 at 1:15pm - May 9 at 3:45pm about 3 hours **Time Limit** 75 Minutes

# Instructions

- The exam on **modules 7, 8, 9, 10, 11 and 12.**
- The exam will be available on **Monday May 09, 2022 from 1:15 PM to 3:45 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **LATEST** | Attempt 1 | 75 minutes | 29 out of 50 * |

* Some questions not yet graded

ⓘ Correct answers will be available on May 10 at 4pm.

Score for this quiz: **29** out of 50 *
Submitted May 9 at 2:30pm
This attempt took 75 minutes.

### Question 1 — 1 / 1 pts

What is true regarding backpropagation rule?

- there is no feedback of signal at any stage
- all of the mentioned
- it is also called generalized delta rule

error in output is propagated backwards only to determine weight updates

## Question 2

1 / 1 pts

The network that involves backward links from output to the input and hidden layers is called as ____.

- perceptrons
- multi layered perceptron
- self organizing maps
- recurrent neural network

## Question 3

1 / 1 pts

What is true regarding backpropagation rule?

- actual output is determined by computing the outputs of units for each hidden layer
- none of the mentioned

hidden layers output is not all important, they are only meant for supporting input and output layers

it is a feedback neural network

## Question 4

1 / 1 pts

What is perceptron?

a neural network that contains feedback

an auto-associative neural network

a single layer feed-forward neural network with pre-processing

a double layer auto-associative neural network

## Question 5

1 / 1 pts

Why is the XOR problem exceptionally interesting to neural network researchers?

because it is complex binary operation that cannot be solved using neural networks

because it is the simplest linearly inseparable problem that exists

because it can be solved by a single layer perceptron

because it can be expressed in a way that allows you to use a neural network

## Question 6

1 / 1 pts

Having multiple perceptrons can actually solve the XOR problem satisfactorily: this is because each perceptron can partition off a linear part of the space itself, and they can then combine their results.

false – perceptrons are mathematically incapable of solving linearly inseparable functions, no matter what you do

true – perceptrons can do this but are unable to learn to do it – they have to be explicitly hand-coded

false – just having a single perceptron is enough

true – this works always, and these multiple perceptrons learn to classify even complex problems.

## Question 7

1 / 1 pts

Which of the following is not a correct statement for SVM.

- (selected) All instances are required to define the maximum margin hyperplane.
- SVMs can be reused as algorithms for learning linear classifiers
- Instances closest to the maximum margin hyperplane are called support vectors
- SVMs are resilient to overfitting

## Question 8

1 / 1 pts

ROC in performance metrics stands for?

- Remote operating characteristic
- (selected) Receiver operating characteristic
- Reverse operating characteristic
- Revise operating characteristic

## Question 9

1 / 1 pts

What’s the objective of the support vector machine algorithm?

to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of samples.

to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of target variables.

to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of features.

None of the mentioned

## Question 10

1 / 1 pts

**What is/are true about kernel in SVM?**

1. Kernel function map low dimensional data to high dimensional space

2. It's a similarity function

None of these

1 and 2

1

2

## Question 11

1 / 1 pts

**Which of the following are real world applications of the SVM?**

- Clustering of News Articles
- Text and Hypertext Categorization
- Image Classification
- All of the mentioned

## Question 12

1 / 1 pts

The effectiveness of an SVM depends upon:

- Kernel Parameters
- All of the mentioned
- Selection of Kernel
- Soft Margin Parameter C

Incorrect

## Question 13

0 / 1 pts

A rule with a lower value of confidence and support could be

preferred because:

- None of the mentioned
- Such rules are more interesting
- It indicates novelty
- Such rules are bound to hold throughout the dataset

## Question 14

1 / 1 pts

**Which of the following is not an example of frequent pattern analysis?**

- Can we predict the winner of match?
- What are the subsequent purchases after buying a PC?
- What kinds of DNA are sensitive to this new drug?
- Can we automatically classify web documents?

## Question 15

1 / 1 pts

Which of the following describes a strategy of frequent Itemset generation?

- Use pruning techniques to reduce the number of candidates
- All of the mentioned
- Use efficient data structures to store the candidates or transactions
- Reduce size of the number of transactions as the size of itemset increases

## Question 16

1 / 1 pts

Which of the following refers to the sequence of pattern that occurs frequently?

- Frequent sub-structure
- Frequent sub-items
- Frequent sub-sequence
- All of the mentioned

Incorrect

## Question 17

0 / 1 pts

Which of the following best describes the Apriori principle?

- [x] If an itemset is frequent, then all of its subsets must also be frequent
- [ ] All of the mentioned
- [ ] Support of an itemset never exceeds the support of its subsets
- [ ] When the anti-monotone property of support holds on given itemset

## Question 18

1 / 1 pts

Which of the following best describes lift in knowledge discovery?

- [ ] An unsupervised learning approach
- [ ] A known class attribute
- [ ] A data mining technique
- [x] A measure of interestingness of a rule

## Question 19

1 / 1 pts

A _____ is a tree diagram for displaying clustering results. Vertical lines represent clusters that are joined together.

- [x] Dendrogram

- Scatter plot
- Tree plot
- Histogram

Incorrect

## Question 20

0 / 1 pts

Which of the following is true about cluster analysis?

- Clustering is referred to as an unsupervised learning method.
- Clustering is referred to as a supervised learning method
- Cluster analysis is the process of ungrouping objects into subsets that have meaning in the context of a particular problem
- It can’t uncover previously undetected relationships in a complex dataset.

## Question 21

1 / 1 pts

_____ Method derives clusters from the number of observations locally falling in a neighborhood of each observation.

- Density-based method

- Hierarchical method
- Partition method
- Grid method

## Question 22

1 / 1 pts

Which of the following statement is NOT true about clustering?

- It uses clusters for data analysis
- It groups the data
- It is an unsupervised learning technique
- It is a supervised learning technique

## Question 23

1 / 1 pts

HAC stands for ----.

- Hierarchical aggregative clustering
- Heightened agglomerative clustering
- Hierarchical agglomerative clustering
- Hierarchical absolute clustering

Unanswered

## Question 24

0 / 1 pts

Which of the following is not an application of cluster analysis?

- Observing earth quake epicenters should be clustered along continent faults.
- Decide about the subsequent purchases after buying a PC.
- Help marketers discover distinct groups in their customer bases.
- Identifying groups of motor insurance policy holders with a high average claim cost.

Incorrect

## Question 25

0 / 1 pts

Which statement is not true about cluster analysis?

- Cluster analysis is a technique for analyzing data when the criterion or dependent variable is categorical and the independent variables are interval in nature.
- Objects in each cluster tend to be similar to each other and dissimilar to objects in the other clusters.

Cluster analysis is also called classification analysis or numerical taxonomy.

Groups or clusters are suggested by the data, not defined a priori.

Incorrect

## Question 26

0 / 1 pts

Which one of the following statements about the K-means clustering is incorrect?

The goal of the k-means clustering is to partition (n) observation into (k) clusters

All of the mentioned

The nearest neighbor is the same as the K-means

K-means clustering can be defined as the method of quantization

## Question 27

1 / 1 pts

One of the drawbacks of using statistical methods in anomaly detection is:

all of the mentioned

Sometime it can be Can difficult to decide on number of clusters

Density may become less meaningful in high-dimensional space.

It may be hard to estimate the true distribution for high dimensional data.

## Question 28

1 / 1 pts

Which of the following will be Euclidean Distance between the two data point A(1, 3) and B(2, 3)?

8

4

2

1

## Question 29

1 / 1 pts

Data may contain erroneous or anomalous values, which are usually referred to as_____.

Inconsistencies

Noise

- Reduction
- Outliers

## Question 30

1 / 1 pts

An observation that is extreme, being distant from the rest of the data is termed a ------.

- Feature
- Class
- Predictor
- Outlier

## Question 31

1 / 1 pts

Which one of the following can be defined as the data object which does not comply with the general behavior (or the model of available data)?

- Prediction
- Evaluation Analysis
- Outlier Analysis

Classification

Incorrect

## Question 32

0 / 1 pts

One of the strengths of using statistical methods in anomaly detection is:

- It can use many dimensionality reduction approaches.
- All of the mentioned
- It is very effective to find the outliers.
- Theoretically it can be Theoretically it can be applicable to all kinds of dataapplicable to all kinds of data

Incorrect

## Question 33

0 / 1 pts

In web mining, _____ is used to know the order in which URLs tend to be accessed.

- associations
- sequential analysis
- classification
- clustering

## Question 34

1 / 1 pts

Which of the following is not part of Web Mining:

- Usage Mining
- Database Mining
- Content Mining
- Structure Mining

## Question 35

1 / 1 pts

Which of the following is a new trend in data mining?

- A Web mining
- Scalable data mining methods
- All of the mentioned
- Invisible data mining

## Question 36

1 / 1 pts

Which one of the following can be considered as the correct application of the data mining?

- (●) All of the mentioned
- ( ) Management and market analysis
- ( ) Corporate Analysis & Risk management
- ( ) Fraud detection

## Question 37

1 / 1 pts

Which of the following is one of the purposes of the visualization?

- ( ) It assists to search for trends and relationships among data.
- ( ) It helps find interesting regions for any further analysis.
- ( ) It can provide qualitative overview of large data sets
- (●) All of the mentioned

Incorrect

## Question 38

0 / 1 pts

Which of the following is not an example application of data mining in science and engineering?

Use data mining in monitoring systems, software bugs and network intrusion

All of the mentioned

Use sequential pattern mining to investigate changes in customer consumption or loyalty

Mining of spatiotemporal, biological, diverse semantics and relationships C)

## Question 39

Not yet graded / 3 pts

Discuss the basic difference between the **agglomerative** and **divisive** hierarchical clustering algorithms and mention which type of hierarchical clustering algorithm is more commonly used.

Your Answer:

**Agglomerative:** This is a "bottom-up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

**Divisive:** This is a "top-down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

**Agglomerative Hierarchical Clustering** is the most common type of hierarchical clustering used to group objects in clusters based on their similarity.

## Question 40

Not yet graded / 3 pts

We generally will be more interested in association rules with **high confidence.** However, often we will not be interested in association rules that have a confidence of **100%**. Why? Then specifically explain why association rules with 99% confidence may be interesting (i.e., **what might they indicate**)?

Your Answer:

While we generally prefer association rules with high confidence, a rule with 100% confidence most likely represents some already known fact or policy (e.g., a checking account → savings account may just indicate that all customers are required to have a checking account if they have a savings account).

