Final Exam

Started: May 6 at 3:29pm

Quiz Instructions

- The exam on **modules 7, 8, 9, 10, 11 and 12.**
- The exam will be available on **Monday May 06, 2024 from 1:00 PM to 5:00 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.**

Question 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 2 1 pts

Layers between the input and output layers are known as:

- ○ Output layer
- ○ Multilayer
- ○ Resultant layer
- ○ Hidden layer

Question 3 1 pts

What is perceptron?

- ○ an auto-associative neural network

○ a double layer auto-associative neural network

○ a neural network that contains feedback

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

Question 4 1 pts

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

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

○ 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

Question 5 1 pts

How can learning process be stopped in backpropagation rule?

○ none of the mentioned

○ on basis of average gradient value

○ no heuristic criteria exist

○ there is convergence involved

Question 6 1 pts

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

○ Multidirectional

○ Bidirectional

○
Unidirectional

○
Directional

Question 7 1 pts

ROC chart is a _____ plot.

○
Two-dimensional

○
Three-dimensional

○
One-dimensional

○
Multi-dimensional

Question 8 1 pts

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

○
Decision tree

○
Linear regression

○
Logistic regression

○
Support vector machine

Question 9 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

○
All instances are required to define the maximum margin hyperplane.

SVMs are resilient to overfitting

Question 10 1 pts

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

y = a0 + a1x12 + a2x2

y = a0 + a1x1 + a2x2

y = a0 + a1/x1 + a2/x2

y = a0 + a1x1 + a2x22

Question 11 1 pts

ROC in performance metrics stands for?

Receiver operating characteristic

Reverse operating characteristic

Remote operating characteristic

Revise operating characteristic

Question 12 1 pts

For SVM, which options are correct?

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

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

Question 13 1 pts

Which of the following can affect the complexity of Apriori?

○
All of the mentioned

○
Maximum number of items in the transactions

○
Number of transactions in the database

○
Dimensionality of the given data set

Question 14 1 pts

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

○
All of the mentioned

○
Frequent sub-sequence

○
Frequent sub-items

○
Frequent sub-structure

Question 15 1 pts

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

○
Maximal frequent itemset

○
Maximal itemset

○
Closed itemset

○
Closed frequent itemset

Question 16 1 pts

Which of the following best describes the Apriori principle?

○
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

○
If an itemset is frequent, then all of its subsets must also be frequent

Question 17 1 pts

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

○
Decision tree.

○
Naïve Bayesian.

○
K-means

○
Apriori.

Question 18 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 automatically classify web documents?

○
What kinds of DNA are sensitive to this new drug?

○
Can we predict the winner of match?

Question 19 1 pts

From the following which method is not the clustering method?

○
Hierarchical

Density based

Divide-and-conquer based

Partition

Question 20 1 pts

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

All of the mentioned

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

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

Question 21 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

Partition method

Grid method

Question 22 1 pts

Which of the following statement is NOT true about clustering?

It groups the data

It is a supervised learning technique

○
It uses clusters for data analysis

○
It is an unsupervised learning technique

Question 23 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?

○
Single Linkage clustering

○
Agglomerative clustering

○
Divisive clustering

○
Complete Linkage clustering

Question 24 1 pts

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

○
Hierarchical method

○
Partition method

○
Grid method

○
Density-based method

Question 25 1 pts

Which statement is not true about cluster analysis?

○
Cluster analysis is also called classification analysis or numerical taxonomy.

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

○
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.

Question 26 1 pts

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

○
Doesn't maximize classification rate

○
Minimizes classification rate

○
Maximizes misclassification rate

○
Provides smoothing that reduces the risk of over fitting

Question 27 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

○
Outlier Analysis

○
Evaluation Analysis

○
Classification

Question 28 1 pts

One of the drawbacks of using clustering in anomaly detection is:

○
Density may become less meaningful in high-dimensional space.

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

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

All of the mentioned

Question 29 1 pts

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

All of the mentioned

Sometime it can be sensitive to variations in density

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

Density may become less meaningful in high-dimensional space.

Question 30 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

Question 31 1 pts

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

8

2

4

1

Question 32 1 pts

One reason of anomaly detection is:

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

Question 33 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 34 1 pts

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

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

Question 35 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.

○
It can provide qualitative overview of large data sets

○
All of the mentioned

Question 36 1 pts

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

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

○
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)

Question 37 1 pts

Which of the following is not part of Web Mining:

○
Structure Mining

○
Usage Mining

○
Database Mining

○
Content Mining

Question 38 1 pts

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

○
All of the mentioned

○
Personal information is encrypted and stored at different locations

Add noise to the data in order to mask some attribute values of records

Removing sensitive features or fields associated with the data

Question 39 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.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 40 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.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words </>

Question 41 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.

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words </>

Question 42 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]*

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Not saved Submit Quiz