Questions for Chapter 1

Multiple Choice Questions

*(1.1)*

1. The process of forming general concept definitions from examples of concepts to be learned.

1. deduction
2. abduction
3. induction
4. conjunction

2. Data mining is best described as the process of

1. identifying structure in data.
2. deducing relationships in data.
3. representing data.
4. simulating trends in data.

3. Data \_\_\_\_\_\_\_ or \_\_\_\_\_\_\_analytics is defined as the process of extracting meaninful knowledge from data.

* 1. mining, machine
  2. discovery, knowledge
  3. mining, scientific
  4. science, data

*(1.2)*

4. Computers are best at learning

1. facts.
2. concepts.
3. procedures.
4. principles.

5. Like the probabilistic view, the \_\_\_\_\_\_\_\_ view allows us to associate a probability of membership with each classification.

a. exemplar

b. deductive

1. classical
2. inductive

6. Data used to build a data mining model.

a. validation data

b. training data

1. test data
2. hidden data

7. Suppose the following rule is derived from a data set of 100 individuals.

IF age < 25 and gender = male THEN life insurance policy = no

Rule precision: 80%

Rule coverage: 50%

What can we conclude from this rule?

1. 80% of all males who are less than 25 do not have life insurance.
2. 50 of the 100 individuals are males less than 25 years of age.
3. 40 individuals satisfy both the rule antecedent and consequent conditions.
4. All of the above statements are correct.

8. Supervised learning differs from unsupervised clustering in that supervised learning requires

1. at least one input attribute.
2. input attributes to be categorical.
3. at least one output attribute.
4. ouput attriubutes to be categorical.

9. Which of the following is a valid rule for the decision tree below?



1. IF Business Appointment = No & Temp above 70 = No

THEN Decision = wear slacks

1. IF Business Appointment = Yes & Temp above 70 = Yes

THEN Decision = wear shorts

1. IF Temp above 70 = No

THEN Decision = wear shorts

1. IF Business Appointment= No & Temp above 70 = No

THEN Decision = wear jeans

*(1.3)*

10. Database query is used to uncover this type of knowledge.

1. deep
2. hidden
3. shallow
4. multidimensional

11. A statement to be tested.

1. theory
2. procedure
3. principle
4. hypothesis

*(1.4)*

1. A person trained to interact with a human expert in order to capture their knowledge.

a. knowledge programmer

b. knowledge developer

c. knowledge engineer

d. knowledge extractor

13. Expert systems are able to ­­­­\_\_\_\_\_ the problem solving methods of a human expert.

a. emulate

b. develop

c. diagnose

d. evaulate

*(1.5)*

1. A nearest neighbor approach is best used
   1. with large-sized datasets.
   2. when irrelevant attributes have been removed from the data.
   3. when a generalized model of the data is desireable.
   4. when an explanation of what has been found is of primary importance.
2. The nearest neighbor approach
   1. is computationally independent of dataset size.
   2. requires a numeric output attribute.
   3. is limited to classifying datasets with numeric input attributes.
   4. stores instances rather than a generalized model of the data.

*(1.6)*

1. Data analytics is often defined as a five-step process. The five steps in their correct order are
   1. preprocess data, model data, interpret results, evaluate results, report results
   2. acquire data, preprocess data, model data, interpret and evaluate results, report results
   3. preprocess data, mine the data, evaluate results, report results, apply results
   4. acquire data, model data, interpret results, evaluate results, apply results
2. Which of the following is not a characteristic of a data warehouse?
   1. contains historical data
   2. designed for decision support
   3. stores data in normalized tables
   4. promotes data redundancy
3. The correlation between the number of years an employee has worked for a company and the salary of the employee is 0.80. What can be said about employee salary and years worked?
   1. There is no relationship between salary and years worked.
   2. Individuals that have worked for the company the longest have higher salaries.
   3. Individuals that have worked for the company the longest have lower salaries.
   4. The majority of employees have been with the company a long time.
   5. The majority of employees have been with the company a short period of time.
4. The correlation coefficient for two real-valued attributes is –0.85. What does this value tell you?
   1. The attributes are not linearly related.
   2. As the value of one attribute increases the value of the second attribute also increases.
   3. As the value of one attribute decreases the value of the second attribute increases.
   4. The attributes show a curvilinear relationship.
5. A structure designed to store data for decision support.
   1. operational database
   2. flat file
   3. decision tree
   4. data warehouse

*(1.7)*

1. A term to describe bias, noise, or abnormality in the data is
   1. Volume
   2. Variety
   3. Veracity
   4. Velocity
2. The primary components of the Hadoop open-source distributed computing environment are
   1. a distributed data storage system and a system for data processing
   2. a distributed data warehouse and a network package
   3. a system for cloud computing and a distributed system for data processing
   4. a preprocessing system and a distributed system for data processing
3. Which statement is true about cloud computing?
   1. Is of limited used in a distributed computing environment.
   2. It delivers computing resources over the internet.
   3. It is a necessary component of a distributed computing environment.
   4. More than one of a, b, or c is a true statement about cloud computing.

*(1.8)*

1. An approach to employee staffing that perceives people as assets.
   1. Human Data Resourcing
   2. Positive Resource Management
   3. Employee Maintenance Quality
   4. Human Capital Management
2. Deducing private information from publicly available data is known as
   1. the privacy issue
   2. assumption deduction
   3. the inference problem
   4. unauthorized supposition

*(1.9)*

1. A term used to describe the process whereby a customer discontinues the use of a service or subscription with one company in order to initialize the same service with another company is known as customer
   1. acumen
   2. bias
   3. prejudice
   4. churn
2. If a customer is spending more than expected, the customer’s intrinsic value is \_\_\_\_\_\_\_\_ their actual value.
   1. greater than
   2. less than
   3. less than or equal to
   4. equal to

Matching Questions

Determine which is the best approach for each problem.

1. *supervised learning*
2. *unsupervised clustering*
3. *data query*
   1. What is the average weekly salary of all female employees under forty years of age?
   2. Develop a profile for credit card customers likely to carry an average monthly balance of more than $1000.00.
   3. Determine the characteristics of a successful used car salesperson.
   4. What attribute similarities group customers holding one or several insurance policies?
   5. Do meaningful attribute relationships exist in a database containing information about credit card customers?
   6. Do single men play more golf than married men?
   7. Determine whether a credit card transaction is valid or fraudulent.

Answers to Chapter 1 Questions

Multiple Choice Questions

* 1. c
  2. a
  3. d
  4. b
  5. a
  6. b
  7. d
  8. c
  9. d
  10. c
  11. d
  12. c
  13. a
  14. b
  15. d
  16. b
  17. c
  18. b
  19. c
  20. d
  21. c
  22. a
  23. b
  24. d
  25. c
  26. d
  27. b

Matching Questions

* 1. c
  2. a
  3. a
  4. a
  5. b
  6. c
  7. a