Rules with 99% confidence are interesting not because of the 99% part but because of the 1% part. These are the exceptions to the rule. They may indicate, for example, that a policy is being violated. They might also indicate that there is a data entry error. Either way, it would be interesting to understand why the 1% do not follow the general pattern.

## Question 41

Not yet graded / 3 pts

Demonstrate how the **perceptron model** can be used to represent the **OR** functions between a pair of Boolean variables.

Your Answer:

From the diagram, the OR gate is 0 only if both inputs are 0.

**Row 1**

| A | B | Output |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

- From w1x1+w2x2+b, initializing w1, w2, as 1 and b as –1, we get;
- > x1(1)+x2(1)–1
  - Passing the first row of the OR logic table (x1=0, x2=0), we get;
- > 0+0–1 = –1
  - From the Perceptron rule, if Wx+b≤0, then y`=0. Therefore, this row is correct.
- **Row 2**
  - Passing (x1=0 and x2=1), we get;
  - > 0+1–1 = 0
    - From the Perceptron rule, if Wx+b <= 0, then y`=0. Therefore, this row is incorrect.
    - So we want values that will make inputs x1=0 and x2=1 give y` a value of 1. If we change w2 to 2, we have;

    > 0+2–1 = 1

    - From the Perceptron rule, this is correct for both the row 1 and 2.

    **Row 3**

    - Passing (x1=1 and x2=0), we get;

    > 1+0–1 = 0

    - From the Perceptron rule, if Wx+b <= 0, then y`=0. Therefore, this row is incorrect.
    - Since it is similar to that of row 2, we can just change w1 to 2, we have;

> 2+0–1 = 1

- From the Perceptron rule, this is correct for both the row 1, 2 and 3.

**Row 4**

- Passing (x1=1 and x2=1), we get;

> 2+2–1 = 3

- Again, from the perceptron rule, this is still valid. Quite Easy!

Therefore, we can conclude that the model to achieve an OR gate, using the Perceptron algorithm is;

> 2x1+2x2–1

## Question 42

**Not yet graded / 3 pts**

Consider a transaction dataset that contains five items, {A, B, C, D, E}. Suppose **the rules {A, B} → C has the same confidence as {A, B} → D**, which one of the following statements are true or not, and why:

1. a) The confidence of the {A, B} → {C, D} is the same as the confidence of {A, B} → {C}.
2. b) All transactions that contain {A, B, C} also contain {A, B, D}.

Your Answer:

Rule a is true because they have same confidence.

Also, Rule b is true because they have same confidence.

The confidence of an association rule is **a percentage value that shows how frequently the rule head occurs among all the groups containing the ruling body.**

Quiz Score: **29** out of 50

# Final Exam

**Due** May 10 at 3:45pm **Points** 50 **Questions** 42
**Available** May 10 at 1:15pm - May 10 at 3:45pm about 3 hours
**Time Limit** 75 Minutes

# Instructions

- The exam on **modules 7, 8, 9, 10, 11 and 12.**
- The exam will be available on **Monday May 10, 2022 from 1:15 PM to 3:45 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **LATEST** | Attempt 1 | 73 minutes | 36 out of 50 * |

* Some questions not yet graded

ⓘ Correct answers will be available on May 10 at 4pm.

Score for this quiz: **36** out of 50 *
Submitted May 10 at 3:05pm
This attempt took 73 minutes.

### Question 1 — 1 / 1 pts

Having multiple perceptrons can actually solve the XOR problem satisfactorily: this is because each perceptron can partition off a linear part of the space itself, and they can then combine their results.

- false – just having a single perceptron is enough

true – this works always, and these multiple perceptrons learn to classify even complex problems.

false – perceptrons are mathematically incapable of solving linearly inseparable functions, no matter what you do

true – perceptrons can do this but are unable to learn to do it – they have to be explicitly hand-coded

## Question 2

1 / 1 pts

Why is the XOR problem exceptionally interesting to neural network researchers?

because it is complex binary operation that cannot be solved using neural networks

because it can be solved by a single layer perceptron

because it is the simplest linearly inseparable problem that exists

because it can be expressed in a way that allows you to use a neural network

## Question 3

1 / 1 pts

The network that involves backward links from output to the input and hidden layers is called as ____.

- self organizing maps
- perceptrons
- recurrent neural network
- multi layered perceptron

## Question 4

1 / 1 pts

What are general limitations of back propagation rule?

- local minima problem
- scaling
- slow convergence
- all of the mentioned

## Question 5

1 / 1 pts

In feed- forward networks, the connections between layers are _____ from input to output.

- Unidirectional

- Directional
- Bidirectional
- Multidirectional

## Question 6

1 / 1 pts

What is the objective of backpropagation algorithm?

- to develop leaning algorithm for single layer feedforward neural network
- none of the mentioned
- to develop learning algorithm for multilayer feedforward neural network, so that network can be trained to capture the mapping implicitly
- to develop learning algorithm for multilayer feedforward neural network

## Question 7

1 / 1 pts

In ROC chart the proportion of false positive fp is on _____ and the proportion of true positive tp is on _____

- The horizontal axis, the x-axis
- The horizontal axis, the vertical axis

- The vertical axis, the y-axis
- The vertical axis, the horizontal axis

## Question 8

1 / 1 pts

ROC chart is a _____ plot.

- Three-dimensional
- One-dimensional
- Multi-dimensional
- Two-dimensional

## Question 9

1 / 1 pts

The effectiveness of an SVM depends upon:

- Selection of Kernel
- All of the mentioned
- Soft Margin Parameter C
- Kernel Parameters

## Question 10

1 / 1 pts

ROC in performance metrics stands for?

- Remote operating characteristic
- Receiver operating characteristic
- Revise operating characteristic
- Reverse operating characteristic

## Question 11

1 / 1 pts

In SVM, the Hyper plane, f(x)=sign(w*x+b), where ‘w’ is a?

- Vector
- Distance
- Constant
- None of the mentioned

## Question 12

1 / 1 pts

For SVM, which options are correct?

- None of the mentioned.

- Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane
- Support vectors are data points that are far away from the hyperplane and influence the position and orientation of the hyperplane
- Deleting the support vectors won't change the position of the hyperplane

## Question 13

1 / 1 pts

The analysis performed to uncover the interesting statistical correlation between associated -attributes value pairs are known as the _____.

- Mining of clusters
- Mining of correlation
- Mining of asociation
- All of the mentioned

## Question 14

1 / 1 pts

In association rules, what is meant by the term support?

- The number of instances correctly covered by the association rule.

- Combinations of attribute-value pairs that have a minimum coverage.
- An attribute-value pair.
- A frequent item set.

## Question 15

1 / 1 pts

Which of the following learning algorithm can be used to predict a combination of attributes?

- Naïve Bayesian.
- K-means
- Decision tree.
- Apriori.

## Question 16

1 / 1 pts

**Which of the following is not an example of frequent pattern analysis?**

- What are the subsequent purchases after buying a PC?
- Can we predict the winner of match?
- Can we automatically classify web documents?
- What kinds of DNA are sensitive to this new drug?

## Question 17

1 / 1 pts

Which of the following best describes the Apriori principle?

- (●) All of the mentioned
- ( ) When the anti-monotone property of support holds on given itemset
- ( ) If an itemset is frequent, then all of its subsets must also be frequent
- ( ) Support of an itemset never exceeds the support of its subsets

## Question 18

1 / 1 pts

Which of the following can affect the complexity of Apriori?

- ( ) Dimensionality of the given data set
- (●) All of the mentioned
- ( ) Maximum number of items in the transactions
- ( ) Number of transactions in the database

Incorrect

## Question 19

0 / 1 pts

Which of the following is true about cluster analysis?

Cluster analysis is the process of ungrouping objects into subsets that have meaning in the context of a particular problem

Clustering is referred to as a unsupervised learning method

Clustering is referred to as an supervised learning method.

It can't uncover previously undetected relationships in a complex dataset.

## Question 20

1 / 1 pts

Which one of the following can be considered as the final output of the hierarchal type of clustering?

None of the mentioned

A tree which displays how the close thing are to each other

Finalize estimation of cluster centroids

Assignment of each point to clusters

## Question 21

1 / 1 pts

In cluster analysis, which of the following is an advantage of choosing k>1?

Maximizes misclassification rate

- Minimizes classification rate
- Provides smoothing that reduces the risk of over fitting
- Doesn't maximize classification rate

## Question 22

1 / 1 pts

_____ Method derives clusters from the number of observations locally falling in a neighborhood of each observation.

- Grid method
- Hierarchical method
- Density-based method
- Partition method

## Question 23

1 / 1 pts

Partition algorithms usually stop when -----.

- All the observation are assigned to the cluster
- Reallocation occurs and subdivision appears unstable
- Subdivisions appear unstable

During the same iteration no reallocation occurs, subdivision appears stable with respect to the evaluation criterion chosen

## Question 24

1 / 1 pts

_____ is a clustering procedure characterized by the development of a tree-like structure.

- K-Means clustering
- Hierarchical clustering
- Non-hierarchical clustering
- K-Medoids clustering

## Question 25

1 / 1 pts

Which statement is not true about cluster analysis?

Cluster analysis is a technique for analyzing data when the criterion or dependent variable is categorical and the independent variables are interval in nature.

Objects in each cluster tend to be similar to each other and dissimilar to objects in the other clusters.

Groups or clusters are suggested by the data, not defined a priori.

Cluster analysis is also called classification analysis or numerical taxonomy.

## Question 26

1 / 1 pts

Which one of the following statements about the K-means clustering is incorrect?

- The nearest neighbor is the same as the K-means
- All of the mentioned
- K-means clustering can be defined as the method of quantization
- The goal of the k-means clustering is to partition (n) observation into (k) clusters

## Question 27

1 / 1 pts

One of the drawbacks of using density methods in anomaly detection is:

It may be hard to estimate the true density distribution for high dimensional data.

- Density may become less meaningful in high-dimensional space.
- All of the mentioned
- Sometime it can be sensitive to variations in density

## Question 28

1 / 1 pts

Data may contain erroneous or anomalous values, which are usually referred to as_____.

- Outliers
- Inconsistencies
- Noise
- Reduction

## Question 29

1 / 1 pts

Which of the following will be Euclidean Distance between the two data point A(1, 3) and B(2, 3)?

- 2

- 1
- 8
- 4

## Question 30

1 / 1 pts

For the purpose of anomaly detection, in the 1-Class SVM approaches we need to -----.

- All of the mentioned
- reduce data to lower dimensional data
- assume data comes from normal distribution
- use certain kernel function on the given data to construct such a model

## Question 31

1 / 1 pts

Which one of the following can be defined as the data object which does not comply with the general behavior (or the model of available data)?

- Classification
- Prediction
- Evaluation Analysis

- Outlier Analysis

## Question 32

1 / 1 pts

For the purpose of anomaly detection, in the reconstruction-based approaches we need to -----.

- assume data comes from normal distribution
- use certain kernel function on the given data to construct such a model
- All of the mentioned
- reduce data to lower dimensional data

Incorrect

## Question 33

0 / 1 pts

Which of the following is an example of the Mining Graphs and Network?

- Sequential Pattern Mining in Symbolic Sequences
- Clustering, Ranking and Classification of Heterogeneous Networks
- Mining Web Data
- All of the mentioned

## Question 34

1 / 1 pts

Which one of the following can be considered as the correct application of the data mining?

- (●) All of the mentioned
- ( ) Fraud detection
- ( ) Corporate Analysis & Risk management
- ( ) Management and market analysis

## Question 35

1 / 1 pts

Which of the following is method of preserving privacy in data mining process?

- (●) All of the mentioned
- ( ) Removing sensitive features or fields associated with the data
- ( ) Add noise to the data in order to mask some attribute values of records
- ( ) Personal information is encrypted and stored at different locations

## Question 36

1 / 1 pts

Data mining projects differ in many respects from both classical statistics and _____analyses.

- OLAM
- OLTP
- OLAP
- HOLAP

## Question 37

1 / 1 pts

Which of the following describes an example of the factor analysi?

- For special type of data, one attempts to determine several discriminant functions (factors) that discriminate among the groups defined by the response variable
- All of the mentioned
- For given experimental data, one analyzes the data for two or more populations described by a numeric response variable and one or more categorical variables (factors)

For certain data, researcher can indirectly measure other quantities that reflect the factor of interest

## Question 38

1 / 1 pts

In web mining, _____ is used to know the order in which URLs tend to be accessed.

- clustering
- classification
- sequential analysis
- associations

## Question 39

Not yet graded / 3 pts

Clustering has been popularly recognized as an important data mining task with broad applications. Give **one application example** for each of the following cases:

a) An application that takes clustering as a **major** data mining function.

b) An application that takes clustering as a **preprocessing tool for data preparation** for other data mining tasks.

Your Answer:

a)City-planning: Identifying groups of houses according to their house type, value, and geographical location..

b)Clustering can be used in process of Semi-Supervised Learning. There are some problems where in you might not have any idea about the data and relationships whatsoever with missing labels or no labels at all. We can use clustering first to find the natural segmentation of data and then apply some other model on it for better accuracy. For eg in the application suppose we have to find how a newly launched product has received response. Then in this case first we can apply clustering to segment users based on the region they belong to and then apply some model region wise for better accuracy.

## Question 40

**Not yet graded / 3 pts**

In your opinion, what are the major **5 trends** in data mining research today? Name one **major issue** in data mining, which in your view, may have a strong impact on society.

Your Answer:

Multimedia Data
Mining This is one of the latest methods which is catching up because of the growing ability to capture useful data accurately.

Ubiquitous Data Mining
This method involves the mining of data from mobile devices to get information about individuals.

Distributed Data Mining
This type of data mining is gaining popularity as it involves the mining of huge amount of information stored in different company locations or at different organizations.

Spatial and Geographic Data Mining
This is new trending type of data mining which includes extracting information from environmental, astronomical, and geographical data which also includes images taken from outer space.

Time Series and Sequence Data Mining
This practice is also helpful in analysing even random events which occur outside the normal series of events.

One major issue in Data Mining in my opinion is:
Mining different kinds of knowledge in databases − Different users may be interested in different kinds of knowledge. Therefore, it is necessary for data mining to cover a broad range of knowledge discovery task.

## Question 41

Not yet graded / 3 pts

Consider the following set of one-dimensional points: **{4, 6, 17, 19, 23, 27, 33, 37}**. Suppose we apply **k-means clustering** to obtain three clusters, A, B, and C. If the three initial centroids are located at **(15, 25, 31)**, respectively. Show: The three clusters and their new three centers after the **first round** of execution.

Your Answer:

cluster A= {4,6,17,19}
centroid = (4+6+17+19)/4=11.5
cluster B= {23,27}
centroid= (23+27)/2=25
cluster C= {33,37}
centroid= (33+37)/2=35
New centroid is (11.5,25, 35)

## Question 42

Not yet graded / 3 pts

Consider a transaction dataset that contains five items, {A, B, C, D, E}. Suppose **the rules {A, B} → C has the same**

**confidence as {A, B} → D**, which one of the following statements are true or not, and why:

1. a) The confidence of the {A, B} → {C, D} is the same as the confidence of {A, B} → {C}.
2. b) All transactions that contain {A, B, C} also contain {A, B, D}.

Your Answer:

Rule a is false because they donot have same confidence.

Also, Rule b is false because they donot have same confidence.

The confidence of an association rule is a percentage value that shows how frequently the rule head occurs among all the groups containing the ruling body.

Quiz Score: **36** out of 50

# Final Exam

**Due** May 10 at 3:45pm **Points** 50 **Questions** 42
**Available** May 10 at 1:15pm - May 10 at 3:45pm about 3 hours
**Time Limit** 75 Minutes

# Instructions

- The exam on **modules 7, 8, 9, 10, 11 and 12.**
- The exam will be available on **Monday May 10, 2022 from 1:15 PM to 3:45 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **LATEST** | Attempt 1 | 73 minutes | 35 out of 50 * |

* Some questions not yet graded

ⓘ Correct answers will be available on May 10 at 4pm.

Score for this quiz: **35** out of 50 *
Submitted May 10 at 3:05pm
This attempt took 73 minutes.

**Question 1** **1 / 1 pts**

What is perceptron?

- a neural network that contains feedback
- an auto-associative neural network

- a double layer auto-associative neural network
- [x] a single layer feed-forward neural network with pre-processing

## Question 2

**1 / 1 pts**

How can learning process be stopped in backpropagation rule?

- there is convergence involved
- no heuristic criteria exist
- none of the mentioned
- [x] on basis of average gradient value

## Question 3

**1 / 1 pts**

Layers between the input and output layers are known as:

- [x] Hidden layer
- Resultant layer
- Output layer
- Multilayer

## Question 4

1 / 1 pts

What is meant by generalized in statement “backpropagation is a generalized delta rule” ?

- (●) because delta rule can be extended to hidden layer units
- ( ) because delta is applied to only input and output layers, thus making it more simple and generalized
- ( ) it has no significance
- ( ) none of the mentioned

## Question 5

1 / 1 pts

What are general limitations of back propagation rule?

- ( ) scaling
- ( ) slow convergence
- ( ) local minima problem
- (●) all of the mentioned

## Question 6

1 / 1 pts

The network that involves backward links from output to the

input and hidden layers is called as ____.

- perceptrons
- self organizing maps
- multi layered perceptron
- recurrent neural network

## Question 7

**1 / 1 pts**

ROC chart is a _____ plot.

- Multi-dimensional
- One-dimensional
- Two-dimensional
- Three-dimensional

## Question 8

**1 / 1 pts**

For SVM, which options are correct?

- Deleting the support vectors won't change the position of the hyperplane

None of the mentioned.

Support vectors are data points that are far away from the hyperplane and influence the position and orientation of the hyperplane

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane

## Question 9

1 / 1 pts

**Which of the following are real world applications of the SVM?**

All of the mentioned

Text and Hypertext Categorization

Image Classification

Clustering of News Articles

## Question 10

1 / 1 pts

What's the objective of the support vector machine algorithm?

None of the mentioned

to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of features.

to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of target variables.

to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of samples.

## Question 11

**1 / 1 pts**

Which of the following is not a correct statement for SVM.

SVMs can be reused as algorithms for learning linear classifiers

All instances are required to define the maximum margin hyperplane.

SVMs are resilient to overfitting

Instances closest to the maximum margin hyperplane are called support vectors

## Question 12

**1 / 1 pts**

If x1, x2 are independent variables and y the dependent variable, which of the following represents a linear regression model?

- [x] y = a0 + a1x1 + a2x2
- [ ] y = a0 + a1/x1 + a2/x2
- [ ] y = a0 + a1x1 + a2x22
- [ ] y = a0 + a1x12 + a2x2

Incorrect

## Question 13

0 / 1 pts

A rule with a lower value of confidence and support could be

preferred because:

- [ ] None of the mentioned
- [x] Such rules are bound to hold throughout the dataset
- [ ] It indicates novelty
- [ ] Such rules are more interesting

## Question 14

1 / 1 pts

**Which of the following is not an example of frequent pattern analysis?**

- [x] Can we predict the winner of match?

- Can we automatically classify web documents?
- What kinds of DNA are sensitive to this new drug?
- What are the subsequent purchases after buying a PC?

## Question 15

**1 / 1 pts**

Which of the following learning algorithm can be used to predict a combination of attributes?

- Naïve Bayesian.
- Decision tree.
- Apriori.
- K-means

## Question 16

**1 / 1 pts**

Which of the following would be appropriate for a data mining

algorithm aimed at discovering which groups of products consumers

would tend to purchase together?

- Decision rules

- Classification rules
- Association rules
- Decision trees

## Question 17

**1 / 1 pts**

Which of the following best describes the Apriori principle?

- If an itemset is frequent, then all of its subsets must also be frequent
- All of the mentioned
- Support of an itemset never exceeds the support of its subsets
- When the anti-monotone property of support holds on given itemset

## Question 18

**1 / 1 pts**

Which of the following can affect the complexity of Apriori?

- All of the mentioned
- Maximum number of items in the transactions
- Dimensionality of the given data set
- Number of transactions in the database

## Question 19

1 / 1 pts

The subdivision of the hierarchical methods are:

- Partition and agglomerative
- Agglomerative and divisive
- Partition and divisive
- Distance based and density based

## Question 20

1 / 1 pts

Which one of the following correctly defines the term cluster?

- Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm
- Symbolic representation of facts or ideas from which information can potentially be extracted
- Group of similar objects that differ significantly from other objects
- All of the mentioned

Incorrect

## Question 21

0 / 1 pts

In cluster analysis, which of the following is an advantage of choosing k>1?

- Minimizes classification rate
- Maximizes misclassification rate
- Provides smoothing that reduces the risk of over fitting
- Doesn't maximize classification rate

## Question 22

1 / 1 pts

Which one of the following statements about the K-means clustering is incorrect?

- All of the mentioned
- The nearest neighbor is the same as the K-means
- The goal of the k-means clustering is to partition (n) observation into (k) clusters
- K-means clustering can be defined as the method of quantization

## Question 23

1 / 1 pts

Partition algorithms usually stop when -----.

- All the observation are assigned to the cluster
- (selected) During the same iteration no reallocation occurs, subdivision appears stable with respect to the evaluation criterion chosen
- Reallocation occurs and subdivision appears unstable
- Subdivisions appear unstable

## Question 24

**1 / 1 pts**

_____ Method derives clusters from the number of observations locally falling in a neighborhood of each observation.

- Grid method
- Partition method
- (selected) Density-based method
- Hierarchical method

## Question 25

**1 / 1 pts**

Which clustering algorithm starts with each cluster comprising exactly one observation and then progressively

combines the two nearest clusters until there is just one cluster left at the end?

- Complete Linkage clustering
- Agglomerative clustering
- Single Linkage clustering
- Divisive clustering

## Question 26

**1 / 1 pts**

Which statement is not true about cluster analysis?

- Groups or clusters are suggested by the data, not defined a priori.
- Objects in each cluster tend to be similar to each other and dissimilar to objects in the other clusters.
- Cluster analysis is also called classification analysis or numerical taxonomy.
- Cluster analysis is a technique for analyzing data when the criterion or dependent variable is categorical and the independent variables are interval in nature.

## Question 27

**1 / 1 pts**

An observation that is extreme, being distant from the rest of the data is termed a ------.

- Feature
- Outlier
- Class
- Predictor

## Question 28

**1 / 1 pts**

Which one of the following can be defined as the data object which does not comply with the general behavior (or the model of available data)?

- Evaluation Analysis
- Classification
- Prediction
- Outlier Analysis

## Question 29

**1 / 1 pts**

For the purpose of anomaly detection, in the statistical-based approaches we need to -----.

All of the mentioned

assume data comes rom normal distribution

reduce data to lower dimensional data

use certain kernel function on the given data to construct such a model

## Question 30

**1 / 1 pts**

Which of the following will be Euclidean Distance between the two data point A(1, 3) and B(2, 3)?

2

8

1

4

## Question 31

**1 / 1 pts**

One of the drawbacks of using statistical methods in anomaly detection is:

Sometime it can be Can difficult to decide on number of clusters

Density may become less meaningful in high-dimensional space.

It may be hard to estimate the true distribution for high dimensional data.

all of the mentioned

## Question 32

**1 / 1 pts**

One of the strengths of using statistical methods in anomaly detection is:

It is very effective to find the outliers.

All of the mentioned

Theoretically it can be Theoretically it can be applicable to all kinds of dataapplicable to all kinds of data

It can use many dimensionality reduction approaches.

## Question 33

**1 / 1 pts**

Which of the following is not trend of data mining?

Using data mining tasks of customers for targeted marketing

All of the mentioned

Mining multimedia, text and web data

- Distributed data mining an real-time data stream mining

## Question 34

1 / 1 pts

Which of the following is not part of Web Mining:

- Usage Mining
- Structure Mining
- Content Mining
- Database Mining

Incorrect

## Question 35

0 / 1 pts

Which of the following is an example of the Mining Graphs and Network?

- Sequential Pattern Mining in Symbolic Sequences
- Clustering, Ranking and Classification of Heterogeneous Networks
- All of the mentioned
- Mining Web Data

## Question 36

1 / 1 pts

In web mining, _____ is used to know the order in which URLs tend to be accessed.

- clustering
- associations
- classification
- sequential analysis

## Question 37

1 / 1 pts

Which of the following is a new trend in data mining?

- Invisible data mining
- All of the mentioned
- Scalable data mining methods
- A Web mining

## Question 38

1 / 1 pts

Which of the following is not an example application of data mining in science and engineering?

All of the mentioned

Mining of spatiotemporal, biological, diverse semantics and relationships C)

Use sequential pattern mining to investigate changes in customer consumption or loyalty

Use data mining in monitoring systems, software bugs and network intrusion

## Question 39

**Not yet graded / 3 pts**

Clustering has been popularly recognized as an important data mining task with broad applications. Give **one application example** for each of the following cases:

a) An application that takes clustering as a **major** data mining function.

b) An application that takes clustering as a **preprocessing tool for data preparation** for other data mining tasks.

Your Answer:

Clustering is recognized as an important data mining task with broad applications.

a.)
ANS: An example that takes clustering as a major data mining function could be a system that identifies groups of houses in a city according to house type, value, and geographical location. More specifically, a clustering algorithm loke CLARANS can be used to discover that, say, the most expensive housing units in Vancouver can be grouped into just

a few clusters.

(b)
ANS: An example application that takes clustering as a pre-processing tool for other data mining is spatial data mining. Spatial data mining is the discovery of interesting relationships and characteristics that may exist implicitly in spatial databases. We can apply cluster analysis only to spatial attributes, where natural notions of similarity exist. Various clustering algorithms, such as PAM, CLARA, or CLARANS may be used. The clusters obtained from the clustering algorithm may trigger the generation of nonspatial components in the sext step if such a generation from interesting groups of objects.

## Question 40

**Not yet graded / 3 pts**

Discuss the basic difference between the **agglomerative** and **divisive** hierarchical clustering algorithms and mention which type of hierarchical clustering algorithm is more commonly used.

Your Answer:

ANS: When compared to agglomerative clustering, divisive clustering is more complicated because we require a flat clustering algorithm as a subroutine to separate each cluster until each data has its own singleton cluster.

Divisive clustering is more efficient if we do not generate a complete hierarchy all the way down to individual data leaves. Time complexity of a naïve agglomerative clustering is $O(n^3)$ because we exhaustively scan N x N matrix dist_mat for lowest distance in each of N-1 iterations. Using priority queue data structure, we can reduce this complexity to $O(n^2 \log n)$. By using some more optimizations, it can be brought down to $O(n^2)$. Whereas for divisive clustering given a fixed number of top levels, using an efficient flat algorithm like k-means, divisible algorithms are linear in the number of patterns and clusters.

The dividing algorithm is also more precise. Without first examining the global distribution of data, agglomerative clustering makes judgments based on local patterns or neighbour points. These early decisions are irreversible. When generating top-level dividing decisions, divisive clustering takes into account the global distribution of data.

## Question 41

**Not yet graded / 3 pts**

A database has 4 transactions, shown below.

| **TID** | **Items bought** |
|---|---|
| T100 | {K, A, D, B} |
| T200 | {D, A, C, E, B} |
| T300 | {C, A, B, E} |
| T400 | {B, A, D} |

Assuming a minimum level of support **min_sup = 60%** and a minimum level of confidence **min_conf = 80%**.

Given the **frequent itemsets** are: **{{A}, {B}, {D}, {A, B}, {B, D}, {A, B, D}}**.

List all of the **strong association rules** (with support *s* and confidence *c*) **matching the following metarule**, where *X* is a variable representing customers, and $item_i$ denotes variables representing items (e.g., *"A", "B"*, etc.):

$\forall X \in transaction, buys(X,item1) \wedge buys(X,item2) \Rightarrow buys(X,item3)[s,c]$

Your Answer:

ANS:We have given the minimum level of support min_sup = 60% and minimum level of confidence min_conf = 80%. So, using this strong association rules are given below:

buys(X, A) ^ buys(X, B) → buys(X, D) (75%, 75%). Hence, it is not Strongly connected (Not Strong)

buys(X, A) ^ buys(X, D) → buys(X, B) (75%, 100%). Hence, it Strongly connected. (Strong)

buys(X, B) ^ buys(X, D) → buys(X, A) (75%, 100%). Hence, it is Strongly connected. (Strong)

## Question 42

**Not yet graded / 3 pts**

Consider the following set of one-dimensional points: **{4, 6, 17, 19, 23, 27, 33, 37}**. Suppose we apply **k-means clustering** to obtain three clusters, A, B, and C. If the three initial centroids are located at **(15, 25, 31)**, respectively. Show: The three clusters and their new three centers after the **first round** of execution.

Your Answer:

ANS: We have given the one-dimensional points: {4,6,17,19,23,27,33,37} and the initial centroids are A :15. B: 25, C: 31. So, by using the initial centroids we calculate the cluster for A, B and C.

For Cluster A= {4,6,17,19}
Centroids = (4+6+17+19)/4 = 11.5

For Cluster B= {23,27}
Centroid= (23+27)/2 = 25

For Cluster C= {33,37}

centroid= (33+37)/2 = 35

Hence, the new centroids for the cluster A, B and C are (11.5, 25, 35) respectively.

Quiz Score: **35** out of 50

# Final Exam

**Due** 10 May at 15:45 **Points** 50 **Questions** 42
**Available** 10 May at 13:15 - 10 May at 15:45 about 3 hours **Time limit** 75 Minutes

# Instructions

- The exam on **modules 7, 8, 9, 10, 11 and 12.**
- The exam will be available on **Monday May 10, 2022 from 1:15 PM to 3:45 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

## Attempt history

| | Attempt | Time | Score |
|---|---|---|---|
| **LATEST** | Attempt 1 | 74 minutes | 35 out of 50 * |

* Some questions not yet graded

ⓘ Correct answers will be available on 10 May at 16:00.

Score for this quiz: **35** out of 50 *
Submitted 10 May at 15:05
This attempt took 74 minutes.

**Question 1** **1 / 1 pts**

Why are linearly separable problems of interest of neural network researchers?

- because they are the only mathematical functions you can draw

because they are the only class of problem that network can solve successfully

because they are the only mathematical functions that are continue

because they are the only class of problem that perceptron can solve successfully

## Question 2

1 / 1 pts

In feed- forward networks, the connections between layers are _____ from input to output.

Directional

Bidirectional

Multidirectional

Unidirectional

## Question 3

1 / 1 pts

What are general limitations of back propagation rule?

local minima problem

- all of the mentioned
- slow convergence
- scaling

## Question 4

1 / 1 pts

What is perceptron?

- a neural network that contains feedback
- a double layer auto-associative neural network
- a single layer feed-forward neural network with pre-processing
- an auto-associative neural network

## Question 5

1 / 1 pts

Why is the XOR problem exceptionally interesting to neural network researchers?

- because it can be solved by a single layer perceptron
- because it is complex binary operation that cannot be solved using neural networks
- because it is the simplest linearly inseparable problem that exists

because it can be expressed in a way that allows you to use a neural network

## Question 6

1 / 1 pts

How can learning process be stopped in backpropagation rule?

- no heuristic criteria exist
- none of the mentioned
- on basis of average gradient value
- there is convergence involved

## Question 7

1 / 1 pts

Which of the following is not a correct statement for SVM.

- All instances are required to define the maximum margin hyperplane.
- Instances closest to the maximum margin hyperplane are called support vectors
- SVMs are resilient to overfitting
- SVMs can be reused as algorithms for learning linear classifiers

## Question 8

1 / 1 pts

ROC chart is a _____ plot.

- Multi-dimensional
- One-dimensional
- Three-dimensional
- Two-dimensional

## Question 9

1 / 1 pts

If x1, x2 are independent variables and y the dependent variable, which of the following represents a linear regression model?

- y = a0 + a1x1 + a2x22
- y = a0 + a1x1 + a2x2
- y = a0 + a1x12 + a2x2
- y = a0 + a1/x1 + a2/x2

## Question 10

1 / 1 pts

In SVM, the Hyper plane, f(x)=sign(w*x+b), where ‘w’ is a?

- Distance
- Constant
- None of the mentioned
- (selected) Vector

## Question 11

1 / 1 pts

ROC in performance metrics stands for?

- Revise operating characteristic
- (selected) Receiver operating characteristic
- Remote operating characteristic
- Reverse operating characteristic

## Question 12

1 / 1 pts

In ROC chart the proportion of false positive fp is on _____ and the proportion of true positive tp is on _____

- The vertical axis, the y-axis

- (●) The horizontal axis, the vertical axis
- ( ) The horizontal axis, the x-axis
- ( ) The vertical axis, the horizontal axis

## Question 13

1 / 1 pts

In association rules, what is meant by the term support?

- ( ) An attribute-value pair.
- ( ) Combinations of attribute-value pairs that have a minimum coverage.
- ( ) A frequent item set.
- (●) The number of instances correctly covered by the association rule.

Incorrect

## Question 14

0 / 1 pts

A rule with a lower value of confidence and support could be

preferred because:

- ( ) Such rules are bound to hold throughout the dataset
- (●) Such rules are more interesting
- ( ) It indicates novelty

- None of the mentioned

## Question 15

1 / 1 pts

Which of the following best describes the Apriori principle?

- All of the mentioned
- If an itemset is frequent, then all of its subsets must also be frequent
- When the anti-monotone property of support holds on given itemset
- Support of an itemset never exceeds the support of its subsets

## Question 16

1 / 1 pts

Which of the following best describes lift in knowledge discovery?

- A known class attribute
- A data mining technique
- An unsupervised learning approach
- A measure of interestingness of a rule

## Question 17

1 / 1 pts

if none of its immediate supersets has support count as X, then X will be:

- Maximal itemset
- Maximal frequent itemset
- Closed itemset
- Closed frequent itemset

## Question 18

1 / 1 pts

Which of the following learning algorithm can be used to predict a combination of attributes?

- Decision tree.
- Apriori.
- K-means
- Naïve Bayesian.

## Question 19

1 / 1 pts

Which one of the following can be considered as the final output of the hierarchal type of clustering?

- None of the mentioned

- Assignment of each point to clusters
- Finalize estimation of cluster centroids
- A tree which displays how the close thing are to each other

## Question 20

1 / 1 pts

A _____ is a tree diagram for displaying clustering results. Vertical lines represent clusters that are joined together.

- Histogram
- Scatter plot
- Tree plot
- Dendrogram

## Question 21

1 / 1 pts

Which clustering method develops a subdivision of the given dataset into a predetermined number K of non empty subset?

- Partition method
- Density-based method
- Hierarchical method

Grid method

## Question 22

1 / 1 pts

Partition algorithms usually stop when -----.

All the observation are assigned to the cluster

During the same iteration no reallocation occurs, subdivision appears stable with respect to the evaluation criterion chosen

Reallocation occurs and subdivision appears unstable

Subdivisions appear unstable

## Question 23

1 / 1 pts

Which one of the following correctly defines the term cluster?

Symbolic representation of facts or ideas from which information can potentially be extracted

Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm

Group of similar objects that differ significantly from other objects

- All of the mentioned

## Question 24

1 / 1 pts

From the following which method is not the clustering method?

- Density based
- Partition
- Hierarchical
- Divide-and-conquer based

## Question 25

1 / 1 pts

Which of the following statement is NOT true about clustering?

- It is a supervised learning technique
- It groups the data
- It is an unsupervised learning technique
- It uses clusters for data analysis

## Question 26

1 / 1 pts

Which of the following is true about cluster analysis?

- Clustering is referred to as an supervised learning method.
- Cluster analysis is the process of ungrouping objects into subsets that have meaning in the context of a particular problem
- It can't uncover previously undetected relationships in a complex dataset.
- Clustering is referred to as a unsupervised learning method

## Question 27

1 / 1 pts

One of the drawbacks of using statistical methods in anomaly detection is:

- It may be hard to estimate the true distribution for high dimensional data.
- Sometime it can be Can difficult to decide on number of clusters
- all of the mentioned
- Density may become less meaningful in high-dimensional space.

## Question 28

1 / 1 pts

Which one of the following can be defined as the data object which does not comply with the general behavior (or the model of available data)?

- Evaluation Analysis
- Classification
- Outlier Analysis
- Prediction

Incorrect

## Question 29

0 / 1 pts

For the purpose of anomaly detection, in the statistical-based approaches we need to -----.

- reduce data to lower dimensional data
- assume data comes rom normal distribution
- All of the mentioned
- use certain kernel function on the given data to construct such a model

## Question 30

1 / 1 pts

An observation that is extreme, being distant from the rest of the data is termed a ------.

- Outlier
- Predictor
- Feature
- Class

## Question 31

1 / 1 pts

Data may contain erroneous or anomalous values, which are usually referred to as_____.

- Inconsistencies
- Reduction
- Noise
- Outliers

Incorrect

## Question 32

0 / 1 pts

For the purpose of anomaly detection, in the 1-Class SVM approaches we need to -----.

- reduce data to lower dimensional data
- assume data comes from normal distribution

use certain kernel function on the given data to construct such a model

All of the mentioned

## Question 33

1 / 1 pts

Which of the following describes an example of the factor analysi?

For given experimental data, one analyzes the data for two or more populations described by a numeric response variable and one or more categorical variables (factors)

All of the mentioned

For special type of data, one attempts to determine several discriminant functions (factors) that discriminate among the groups defined by the response variable

For certain data, researcher can indirectly measure other quantities that reflect the factor of interest

## Question 34

1 / 1 pts

Which of the following is method of preserving privacy in data mining process?

- [x] All of the mentioned
- [ ] Add noise to the data in order to mask some attribute values of records
- [ ] Removing sensitive features or fields associated with the data
- [ ] Personal information is encrypted and stored at different locations

## Question 35

1 / 1 pts

Which of the following is not type of data mining in recommender systems?

- [ ] Model-based method uses a collection of ratings to learn a model
- [ ] All of the mentioned
- [x] Mining of spatiotemporal, biological, diverse semantics and relationships
- [ ] Extract from known to unknown ratings to predict user-item combinations

## Question 36

1 / 1 pts

Which of the following is a new trend in data mining?

- [ ] A Web mining
- [x] All of the mentioned

- Invisible data mining
- Scalable data mining methods

## Question 37

1 / 1 pts

Which of the following is not part of Web Mining:

- Database Mining
- Content Mining
- Usage Mining
- Structure Mining

## Question 38

1 / 1 pts

Which of the following is one of the purposes of the visualization?

- It helps find interesting regions for any further analysis.
- It assists to search for trends and relationships among data.
- All of the mentioned
- It can provide qualitative overview of large data sets

## Question 39

Not yet graded / 3 pts

In your opinion, what are the major **5 trends** in data mining research today? Name one **major issue** in data mining, which in your view, may have a strong impact on society.

Your answer:

Time Series and Sequence Data Mining: This is one of the important trending research which can be contributing to predicting the stock prices i.e. continuous data.

Spatial and Geographic Data Mining: This is a novel sort of data mining that involves collecting information from environmental, astronomical, and geographical data, as well as visuals from space.

Distributed Data Mining: This sort of data mining is gaining popularity since it includes mining a large volume of data held across several firm sites or organizations.

Ubiquitous Data Mining: This strategy entails obtaining information about individuals by mining data from mobile devices.

Multimedia Data Mining: This is one of the most recent ways that is taking on due to the increasing capacity to reliably gather meaningful data.

One major issue in Data Mining in my opinion is:

Mining different kinds of knowledge in databases − In today's era, there are lot of working domains and the presence of data mining techniques for each domain is not possible but sometimes generalized also fails, as each domain requires some particular set of mining techniques.

## Question 40

Not yet graded / 3 pts

Discuss the basic difference between the **agglomerative** and **divisive** hierarchical clustering algorithms and mention which type of hierarchical clustering algorithm is more commonly used.

Your answer:

Divisive clustering is more difficult than agglomerative clustering because it requires a flat clustering method as a subroutine to separate each cluster until each data has its own singleton cluster.

If we don't create a full hierarchy all the way down to individual data leaves, divisive clustering is more efficient. Because researchers thoroughly scan the N x N matrix dist mat for the lowest distance in each of the N-1 iterations, the time complexity of a naive agglomerative clustering is O(n3). We can decrease this complexity to O(n2logn) by using a priority queue data structure. It can be reduced to O(n2) with some further improvements  Divisible techniques are linear in the number of patterns and clusters, whereas divisive clustering given a fixed number of top levels uses an efficient flat approach like k-means.

The algorithm for dividing is also more exact. Agglomerative clustering makes decisions based on local patterns of neighbor points rather than the global distribution of data. These early judgments cannot be changed. Divisive clustering considers the global distribution of data while making top-level division choices.

## Question 41

**Not yet graded / 3 pts**

For each of the Boolean functions given below, state whether the problem is **linearly separable** or not, and why?

1. a) A AND B AND C
2. b) NOT A AND B
3. c) (A OR B) AND (A OR C)

Your answer:

1. The statement (A AND B AND C) contains the AND operation. So it is linearly separable.
2. The statement NOT A AND B contains the NOT and AND operation, so it is linearly separable.
3. The statement (A OR B) AND (A OR C) contains the OR and AND operation so it is linearly separable.

Because for all the inputs of A, B, and C the X-Y plane are separated by one single straight line in AND, OR, NOT operation

## Question 42

**Not yet graded / 3 pts**

Consider a transaction dataset that contains five items, {A, B, C, D, E}. Suppose **the rules {A, B} → C has the same confidence as {A, B} → D**, which one of the following statements are true or not, and why:

1. a) The confidence of the {A, B} → {C, D} is the same as the confidence of {A, B} → {C}.
2. b) All transactions that contain {A, B, C} also contain {A, B, D}.

Your answer:

The support is the probability that a randomly selected transaction from the database will contain all items in the antecedent and the consequent, whereas the confidence is the conditional probability that a randomly selected transaction will include all the items in the consequent, given that the transaction includes all the items in the antecedent.

1. False: because the antecedent in the 1st item set is same as the 2nd item set. And the consequents are different. From the confidence calculations from P(item set)/P(item), we can determine that the confidence of {A,B} -> {C,D} is not same as the confidence of {A, B} -> {C}.

2. False: because from the rules of {A,B} -> C has the same confidence as {A.B} -> D, we can confirm from the confidence from P(item set)/P(item), that all the transactions that contain {A,B,C} also contains {A,B,C} because the confidence would be same for both the item sets.

Quiz score: **35** out of 50

# Final Exam

**Due** May 9 at 3:45pm **Points** 50 **Questions** 42
**Available** May 9 at 1:15pm - May 9 at 3:45pm about 3 hours **Time Limit** 75 Minutes

# Instructions

- The exam on **modules 7, 8, 9, 10, 11 and 12.**
- The exam will be available on **Monday May 09, 2022 from 1:15 PM to 3:45 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **LATEST** | Attempt 1 | 75 minutes | 34 out of 50 * |

* Some questions not yet graded

ⓘ Correct answers will be available on May 10 at 4pm.

Score for this quiz: **34** out of 50 *
Submitted May 9 at 2:30pm
This attempt took 75 minutes.

### Question 1 — 1 / 1 pts

What is true regarding backpropagation rule?

- there is no feedback of signal at any stage
- all of the mentioned
- it is also called generalized delta rule

error in output is propagated backwards only to determine weight updates

## Question 2

1 / 1 pts

Why are linearly separable problems of interest of neural network researchers?

- because they are the only mathematical functions you can draw
- because they are the only mathematical functions that are continue
- **because they are the only class of problem that perceptron can solve successfully**
- because they are the only class of problem that network can solve successfully

## Question 3

1 / 1 pts

What is perceptron?

- a neural network that contains feedback
- a double layer auto-associative neural network

- a single layer feed-forward neural network with pre-processing
- an auto-associative neural network

## Question 4

1 / 1 pts

What are general limitations of back propagation rule?

- local minima problem
- scaling
- all of the mentioned
- slow convergence

## Question 5

1 / 1 pts

How can learning process be stopped in backpropagation rule?

- no heuristic criteria exist
- none of the mentioned
- on basis of average gradient value
- there is convergence involved

## Question 6

1 / 1 pts

Why is the XOR problem exceptionally interesting to neural network researchers?

- (●) because it is the simplest linearly inseparable problem that exists
- ( ) because it can be solved by a single layer perceptron
- ( ) because it can be expressed in a way that allows you to use a neural network
- ( ) because it is complex binary operation that cannot be solved using neural networks

## Question 7

1 / 1 pts

**Which of the following are real world applications of the SVM?**

- ( ) Clustering of News Articles
- ( ) Text and Hypertext Categorization
- ( ) Image Classification
- (●) All of the mentioned

## Question 8

1 / 1 pts

For SVM, which options are correct?

- Support vectors are data points that are far away from the hyperplane and influence the position and orientation of the hyperplane
- Deleting the support vectors won't change the position of the hyperplane
- None of the mentioned.
- **Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane**

## Question 9

1 / 1 pts

**The SVM's are less effective when -----.**

- The data is linearly separable
- All of the mentioned
- The data is clean and ready to use
- **The data is noisy and contains overlapping points**

## Question 10

1 / 1 pts

ROC in performance metrics stands for?

- Remote operating characteristic
- Reverse operating characteristic
- Revise operating characteristic
- Receiver operating characteristic

## Question 11

1 / 1 pts

ROC chart is a _____ plot.

- Multi-dimensional
- Two-dimensional
- Three-dimensional
- One-dimensional

## Question 12

1 / 1 pts

If x1, x2 are independent variables and y the dependent variable, which of the following represents a linear regression model?

- y = a0 + a1x1 + a2x22
- y = a0 + a1/x1 + a2/x2
- y = a0 + a1x12 + a2x2
- y = a0 + a1x1 + a2x2

Incorrect

## Question 13

0 / 1 pts

In association rules, what is meant by the term support?

- The number of instances correctly covered by the association rule.
- An attribute-value pair.
- A frequent item set.
- Combinations of attribute-value pairs that have a minimum coverage.

## Question 14

1 / 1 pts

Which of the following refers to the sequence of pattern that occurs frequently?

- Frequent sub-items
- Frequent sub-sequence
- All of the mentioned
- Frequent sub-structure

Incorrect

## Question 15

0 / 1 pts

Which of the following learning algorithm can be used to predict a combination of attributes?

- Decision tree.
- Apriori.
- K-means
- Naïve Bayesian.

## Question 16

1 / 1 pts

Which of the following describes a strategy of frequent Itemset generation?

- Use efficient data structures to store the candidates or transactions

- All of the mentioned
- Use pruning techniques to reduce the number of candidates
- Reduce size of the number of transactions as the size of itemset increases

## Question 17

1 / 1 pts

**Which of the following is not an example of frequent pattern analysis?**

- Can we automatically classify web documents?
- What are the subsequent purchases after buying a PC?
- Can we predict the winner of match?
- What kinds of DNA are sensitive to this new drug?

Incorrect

## Question 18

0 / 1 pts

if none of its immediate supersets has support count as X, then X will be:

- Closed frequent itemset
- Closed itemset

Maximal itemset

Maximal frequent itemset

## Question 19

1 / 1 pts

Which clustering method develops a subdivision of the given dataset into a predetermined number K of non empty subset?

Density-based method

Hierarchical method

Grid method

Partition method

## Question 20

1 / 1 pts

Which statement is not true about cluster analysis?

Cluster analysis is a technique for analyzing data when the criterion or dependent variable is categorical and the independent variables are interval in nature.

Objects in each cluster tend to be similar to each other and dissimilar to objects in the other clusters.

Groups or clusters are suggested by the data, not defined a priori.

Cluster analysis is also called classification analysis or numerical taxonomy.

## Question 21

1 / 1 pts

The aim of clustering models is to subdivide the records of a dataset into_____.

Heterogeneous groups

Homogeneous groups

Problem-Solving groups

Learning groups

## Question 22

1 / 1 pts

Which one of the following statements about the K-means clustering is incorrect?

The nearest neighbor is the same as the K-means

K-means clustering can be defined as the method of quantization

The goal of the k-means clustering is to partition (n) observation into (k) clusters

All of the mentioned

## Question 23

1 / 1 pts

Which one of the following can be considered as the final output of the hierarchal type of clustering?

None of the mentioned

Assignment of each point to clusters

Finalize estimation of cluster centroids

A tree which displays how the close thing are to each other

## Question 24

1 / 1 pts

A _____ is a tree diagram for displaying clustering results. Vertical lines represent clusters that are joined together.

Histogram

Dendrogram

Tree plot

Scatter plot

## Question 25

1 / 1 pts

_____ Method derives clusters from the number of observations locally falling in a neighborhood of each observation.

Hierarchical method

Density-based method

Partition method

Grid method

## Question 26

1 / 1 pts

Which of the following is not an application of cluster analysis?

Identifying groups of motor insurance policy holders with a high average claim cost.

Decide about the subsequent purchases after buying a PC.

Help marketers discover distinct groups in their customer bases.

Observing earth quake epicenters should be clustered along continent faults.

## Question 27

1 / 1 pts

One reason of anomaly detection is:

- Errors from collecting data
- Data coming from different classes
- Normal variations can be seen on data
- All of the mentioned

## Question 28

1 / 1 pts

For the purpose of anomaly detection, in the 1-Class SVM approaches we need to -----.

- All of the mentioned
- assume data comes from normal distribution
- use certain kernel function on the given data to construct such a model
- reduce data to lower dimensional data

## Question 29

1 / 1 pts

For the purpose of anomaly detection, in the statistical-based approaches we need to -----.

- All of the mentioned
- use certain kernel function on the given data to construct such a model
- assume data comes rom normal distribution
- reduce data to lower dimensional data

## Question 30

1 / 1 pts

One of the drawbacks of using density methods in anomaly detection is:

- It may be hard to estimate the true density distribution for high dimensional data.
- Density may become less meaningful in high-dimensional space.
- Sometime it can be sensitive to variations in density
- All of the mentioned

## Question 31

1 / 1 pts

Which one of the following can be defined as the data object which does not comply with the general behavior (or the model of available data)?

- Classification
- Prediction
- Outlier Analysis
- Evaluation Analysis

## Question 32

1 / 1 pts

An observation that is extreme, being distant from the rest of the data is termed a ------.

- Outlier
- Predictor
- Class
- Feature

## Question 33

1 / 1 pts

In web mining, _____ is used to know the order in which URLs tend to be accessed.

- associations
- classification
- sequential analysis
- clustering

## Question 34

1 / 1 pts

Which of the following is one of the purposes of the visualization?

- It assists to search for trends and relationships among data.
- All of the mentioned
- It helps find interesting regions for any further analysis.
- It can provide qualitative overview of large data sets

## Question 35

1 / 1 pts

Which of the following is not type of data mining in recommender systems?

- Mining of spatiotemporal, biological, diverse semantics and relationships
- All of the mentioned
- Extract from known to unknown ratings to predict user-item combinations
- Model-based method uses a collection of ratings to learn a model

Incorrect

## Question 36

0 / 1 pts

Which of the following is method of preserving privacy in data mining process?

- Add noise to the data in order to mask some attribute values of records
- All of the mentioned Mostly
- Removing sensitive features or fields associated with the data
- Personal information is encrypted and stored at different locations

## Question 37

1 / 1 pts

Which of the following is not trend of data mining?

- All of the mentioned

- Distributed data mining an real-time data stream mining
- Mining multimedia, text and web data
- Using data mining tasks of customers for targeted marketing

## Question 38

1 / 1 pts

Which of the following describes an example of the factor analysi?

- For given experimental data, one analyzes the data for two or more populations described by a numeric response variable and one or more categorical variables (factors)
- All of the mentioned
- For special type of data, one attempts to determine several discriminant functions (factors) that discriminate among the groups defined by the response variable
- For certain data, researcher can indirectly measure other quantities that reflect the factor of interest

## Question 39

Not yet graded / 3 pts

For each of the following questions, provide an example of **an association rule from the market basket domain** that satisfies the following conditions. Also, describe whether such rules are subjectively interesting or not.

a) A rule that has reasonably high support but low confidence.

b) A rule that has low support and high confidence.

Your Answer:

a)A rule that has reasonably high support but low confidence.

X = Milk Y = Cola {Milk} {Cola} Support = 2/5 = 40% Confidence = 2/4 = 50%

Such rule can be interesting besides the fact that both products are liquid that can be consumed, although one is a healthy drink well the other is a high sugar drink.

b)A rule that has low support and high confidence

X = Eggs Y = Diapers {Eggs} →{Diapers} Support = 1/5 = 20%
Confidence = 1/1 = 100%
Such rule set is subjectively interesting due to its not being expected.

## Question 40

**Not yet graded / 3 pts**

Discuss the basic difference between the **agglomerative** and **divisive** hierarchical clustering algorithms and mention which type of hierarchical clustering algorithm is more commonly used.

Your Answer:

Agglomerative Clustering

It is a bottom-up approach in which the sub-components are read first and then move to the parent.Clusters are grouped together to create a large cluster as one moves up the hierarchy.The process of merging continues until all clusters are merged as a single cluster.

Divisive hierarchical clustering;

It is a top-down approach which starts with all objects in one cluster. Clusters are divided again and again to create a smaller cluster. Division continues until all the clusters are having a single object.

Divisive hierarchical algorithm is more complex, efficient and accurate compared to agglomerative clustering as it requires a method for splitting the cluster.

## Question 41

**Not yet graded / 3 pts**

Consider a transaction dataset that contains five items, {A, B, C, D, E}. Suppose **the rules {A, B} → C has the same confidence as {A, B} → D**, which one of the following statements are true or not, and why:

1. a) The confidence of the {A, B} → {C, D} is the same as the confidence of {A, B} → {C}.
2. b) All transactions that contain {A, B, C} also contain {A, B, D}.

Your Answer:

The support is the probability that a randomly selected transaction from the database will contain all items in the antecedent and the consequent, whereas the **confidence is the conditional probability that a randomly selected transaction will include all the items in the consequent**, given that the transaction includes all the items in the antecedent.

1. a.) False: because the antecedent in 1st item set is same as the 2nd item set. And the consequents are different. From the confidence calculations from P(item set)/P(item), we can determine that the

confidence of {A,B} -> {C,D} is not same as the confidence of {A, B} -> {C}.

2. b.) True: because from the rules of {A,B} -> C has the same confidence as {A.B} -> D, we can confirm from the confidence from P(item set)/P(item), that all the transactions that contain {A,B,C} also contains {A,B,C} because the confidence would be same for both the item sets.

## Question 42

**Not yet graded / 3 pts**

Consider the following set of one-dimensional points: **{4, 6, 17, 19, 23, 27, 33, 37}**. Suppose we apply **k-means clustering** to obtain three clusters, A, B, and C. If the three initial centroids are located at **(15, 25, 31)**, respectively. Show: The three clusters and their new three centers after the **first round** of execution.

Your Answer:

| | 15. | 25. | 31 | |
|---|---|---|---|---|
| 4. | 11. | 21. | 27. | A |
| 6 | 9. | 19. | 25. | A |
| 17 | 2. | 8. | 14. | A |
| 19 | 4. | 6. | 12. | A |
| 23 | 8. | 2. | 8. | B |
| 27 | 12. | 2. | 4. | B |
| 33 | 18. | 8. | 2. | C |
| 37. | 22. | 12. | 6. | C |

cluster A= {4,6,17,19}

centroid = (4+6+17+19)/4=11.5

cluster B={23,27}

centroid=(23+27)/2=25

cluster C={33,37}

centroid=(33+37)/2=35

New centroid is (11.5,25, 35)

Quiz Score: **34** out of 50

# Final Exam

**Due** May 9 at 3:45pm **Points** 50 **Questions** 42
**Available** May 9 at 1:15pm - May 9 at 3:45pm about 3 hours **Time Limit** 75 Minutes

# Instructions

- The exam on **modules 7, 8, 9, 10, 11 and 12.**
- The exam will be available on **Monday May 09, 2022 from 1:15 PM to 3:45 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **LATEST** | Attempt 1 | 75 minutes | 32 out of 50 * |

* Some questions not yet graded

ⓘ Correct answers will be available on May 10 at 4pm.

Score for this quiz: **32** out of 50 *
Submitted May 9 at 3:27pm
This attempt took 75 minutes.

### Question 1 — 1 / 1 pts

Why is the XOR problem exceptionally interesting to neural network researchers?

- (●) because it is the simplest linearly inseparable problem that exists
- ( ) because it can be solved by a single layer perceptron

because it is complex binary operation that cannot be solved using neural networks

because it can be expressed in a way that allows you to use a neural network

Incorrect

## Question 2

0 / 1 pts

In feed- forward networks, the connections between layers are _____ from input to output.

- Multidirectional
- Bidirectional
- Directional
- Unidirectional

## Question 3

1 / 1 pts

Why are linearly separable problems of interest of neural network researchers?

because they are the only mathematical functions you can draw

because they are the only mathematical functions that are continue

because they are the only class of problem that perceptron can solve successfully

because they are the only class of problem that network can solve successfully

## Question 4

1 / 1 pts

What is perceptron?

- a single layer feed-forward neural network with pre-processing
- a double layer auto-associative neural network
- a neural network that contains feedback
- an auto-associative neural network

## Question 5

1 / 1 pts

What is true regarding backpropagation rule?

- error in output is propagated backwards only to determine weight updates

- it is also called generalized delta rule
- all of the mentioned
- there is no feedback of signal at any stage

## Question 6

1 / 1 pts

What is meant by generalized in statement “backpropagation is a generalized delta rule” ?

- because delta is applied to only input and output layers, thus making it more simple and generalized
- because delta rule can be extended to hidden layer units
- it has no significance
- none of the mentioned

## Question 7

1 / 1 pts

The maximum margin classifier is associated with which of the following:

- Decision tree

- (selected) Support vector machine
- Linear regression
- Logistic regression

Incorrect

## Question 8

0 / 1 pts

In ROC chart the proportion of false positive fp is on _____ and the proportion of true positive tp is on _____

- The horizontal axis, the vertical axis
- (selected) The vertical axis, the horizontal axis
- The horizontal axis, the x-axis
- The vertical axis, the y-axis

## Question 9

1 / 1 pts

What's the objective of the support vector machine algorithm?

- None of the mentioned
- to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of target variables.

to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of samples.

to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points where N is the number of features.

## Question 10

1 / 1 pts

ROC in performance metrics stands for?

Reverse operating characteristic

Revise operating characteristic

Remote operating characteristic

Receiver operating characteristic

## Question 11

1 / 1 pts

**The SVM's are less effective when -----.**

The data is clean and ready to use

The data is noisy and contains overlapping points

The data is linearly separable

All of the mentioned

## Question 12

1 / 1 pts

Which of the following is not a correct statement for SVM.

SVMs can be reused as algorithms for learning linear classifiers

Instances closest to the maximum margin hyperplane are called support vectors

SVMs are resilient to overfitting

All instances are required to define the maximum margin hyperplane.

## Question 13

1 / 1 pts

Which of the following can affect the complexity of Apriori?

Dimensionality of the given data set

Number of transactions in the database

All of the mentioned

Maximum number of items in the transactions

## Question 14

1 / 1 pts

Which of the following best describes lift in knowledge discovery?

- (●) A measure of interestingness of a rule
- ( ) An unsupervised learning approach
- ( ) A known class attribute
- ( ) A data mining technique

## Question 15

1 / 1 pts

In association rules, what is meant by the term support?

- ( ) A frequent item set.
- (●) The number of instances correctly covered by the association rule.
- ( ) An attribute-value pair.
- ( ) Combinations of attribute-value pairs that have a minimum coverage.

Incorrect

## Question 16

0 / 1 pts

A rule with a lower value of confidence and support could be

preferred because:

- It indicates novelty
- None of the mentioned
- Such rules are bound to hold throughout the dataset
- Such rules are more interesting

## Question 17

1 / 1 pts

Which of the following would be appropriate for a data mining

algorithm aimed at discovering which groups of products consumers

would tend to purchase together?

- Decision trees
- Classification rules
- Decision rules
- Association rules

Incorrect

## Question 18

0 / 1 pts

Which of the following best describes the Apriori principle?

- When the anti-monotone property of support holds on given itemset
- All of the mentioned
- If an itemset is frequent, then all of its subsets must also be frequent
- Support of an itemset never exceeds the support of its subsets

## Question 19

1 / 1 pts

_____ is a clustering procedure characterized by the development of a tree-like structure.

- Hierarchical clustering
- Non-hierarchical clustering
- K-Means clustering
- K-Medoids clustering

## Question 20

1 / 1 pts

Which one of the following correctly defines the term cluster?

- Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm
- (selected) Group of similar objects that differ significantly from other objects
- Symbolic representation of facts or ideas from which information can potentially be extracted
- All of the mentioned

## Question 21

1 / 1 pts

A _____ is a tree diagram for displaying clustering results. Vertical lines represent clusters that are joined together.

- (selected) Dendrogram
- Tree plot
- Histogram
- Scatter plot

Incorrect

## Question 22

0 / 1 pts

Which of the following is true about cluster analysis?

- Cluster analysis is the process of ungrouping objects into subsets that have meaning in the context of a particular problem
- It can't uncover previously undetected relationships in a complex dataset.
- Clustering is referred to as a supervised learning method
- [x] Clustering is referred to as an unsupervised learning method.

## Question 23

1 / 1 pts

Which of the following is not an application of cluster analysis?

- [x] Decide about the subsequent purchases after buying a PC.
- Observing earth quake epicenters should be clustered along continent faults.
- Identifying groups of motor insurance policy holders with a high average claim cost.
- Help marketers discover distinct groups in their customer bases.

## Question 24

1 / 1 pts

Which clustering algorithm starts with each cluster comprising exactly one observation and then progressively combines the two nearest clusters until there is just one cluster left at the end?

- (●) Agglomerative clustering
- ( ) Divisive clustering
- ( ) Complete Linkage clustering
- ( ) Single Linkage clustering

## Question 25

1 / 1 pts

Which one of the following statements about the K-means clustering is incorrect?

- ( ) K-means clustering can be defined as the method of quantization
- ( ) The goal of the k-means clustering is to partition (n) observation into (k) clusters
- (●) The nearest neighbor is the same as the K-means
- ( ) All of the mentioned

## Question 26

1 / 1 pts

Which one of the following can be considered as the final output of the hierarchal type of clustering?

- (●) A tree which displays how the close thing are to each other
- ( ) Finalize estimation of cluster centroids
- ( ) Assignment of each point to clusters
- ( ) None of the mentioned

## Question 27

1 / 1 pts

One reason of anomaly detection is:

- ( ) Errors from collecting data
- ( ) Normal variations can be seen on data
- (●) All of the mentioned
- ( ) Data coming from different classes

## Question 28

1 / 1 pts

Which of the following will be Euclidean Distance between the two data point A(1, 3) and B(2, 3)?

- 4
- 1
- 2
- 8

## Question 29

1 / 1 pts

Data may contain erroneous or anomalous values, which are usually referred to as_____.

- Reduction
- Noise
- Outliers
- Inconsistencies

## Question 30

1 / 1 pts

For the purpose of anomaly detection, in the reconstruction-based approaches we need to -----.

use certain kernel function on the given data to construct such a model

reduce data to lower dimensional data

All of the mentioned

assume data comes from normal distribution

## Question 31

1 / 1 pts

One of the drawbacks of using density methods in anomaly detection is:

All of the mentioned

Density may become less meaningful in high-dimensional space.

It may be hard to estimate the true density distribution for high dimensional data.

Sometime it can be sensitive to variations in density

## Question 32

1 / 1 pts

An observation that is extreme, being distant from the rest of the data is termed a ------.

- Feature
- Class
- Outlier
- Predictor

## Question 33

1 / 1 pts

Which of the following is not an example application of data mining in science and engineering?

- Use sequential pattern mining to investigate changes in customer consumption or loyalty
- Mining of spatiotemporal, biological, diverse semantics and relationships C)
- All of the mentioned
- Use data mining in monitoring systems, software bugs and network intrusion

## Question 34

1 / 1 pts

Which of the following is not trend of data mining?

- Mining multimedia, text and web data
- Distributed data mining an real-time data stream mining
- All of the mentioned
- Using data mining tasks of customers for targeted marketing

Incorrect

## Question 35

0 / 1 pts

Which of the following is a new trend in data mining?

- Invisible data mining
- A Web mining
- All of the mentioned
- Scalable data mining methods

## Question 36

1 / 1 pts

Which of the following is one of the purposes of the visualization?

- All of the mentioned
- It assists to search for trends and relationships among data.
- It helps find interesting regions for any further analysis.
- It can provide qualitative overview of large data sets

## Question 37

1 / 1 pts

Which of the following describes an example of the factor analysi?

- All of the mentioned
- For certain data, researcher can indirectly measure other quantities that reflect the factor of interest
- For given experimental data, one analyzes the data for two or more populations described by a numeric response variable and one or more categorical variables (factors)
- For special type of data, one attempts to determine several discriminant functions (factors) that discriminate among the groups defined by the response variable

## Question 38

1 / 1 pts

Which of the following is not part of Web Mining:

- Structure Mining
- Content Mining
- Database Mining
- Usage Mining

## Question 39

Not yet graded / 3 pts

In your opinion, what are the major **5 trends** in data mining research today? Name one **major issue** in data mining, which in your view, may have a strong impact on society.

Your Answer:

Multimedia Data

Mining This is one of the latest methods which is catching up because of the growing ability to capture useful data accurately.

Ubiquitous Data Mining

This method involves the mining of data from mobile devices to get information about individuals.

Distributed Data Mining

This type of data mining is gaining popularity as it involves the mining of huge amount of information stored in different company locations or at different organizations.

Spatial and Geographic Data Mining

This is new trending type of data mining which includes extracting information from environmental, astronomical, and geographical data

which also includes images taken from outer space.

Time Series and Sequence Data Mining

This practice is also helpful in analyzing even random events which occur outside the normal series of events.

One major issue in Data Mining in my opinion is:

Mining different kinds of knowledge in databases − Different users may be interested in different kinds of knowledge. Therefore it is necessary for data mining to cover a broad range of knowledge discovery task.

## Question 40

**Not yet graded / 3 pts**

We generally will be more interested in association rules with **high confidence.** However, often we will not be interested in association rules that have a confidence of **100%**. Why? Then specifically explain why association rules with 99% confidence may be interesting (i.e., **what might they indicate**)?

Your Answer:

Confidence levels tell us how many times our parameter of interest is captured by the confidence interval on a long run. It means that 99% confidence means we expect 99 times out of 100 that our confidence interval contains the actual parameter. So, 100% means we expect 100 times out of 100 that our confidence interval contains the actual parameter which means our parameter must be present in the interval always. If we take samples 100 times or 100 times or even 100000 times, we should always get success and our results must never be wrong. But this is not true in practice. Real life situations are always surrounded with some

uncertainties and hence inaccuracies. So, we almost always get some error and are never 100% true. It's like "nobody is perfect" in literal sense.

So, we prefer rules with 99% confidence to allow inaccuracies up-to 1% and produce correct results rest of the 99% of the times. This is more close to real life scenario as compared to 100%. In some cases, where we know uncertainty is higher, we may choose confidence levels of 95% or even 90%.

## Question 41

**Not yet graded / 3 pts**

A database has 4 transactions, shown below.

| TID | Items bought |
|---|---|
| T100 | {K, A, D, B} |
| T200 | {D, A, C, E, B} |
| T300 | {C, A, B, E} |
| T400 | {B, A, D} |

Assuming a minimum level of support **min_sup = 60%** and a minimum level of confidence **min_conf = 80%**.

Given the **frequent itemsets** are: **{{A}, {B}, {D}, {A, B}, {B, D}, {A, B, D}}**.

List all of the **strong association rules** (with support *s* and confidence *c*) **matching the following metarule**, where *X* is a variable representing customers, and $item_i$ denotes variables representing items (e.g., *"A", "B"*, etc.):

$\forall X \in transaction,\ buys(X,item1) \wedge buys(X,item2) \Rightarrow buys(X,item3)[s,c]$

Your Answer:

buys(X, A) ^ buys(X, B) → buys(X, D) (75%, 75%) Not Strong

buys(X, A) ^ buys(X, D) → buys(X, B) (75%, 100%) Strong

buys(X, B) ^ buys(X, D) → buys(X, A) (75%, 100%) Strong

## Question 42

**Not yet graded / 3 pts**

For each of the Boolean functions given below, state whether the problem is **linearly separable** or not, and why?

1. a) A AND B AND C
2. b) NOT A AND B
3. c) (A OR B) AND (A OR C)

Your Answer:

1) The statement (A AND B AND C) contains the AND operation. So it is linearly separable.

2) The statement NOT A AND B contains the NOT and AND operation, so it is linearly separable.

3) The statement (A OR B) AND (A OR C) contains the OR and AND operation so it is linearly separable.

Because for all the inputs of A, B, C the X-Y plan is separated by one single straight line in AND, OR, NOT Operation

Quiz Score: **32** out of 50

# Practice Test for Final Exam

**Due** May 9 at 11:59pm **Points** 8 **Questions** 6
**Available** after May 1 at 10pm **Time Limit** 10 Minutes **Allowed Attempts** 3

# Instructions

- The practice test for the final exam on **modules 7, 8, 9, 10, 11, and 12.**
- You need to answer **5 MCQs** + **1 Short question .**
- You will have only **10 minutes** to complete your test in **one sitting.**
- **You have up to 3 attempts, and be aware that in the final you have only one attempt**

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **KEPT** | Attempt 3 | less than 1 minute | 5 out of 8 * |
| **LATEST** | Attempt 3 | less than 1 minute | 5 out of 8 * |
| | Attempt 2 | less than 1 minute | 4 out of 8 * |
| | Attempt 1 | 7 minutes | 3 out of 8 * |

* Some questions not yet graded

Submitted May 10 at 10:30am

## Question 1 — 1 / 1 pts

**What is true regarding backpropagation rule?**

Correct!

- (●) all of the mentioned
- ( ) it is also called generalized delta rule
- ( ) error in output is propagated backwards only to determine weight updates
- ( ) there is no feedback of signal at any stage

## Question 2

1 / 1 pts

**For SVM, which options are correct?**

Correct!

- [x] Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane
- [ ] None of the mentioned.
- [ ] Deleting the support vectors won't change the position of the hyperplane
- [ ] Support vectors are data points that are far away from the hyperplane and influence the position and orientation of the hyperplane

## Question 3

1 / 1 pts

**Which of the following is not an example of frequent pattern analysis?**

- [ ] What kinds of DNA are sensitive to this new drug?
- [ ] What are the subsequent purchases after buying a PC?
- [ ] Can we automatically classify web documents?

Correct!

- [x] Can we predict the winner of match?

## Question 4

1 / 1 pts

**Which one of the following can be considered as the final output of the hierarchal type of clustering?**

- Finalize estimation of cluster centroids
- Assignment of each point to clusters
- None of the mentioned

Correct!

- A tree which displays how the close thing are to each other

## Question 5

1 / 1 pts

**One of the drawbacks of using statistical methods in anomaly detection is:**

- Sometime it can be Can difficult to decide on number of clusters

Correct!

- It may be hard to estimate the true distribution for high dimensional data.
- Density may become less meaningful in high-dimensional space.
- all of the mentioned

Unanswered

## Question 6

Not yet graded / 3 pts

**For each of the following questions, provide an example of an association rule from the market basket domain that satisfies the following conditions. Also, describe whether such rules are subjectively interesting or not.**

a) A rule that has reasonably high support but low confidence.

b) A rule that has low support and high confidence.

Your Answer: