

Chapter 1

Multiple Choice Questions

(1.1)

1. The process of forming general concept definitions from examples of concepts to be learned.
 - a. deduction
 - b. abduction
 - c. induction
 - d. conjunction
2. Data mining is best described as the process of
 - a. identifying patterns in data.
 - b. deducing relationships in data.
 - c. representing data.
 - d. simulating trends in data.

(1.2)

3. Computers are best at learning
 - a. facts.
 - b. concepts.
 - c. procedures.
 - d. principles.

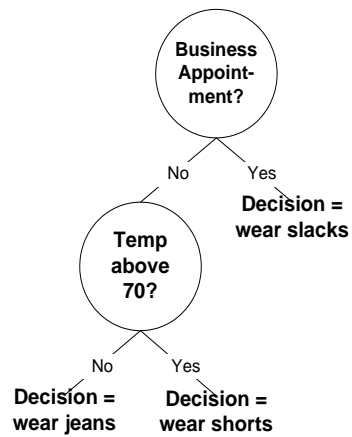
4. Like the probabilistic view, the _____ view allows us to associate a probability of membership with each classification.
 - a. exemplar
 - b. deductive
 - c. classical
 - d. inductive

5. Data used to build a data mining model.
 - a. validation data
 - b. training data
 - c. test data
 - d. hidden data

6. Supervised learning and unsupervised clustering both require at least one
 - a. hidden attribute.
 - b. output attribute.
 - c. input attribute.
 - d. categorical attribute.

7. Supervised learning differs from unsupervised clustering in that supervised learning requires
 - a. at least one input attribute.
 - b. input attributes to be categorical.
 - c. at least one output attribute.
 - d. output attributes to be categorical.

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



- a. IF Business Appointment = No & Temp above 70 = No
THEN Decision = wear slacks
- b. IF Business Appointment = Yes & Temp above 70 = Yes
THEN Decision = wear shorts
- c. IF Temp above 70 = No
THEN Decision = wear shorts
- d. IF Business Appointment= No & Temp above 70 = No
THEN Decision = wear jeans

(1.3)

9. Database query is used to uncover this type of knowledge.
- a. deep
 - b. hidden
 - c. shallow
 - d. multidimensional

10. A statement to be tested.

- a. theory
- b. procedure
- c. principle
- d. hypothesis

(1.4)

11. 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

(1.5)

12. Which of the following is not a characteristic of a data warehouse?

- a. contains historical data
- b. designed for decision support
- c. stores data in normalized tables
- d. promotes data redundancy

13. A structure designed to store data for decision support.

- a. operational database
- b. flat file
- c. decision tree
- d. data warehouse

(1.6)

14. A nearest neighbor approach is best used
- with large-sized datasets.
 - when irrelevant attributes have been removed from the data.
 - when a generalized model of the data is desirable.
 - when an explanation of what has been found is of primary importance.

(1.7)

15. If a customer is spending more than expected, the customer's intrinsic value is _____ their actual value.
- greater than
 - less than
 - less than or equal to
 - equal to

Matching Questions

Determine which is the best approach for each problem.

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

Answers to Chapter 1 Questions

Multiple Choice Questions

1. c
2. a
3. b
4. a
5. b
6. c
7. c
8. d
9. c
10. d
11. c
12. c
13. d
14. b
15. b

Matching Questions

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

Chapter 2

Multiple Choice Questions

(2.1)

1. Another name for an output attribute.
 - a. predictive variable
 - b. independent variable
 - c. estimated variable
 - d. dependent variable
2. Classification problems are distinguished from estimation problems in that
 - a. classification problems require the output attribute to be numeric.
 - b. classification problems require the output attribute to be categorical.
 - c. classification problems do not allow an output attribute.
 - d. classification problems are designed to predict future outcome.
3. Which statement is true about prediction problems?
 - a. The output attribute must be categorical.
 - b. The output attribute must be numeric.
 - c. The resultant model is designed to determine future outcomes.
 - d. The resultant model is designed to classify current behavior.
4. Which statement about outliers is true?
 - a. Outliers should be identified and removed from a dataset.
 - b. Outliers should be part of the training dataset but should not be present in the test data.
 - c. Outliers should be part of the test dataset but should not be present in the training data.
 - d. The nature of the problem determines how outliers are used.
 - e. More than one of a,b,c or d is true.

(2.2)

5. Assume that we have a dataset containing information about 200 individuals. One hundred of these individuals have purchased life insurance. A supervised data mining session has discovered the following rule:

IF age < 30 & credit card insurance = yes

THEN life insurance = yes

Rule Accuracy: 70%

Rule Coverage: 63%

How many individuals in the class *life insurance* = no have credit card insurance and are less than 30 years old?

- a. 63
 - b. 70
 - c. 30
 - d. 27
6. Which statement is true about neural network and linear regression models?
- a. Both models require input attributes to be numeric.
 - b. Both models require numeric attributes to range between 0 and 1.
 - c. The output of both models is a categorical attribute value.
 - d. Both techniques build models whose output is determined by a linear sum of weighted input attribute values.
 - e. More than one of a,b,c or d is true.

(2.3)

7. Unlike traditional production rules, association rules
- a. allow the same variable to be an input attribute in one rule and an output attribute in another rule.
 - b. allow more than one input attribute in a single rule.
 - c. require input attributes to take on numeric values.
 - d. require each rule to have exactly one categorical output attribute.

(2.4)

8. Which of the following is a common use of unsupervised clustering?
- a. detect outliers
 - b. determine a best set of input attributes for supervised learning
 - c. evaluate the likely performance of a supervised learner model
 - d. determine if meaningful relationships can be found in a dataset
 - e. All of a,b,c, and d are common uses of unsupervised clustering.

(2.5)

9. The average positive difference between computed and desired outcome values.
- a. root mean squared error
 - b. mean squared error
 - c. mean absolute error
 - d. mean positive error
10. Given desired class C and population P , lift is defined as
- a. the probability of class C given population P divided by the probability of C given a sample taken from the population.
 - b. the probability of population P given a sample taken from P .
 - c. the probability of class C given a sample taken from population P .
 - d. the probability of class C given a sample taken from population P divided by the probability of C within the entire population P .

Fill in the Blank

Use the three-class confusion matrix below to answer questions 1 through 3.

Computed Decision			
	Class 1	Class 2	Class 3
Class 1	10	5	3
Class 2	5	15	3
Class 3	2	2	5

1. What percent of the instances were correctly classified?
2. How many *class 2* instances are in the dataset?
3. How many instances were incorrectly classified with *class 3*?

Use the confusion matrix for Model X and confusion matrix for Model Y to answer questions 4 through 6.

Model X	Computed Accept	Computed Reject		Model Y	Computed Accept	Computed Reject
Accept	10	5		Accept	6	9
Reject	25	60		Reject	15	70

4. How many instances were classified as an accept by Model X?
5. Compute the lift for Model Y.
6. You will notice that the lift for both models is the same. Assume that the cost of a false reject is significantly higher than the cost of a false accept. Which model is the better choice?

Answers to Chapter 2 Questions

Multiple Choice Questions

1. d
2. b
3. c
4. d
5. d
6. a
7. a
8. e
9. c
10. d

Fill in the Blank

1. 60%
2. 23
3. 6
4. 35
5. $8/7$
6. Model X

Chapter 3

Multiple Choice Questions

(3.1)

1. A data mining algorithm is *unstable* if
 - a. test set accuracy depends on the ordering of test set instances.
 - b. the algorithm builds models unable to classify outliers.
 - c. the algorithm is highly sensitive to small changes in the training data.
 - d. test set accuracy depends on the choice of input attributes.
2. Which statement is true about the decision tree attribute selection process described in your book?
 - a. A categorical attribute may appear in a tree node several times but a numeric attribute may appear at most once.
 - b. A numeric attribute may appear in several tree nodes but a categorical attribute may appear at most once.
 - c. Both numeric and categorical attributes may appear in several tree nodes.
 - d. Numeric and categorical attributes may appear in at most one tree node.

(3.2)

3. Given a rule of the form IF X THEN Y, rule *confidence* is defined as the conditional probability that
 - a. Y is true when X is known to be true.
 - b. X is true when Y is known to be true.
 - c. Y is false when X is known to be false.
 - d. X is false when Y is known to be false.

4. Association rule *support* is defined as
- the percentage of instances that contain the antecedent conditional items listed in the association rule.
 - the percentage of instances that contain the consequent conditions listed in the association rule.
 - the percentage of instances that contain all items listed in the association rule.
 - the percentage of instances in the database that contain at least one of the antecedent conditional items listed in the association rule.

Use these tables to answer questions 5 and 6.

Single Item Sets	Number of Items
Magazine Promo = Yes	7
Watch Promo = No	6
Life Ins Promo = Yes	5
Life Ins Promo = No	5
Card Insurance = No	8
Sex = Male	6

Two Item Sets	Number of Items
Magazine Promo = Yes & Watch Promo = No	4
Magazine Promo = Yes & Life Ins Promo = Yes	5
Magazine Promo = Yes & Card Insurance = No	5
Watch Promo = No & Card Insurance = No	5

5. One two-item set rule that can be generated from the tables above is:

If Magazine Promo = Yes Then Life Ins promo = Yes

The confidence for this rule is:

- 5 / 7
- 5 / 12
- 7 / 12
- 1

6. Based on the two-item set table, which of the following is *not* a possible two-item set rule?
- a. IF Life Ins Promo = Yes THEN Magazine Promo = Yes
 - b. IF Watch Promo = No THEN Magazine Promo = Yes
 - c. IF Card Insurance = No THEN Magazine Promo = Yes
 - d. IF Life Ins Promo = No THEN Card Insurance = No

(3.3)

7. Which statement is true about the K-Means algorithm?
- a. All attribute values must be categorical.
 - b. The output attribute must be categorical.
 - c. Attribute values may be either categorical or numeric.
 - d. All attributes must be numeric.
8. The K-Means algorithm terminates when
- a. a user-defined minimum value for the summation of squared error differences between instances and their corresponding cluster center is seen.
 - b. the cluster centers for the current iteration are identical to the cluster centers for the previous iteration.
 - c. the number of instances in each cluster for the current iteration is identical to the number of instances in each cluster of the previous iteration.
 - d. the number of clusters formed for the current iteration is identical to the number of clusters formed in the previous iteration.

(3.4)

9. A genetic learning operation that creates new population elements by combining parts of two or more existing elements.
- a. selection
 - b. crossover
 - c. mutation
 - d. absorption

10. An evolutionary approach to data mining.
 - a. backpropagation learning
 - b. genetic learning
 - c. decision tree learning
 - d. linear regression
11. The computational complexity as well as the explanation offered by a genetic algorithm is largely determined by the
 - a. fitness function
 - b. techniques used for crossover and mutation
 - c. training data
 - d. population of elements

(3.5)

12. This approach is best when we are interested in finding all possible interactions among a set of attributes.
 - a. decision tree
 - b. association rules
 - c. K-Means algorithm
 - d. genetic learning

Computational Questions

1. Construct a decision tree with root node *Type* from the data in the table below. The first row contains attribute names. Each row after the first represents the values for one data instance. The output attribute is *Class*.

Scale	Type	Shade	Texture	Class
One	One	Light	Thin	A
Two	One	Light	Thin	A
Two	Two	Light	Thin	B
Two	Two	Dark	Thin	B
Two	One	Dark	Thin	C
One	One	Dark	Thin	C
One	Two	Light	Thin	C

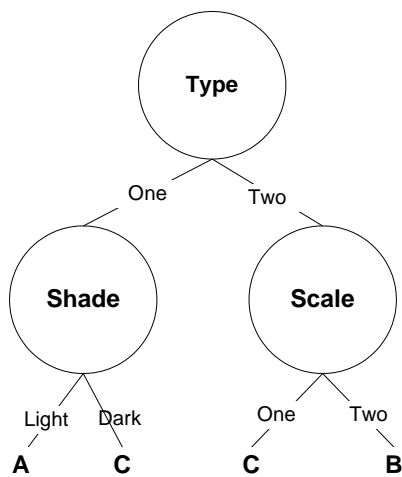
Answers to Chapter 3 Questions

Multiple Choice Questions

1. c
2. b
3. a
4. c
5. a
6. d
7. d
8. b
9. b
10. b
11. a
12. b

Computational Questions

1. Here is the tree.



Chapter 4

Multiple Choice Questions

(4.1)

1. This iDA component allows us to decide if we wish to process an entire dataset or to extract a representative subset of the data for mining.
 - a. preprocessor
 - b. heuristic agent
 - c. ESX
 - d. RuleMaker

(4.2)

2. ESX represents the overall similarity of the exemplars contained in an individual class by computing a ____ score.
 - a. class resemblance
 - b. class predictability
 - c. class predictiveness
 - d. typicality

(4.3)

3. The first row of an *iDAV* formatted file contains attribute names. The second row reflects attribute types. What is specified in the third row of an *iDAV* formatted file?
 - a. attribute predictability
 - b. attribute tolerance
 - c. attribute similarity
 - d. attribute usage

(4.4)

4. A dataset of 1000 instances contains one attribute specifying the color of an object. Suppose that 800 of the instances contain the value red for the color attribute. The remaining 200 instances hold green as the value of the color attribute. What is the domain predictability score for *color = green*?
 - a. 0.80
 - b. 0.20
 - c. 0.60
 - d. 0.40

5. Which relationship is likely to be seen with an interesting clustering of data instances?
 - a. The domain resemblance score is greater than the resemblance scores for the individual clusters.
 - b. The domain resemblance score is equal to the average of the resemblance scores for the individual clusters.
 - c. The resemblance scores for all formed clusters are greater than zero.
 - d. The domain resemblance score is less than the resemblance scores for the individual clusters.

6. A particular categorical attribute value has a predictiveness score of 1.0 and a predictability score of 0.50. The attribute value is
 - a. necessary but not sufficient for class membership.
 - b. sufficient but not necessary for class membership.
 - c. necessary and sufficient for class membership.
 - d. neither necessary nor sufficient for class membership.

7. A particular categorical attribute value has a predictiveness score of 0.5 and a predictability score of 1.0. The attribute value is
 - a. necessary but not sufficient for class membership.
 - b. sufficient but not necessary for class membership.
 - c. necessary and sufficient for class membership.
 - d. neither necessary nor sufficient for class membership.

8. A particular categorical attribute value has a predictiveness score of 0.3 and a predictability score of 0.3. The attribute value is
- necessary but not sufficient for class membership.
 - sufficient but not necessary for class membership.
 - necessary and sufficient for class membership.
 - neither necessary nor sufficient for class membership.

Concept class C_1 shows the following information for the categorical attribute Risk Factor. Use this information to answer questions 9 and 10.

<u>Attribute Name</u>	<u>Value</u>	<u>Frequency</u>
Risk factor	High Risk	25
	Medium Risk	10
	Low Risk	5

9. What is the predictability score for the attribute value *medium risk*?
- 0.10
 - 0.20
 - 0.25
 - 0.50
10. Suppose that the predictiveness score for *risk factor = medium risk* is 0.50. How many domain instances have a value of medium risk for the risk factor attribute?
- 10
 - 20
 - 30
 - 40

(4.5)

11. A certain dataset contains two classes—*class A* and *class B*—each having 100 instances. RuleMaker generates several rules for each class. One rule for *class A* is given as

```
att1 = value1
# covered = 20
# remaining =60
```

What percent of the class A instances are covered by this rule?

- a. 20
- b. 40
- c. 60
- d. 70
- e. 80

(4.7)

12. The single best representative of a class.
- a. mean
 - b. centroid
 - c. signature
 - d. prototype

Answers to Chapter 4 Questions

Multiple Choice Questions

- 1. b
- 2. a
- 3. d
- 4. b
- 5. d
- 6. b
- 7. a
- 8. d
- 9. c
- 10. b
- 11. a
- 12. d

Chapter 5

Multiple Choice Questions

(5.1)

1. KDD has been described as the application of ____ to data mining.
 - a. the waterfall model
 - b. object-oriented programming
 - c. the scientific method
 - d. procedural intuition

(5.2)

2. The choice of a data mining tool is made at this step of the KDD process.
 - a. goal identification
 - b. creating a target dataset
 - c. data preprocessing
 - d. data mining
3. Attributes may be eliminated from the target dataset during this step of the KDD process.
 - a. creating a target dataset
 - b. data preprocessing
 - c. data transformation
 - d. data mining
4. This step of the KDD process model deals with noisy data.
 - a. Creating a target dataset
 - b. data preprocessing
 - c. data transformation
 - d. data mining

(5.3)

5. The relational database model is designed to
- promote data redundancy.
 - minimize data redundancy.
 - eliminate the need for data transformations.
 - eliminate the need for data preprocessing.

(5.4)

6. A common method used by some data mining techniques to deal with missing data items during the learning process.
- replace missing real-valued data items with class means
 - discard records with missing data
 - replace missing attribute values with the values found within other similar instances
 - ignore missing attribute values

(5.5)

7. This data transformation technique works well when minimum and maximum values for a real-valued attribute are known.
- min-max normalization
 - decimal scaling
 - z-score normalization
 - logarithmic normalization
8. This technique uses mean and standard deviation scores to transform real-valued attributes.
- decimal scaling
 - min-max normalization
 - z-score normalization
 - logarithmic normalization

9. A data normalization technique for real-valued attributes that divides each numerical value by the same power of 10.
- a. min-max normalization
 - b. z-score normalization
 - c. decimal scaling
 - d. decimal smoothing
10. The price of a 12 ounce box of cereal decreases from \$3.50 to \$3.00. What fraction is used to compute the percent decrease in the price of the cereal?
- a. $\frac{1}{3}$
 - b. $\frac{1}{5}$
 - c. $\frac{1}{6}$
 - d. $\frac{1}{7}$

Answers to Chapter 5 Questions

Multiple Choice Questions

- 1. c
- 2. a
- 3. c
- 4. b
- 5. b
- 6. d
- 7. a
- 8. c
- 9. c
- 10. d

Chapter 6

Multiple Choice Questions

(6.1)

1. Operational databases are designed to support _____ whereas decision support systems are designed to support _____.
 - a. transactional processing, data analysis
 - b. data analysis, transactional processing
 - c. independent data marts, dependent data marts
 - d. dependent data marts, independent data marts
2. A one-to-many relationship.
 - a. brother-sister
 - b. father-son
 - c. uncle-nephew
 - d. teacher-student
3. The purpose of an intersection entity is to replace
 - a. two one-to-one relationships with a one-to-many relationship
 - b. two one-to-many relationships with one many-to-many relationship
 - c. a many-to-many relationship with two one-to-many relationships
 - d. a one-to-many relationship with two one-to-one relationships
4. This process removes redundancies that may be present in a data model.
 - a. abstraction
 - b. granularization
 - c. standardization
 - d. normalization

(6.2)

5. Which of the following is not a characteristic of a data warehouse?
 - a. contains nonvolatile data
 - b. is subject oriented
 - c. supports data processing, collection and management
 - d. stores data to be reported on, analyzed and tested

6. The level of detail of the information stored in a data warehouse.
 - a. granularity
 - b. scope
 - c. functionality
 - d. level of query

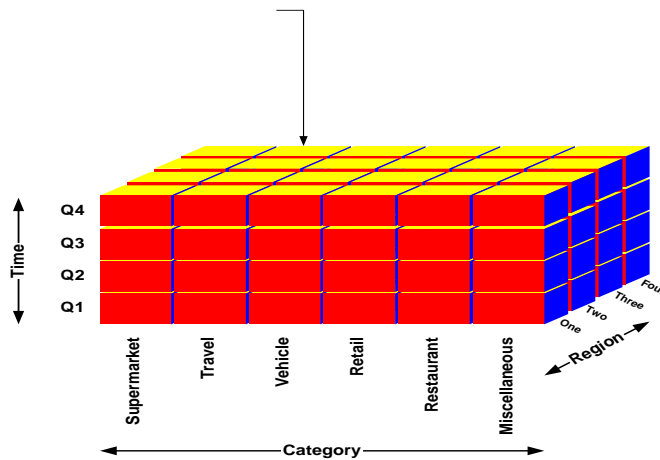
7. Defines each dimension of the multidimensional space of a star schema.
 - a. ELT table
 - b. dimension table
 - c. fact table
 - d. decision table

8. A variation of the star schema that allows more than one central fact table.
 - a. snowflake schema
 - b. linked star schema
 - c. distributed star schema
 - d. constellation schema

9. Which of the following is *not* a primary function of a decision support system?
 - a. knowledge discovery
 - b. reporting data
 - c. processing data
 - d. analyzing data

(6.3)

Use the OLAP cube to answer questions 10 through 12. Each cell of the cube contains a dollar amount representing total dollars spent for the specified time, region and category.



10. Consider the OLAP cube shown above. The vertical arrow points to:
 - a. region four, travel
 - b. region two, travel
 - c. Q2, travel
 - d. Q1, travel
11. Which of the following is *not* an example of a *slice* operation?
 - a. Select all cells where purchase category = retail.
 - b. Select all cells where purchase category = retail or vehicle.
 - c. Provide a spreadsheet of quarter and region information for all cells pertaining to restaurant.
 - d. Identify the region of peak travel expenditure for each quarter.
12. Which of the following is an example of a *dice* operation?
 - a. Which region shows the smallest amount of total dollars spent on restaurant and travel for all quarters?
 - b. Select all cells where time=Q1 or Q2.
 - c. Provide a spreadsheet of category and region information for Q1.
 - d. Select all cells where category = travel, vehicle or retail.

13. A mapping that allows the attributes of an OLAP cube to be viewed from varying levels of detail.
- a. reflection
 - b. concept hierarchy
 - c. rotation
 - d. semantic network

Answers to Chapter 6 Questions

Multiple Choice Questions

- 1. a
- 2. b
- 3. c
- 4. d
- 5. c
- 6. a
- 7. c
- 8. d
- 9. c
- 10. a
- 11. d
- 12. a
- 13. b

Chapter 7

Multiple Choice Questions

(7.1)

1. Selecting data so as to assure that each class is properly represented in both the training and test set.
 - a. cross validation
 - b. stratification
 - c. verification
 - d. bootstrapping

(7.2)

2. The standard error is defined as the square root of this computation.
 - a. The sample variance divided by the total number of sample instances.
 - b. The population variance divided by the total number of sample instances.
 - c. The sample variance divided by the sample mean.
 - d. The population variance divided by the sample mean.
3. If a real-valued attribute is normally distributed, we know that approximately 95% of all attribute values lie within
 - a. one standard deviation of the mean.
 - b. two standard deviations of the mean.
 - c. three standard deviations of the mean.
 - d. four standard deviations of the mean.
4. The hypothesis of no significant difference.
 - a. nil
 - b. invalid
 - c. null
 - d. void

5. A decision tree is built to determine individuals likely to default on an unsecured loan. The null hypothesis states that an individual will not default on the loan. The decision tree correctly classifies 80% of the instances in a test dataset. Fifteen percent of the mistakes made by the model are type 1 errors. What can be said about the performance of the model?
- The accuracy of the model for correctly determining those individuals who did not default on their loan was at least 75%.
 - The accuracy of the model for correctly determining those individuals who defaulted on their loan was at least 75%.
 - The majority of errors made by the model accepted individuals who defaulted.
 - The majority of errors made by the model rejected individuals who did not default.
 - More than one of a,b,c or d is correct.

(7.3)

6. Data used to optimize the parameter settings of a supervised learner model.
- training
 - test
 - verification
 - validation
7. We have performed a supervised classification on a dataset containing 100 test set instances. Eighty of the test set instances were correctly classified. The 95% test set accuracy confidence boundaries are:
- 76% and 84%
 - 72% and 88%
 - 78% and 82%
 - 70% and 90%
8. Bootstrapping allows us to
- choose the same training instance several times.
 - choose the same test set instance several times.
 - build models with alternative subsets of the training data several times.
 - test a model with alternative subsets of the test data several times.

(7.4)

9. We have built and tested two supervised learner models— M_1 and M_2 . We compare the test set accuracy of the models using the classical hypothesis testing paradigm using a 95% confidence setting. The computed value of P is 2.53. What can we say about this result?
- Model M_1 performs significantly better than M_2 .
 - Model M_2 performs significantly better than M_1 .
 - Both models perform at the same level of accuracy.
 - The models differ significantly in their performance.
 - More than one of a,b,c or d is correct.

(7.5)

10. The correlation between the number of years an employee has worked for a company and the salary of the employee is 0.75. What can be said about employee salary and years worked?
- There is no relationship between salary and years worked.
 - Individuals that have worked for the company the longest have higher salaries.
 - Individuals that have worked for the company the longest have lower salaries.
 - The majority of employees have been with the company a long time.
 - The majority of employees have been with the company a short period of time.
11. The correlation coefficient for two real-valued attributes is -0.85 . What does this value tell you?
- The attributes are not linearly related.
 - As the value of one attribute increases the value of the second attribute also increases.
 - As the value of one attribute decreases the value of the second attribute increases.
 - The attributes show a curvilinear relationship.

(7.6)

12. Unsupervised evaluation can be internal or external. Which of the following is an internal method for evaluating alternative clusterings produced by the K-Means algorithm?
- a. Use a production rule generator to compare the rule sets generated for each clustering.
 - b. Compute and compare class resemblance scores for the clusters formed by each clustering.
 - c. Compare the sum of squared error differences between instances and their corresponding cluster centers for each alternative clustering.
 - d. Create and compare the decision trees determined by each alternative clustering.

(7.7)

13. The average squared difference between classifier predicted output and actual output.
- a. mean squared error
 - b. root mean squared error
 - c. mean absolute error
 - d. mean relative error

Answers to Chapter 7 Questions

Multiple Choice Questions

- 1. b
- 2. a
- 3. b
- 4. c
- 5. d
- 6. d
- 7. b
- 8. a
- 9. d
- 10. b
- 11. c
- 12. c
- 13. a

Chapter 8

Multiple Choice Questions

(8.1)

1. A feed-forward neural network is said to be *fully connected* when
 - a. all nodes are connected to each other.
 - b. all nodes at the same layer are connected to each other.
 - c. all nodes at one layer are connected to all nodes in the next higher layer.
 - d. all hidden layer nodes are connected to all output layer nodes.
2. The values input into a feed-forward neural network
 - a. may be categorical or numeric.
 - b. must be either all categorical or all numeric but not both.
 - c. must be numeric.
 - d. must be categorical.

(8.2)

3. Neural network training is accomplished by repeatedly passing the training data through the network while
 - a. individual network weights are modified.
 - b. training instance attribute values are modified.
 - c. the ordering of the training instances is modified.
 - d. individual network nodes have the coefficients on their corresponding functional parameters modified.

4. Genetic learning can be used to train a feed-forward network. This is accomplished by having each population element represent one possible
 - a. network configuration of nodes and links.
 - b. set of training data to be fed through the network.
 - c. set of network output values.
 - d. set of network connection weights.

5. With a Kohonen network, the output layer node that wins an input instance is rewarded by having
 - a. a higher probability of winning the next training instance to be presented.
 - b. its connect weights modified to more closely match those of the input instance.
 - c. its connection weights modified to more closey match those of its neighbors.
 - d. neighoring connection weights modified to become less similar to its own connection weights.

6. A two-layered neural network used for unsupervised clustering.
 - a. backpropagation network
 - b. Kohonen network
 - c. perceptron network
 - d. agglomerative network

(8.3)

7. This neural network explanation technique is used to determine the relative importance of individual input attributes.
 - a. sensitivity analysis
 - b. average member technique
 - c. mean squared error analysis
 - d. absolute average technique

(8.4)

8. Which one of the following is not a major strength of the neural network approach?
- a. Neural networks work well with datasets containing noisy data.
 - b. Neural networks can be used for both supervised learning and unsupervised clustering.
 - c. Neural network learning algorithms are guaranteed to converge to an optimal solution.
 - d. Neural networks can be used for applications that require a time element to be included in the data.

(8.5)

9. During backpropagation training, the purpose of the delta rule is to make weight adjustments so as to
- a. minimize the number of times the training data must pass through the network.
 - b. minimize the number of times the test data must pass through the network.
 - c. minimize the sum of absolute differences between computed and actual outputs.
 - d. minimize the sum of squared error differences between computed and actual output.
10. Epochs represent the total number of
- a. input layer nodes.
 - b. passes of the training data through the network.
 - c. network nodes.
 - d. passes of the test data through the network.

Answers to Chapter 8 Questions

Multiple Choice Questions

1. c
2. c
3. a
4. d
5. b
6. b
7. a
8. c
9. d
10. b

Chapter 9

Multiple Choice Questions

(9.1)

1. Two classes each of which is represented by the same pair of numeric attributes are linearly separable if
 - a. at least one of the pairs of attributes shows a curvilinear relationship between the classes.
 - b. at least one of the pairs of attributes shows a high positive correlation between the classes.
 - c. at least one of the pairs of attributes shows a high positive correlation between the classes.
 - d. a straight line partitions the instances of the two classes.
2. The test set accuracy of a backpropagation neural network can often be improved by
 - a. increasing the number of epochs used to train the network.
 - b. decreasing the number of hidden layer nodes.
 - c. increasing the learning rate.
 - d. decreasing the number of hidden layers.
3. This type of supervised network architecture does not contain a hidden layer.
 - a. backpropagation
 - b. perceptron
 - c. self-organizing map
 - d. genetic

(9.2)

4. The total delta measures the total absolute change in network connection weights for each pass of the training data through a neural network. This value is most often used to determine the convergence of a
 - a. perceptron network.
 - b. feed-forward network.
 - c. backpropagation network.
 - d. self-organizing network.

Answers to Chapter 9 Questions

Multiple Choice Questions

1. d
2. a
3. b
4. c

Chapter 10

Multiple Choice Questions

(10.1)

1. Simple regression assumes a _____ relationship between the input attribute and output attribute.
 - a. linear
 - b. quadratic
 - c. reciprocal
 - d. inverse

2. Regression trees are often used to model _____ data.
 - a. linear
 - b. nonlinear
 - c. categorical
 - d. symmetrical

3. The leaf nodes of a model tree are
 - a. averages of numeric output attribute values.
 - b. nonlinear regression equations.
 - c. linear regression equations.
 - d. sums of numeric output attribute values.

(10.2)

4. Logistic regression is a _____ regression technique that is used to model data having a _____ outcome.
- a. linear, numeric
 - b. linear, binary
 - c. nonlinear, numeric
 - d. nonlinear, binary
5. This technique associates a conditional probability value with each data instance.
- a. linear regression
 - b. logistic regression
 - c. simple regression
 - d. multiple linear regression

(10.3)

6. The probability of a hypothesis before the presentation of evidence.
- a. a priori
 - b. subjective
 - c. posterior
 - d. conditional
7. This supervised learning technique can process both numeric and categorical input attributes.
- a. linear regression
 - b. Bayes classifier
 - c. logistic regression
 - d. backpropagation learning
8. With Bayes classifier, missing data items are
- a. treated as equal compares.
 - b. treated as unequal compares.
 - c. replaced with a default value.
 - d. ignored.

9. The table below contains counts and ratios for a set of data instances to be used for supervised Bayesian learning. The output attribute is sex with possible values *male* and *female*. Consider an individual who has said *no* to the life insurance promotion, *yes* to the magazine promotion, *yes* to the watch promotion and has credit card insurance. Use the values in the table together with Bayes classifier to determine which of a,b,c or d represents the probability that this individual is *male*.

	Magazine Promotion		Watch Promotion		Life Insurance Promotion		Credit Card Insurance	
	<i>male</i>	<i>female</i>	<i>male</i>	<i>female</i>	<i>male</i>	<i>female</i>	<i>male</i>	<i>female</i>
<i>Yes</i>	4	3	2	2	2	3	2	1
<i>No</i>	2	1	4	2	4	1	4	3
<i>Yes</i>	4/6	3 /4	2/6	2/4	2/6	3/4	2/6	1/4
<i>No</i>	2/6	1 /4	4/6	2/4	4/6	1/4	4/6	3/4

- $(4/6) (2/6) (2/6) (2/6) (6/10) / P(E)$
- $(4/6) (2/6) (3/4) (2/6) (3/4) / P(E)$
- $(4/6) (4/6) (2/6) (2/6) (6/10) / P(E)$
- $(2/6) (4/6) (4/6) (2/6) (4/10) / P(E)$

(10.4)

10. This clustering algorithm merges and splits nodes to help modify nonoptimal partitions.
- agglomerative clustering
 - expectation maximization
 - conceptual clustering
 - K-Means clustering
11. This clustering algorithm initially assumes that each data instance represents a single cluster.
- agglomerative clustering
 - conceptual clustering
 - K-Means clustering
 - expectation maximization

12. This unsupervised clustering algorithm terminates when mean values computed for the current iteration of the algorithm are identical to the computed mean values for the previous iteration.
- agglomerative clustering
 - conceptual clustering
 - K-Means clustering
 - expectation maximization

(10.5)

13. Machine learning techniques differ from statistical techniques in that machine learning methods
- typically assume an underlying distribution for the data.
 - are better able to deal with missing and noisy data.
 - are not able to explain their behavior.
 - have trouble with large-sized datasets.

Answers to Chapter 10 Questions

Multiple Choice Questions

- a
- b
- c
- d
- b
- a
- b
- d
- c
- c
- a
- c
- b

Chapter 11

Multiple Choice Questions

(11.1)

1. Which of the following problems is best solved using time-series analysis?
 - a. Predict whether someone is a likely candidate for having a stroke.
 - b. Determine if an individual should be given an unsecured loan.
 - c. Develop a profile of a star athlete.
 - d. Determine the likelihood that someone will terminate their cell phone contract.

(11.2)

2. At least eighty percent of the time spent on a Web-based data mining project is devoted to this.
 - a. goal identification
 - b. data preparation
 - c. data mining
 - d. interpretation of results
3. A set of pageviews requested by a single user from a Web server.
 - a. index page
 - b. common log
 - c. session
 - d. page frame
4. A data file that contains session information.
 - a. cookie
 - b. pageview
 - c. page frame
 - d. common log

5. Usage profiles for Web-based personalization contain several
 - a. pageviews
 - b. clickstreams
 - c. cookies
 - d. session files
6. The automation of Web site adaptation involves creating and deleting
 - a. index pages
 - b. cookies
 - c. pageviews
 - d. clickstreams
7. A data mining algorithm designed to discover frequently accessed Web pages that occur in the same order.
 - a. serial miner
 - b. association rule miner
 - c. sequence miner
 - d. decision miner

(11.3)

8. The training phase of a textual data mining process involves
 - a. removing common words from a dictionary.
 - b. creating an attribute dictionary.
 - c. determining whether a document is about the topic under investigation.
 - d. modifying an initially created attribute dictionary.

(11.4)

9. These can be used to help select a best subset of training data.
 - a. domain resemblance scores
 - b. class resemblance scores
 - c. instance typicality scores
 - d. standard deviation scores

10. Which of the following is a fundamental difference between bagging and boosting?
 - a. Bagging is used for supervised learning. Boosting is used with unsupervised clustering.
 - b. Bagging gives varying weights to training instances. Boosting gives equal weight to all training instances.
 - c. Bagging does not take the performance of previously built models into account when building a new model. With boosting each new model is built based upon the results of previous models.
 - d. Boosting is used for supervised learning. Bagging is used with unsupervised clustering.

Answers to Chapter 11 Questions

Multiple Choice Questions

1. d
2. b
3. c
4. d
5. a
6. a
7. c
8. b
9. c
10. c

11.

Chapter 12

Multiple Choice Questions

(12.1)

1. A problem that cannot be solved with a computer using a traditional algorithmic technique.
 - a. exponentially hard problem
 - b. recursive problem
 - c. non-transformable problem
 - d. combinatorial problem
2. Any technique that helps limit the size of a search space.
 - a. top-down technique
 - b. conflict resolution strategy
 - c. bottom-up technique
 - d. heuristic technique

12.2)

3. Problems posed for solution by an artificial intelligence technique are often represented using
 - a. a truth space.
 - b. a rule space.
 - c. a state space.
 - d. a knowledge space.
4. This reasoning strategy works best for problems where the goal can be stated as a question.
 - a. forward chaining
 - b. depth-first search
 - c. backward chaining
 - d. breadth-first search

(12.3)

5. The main components of an expert system include a Knowledge base, a user interface and
 - a. a database
 - b. an explanation facility
 - c. an inference engine
 - d. a constraint propagator
6. Encoding the knowledge extracted from an expert into a knowledge base.
 - a. knowledge programming
 - b. knowledge representation
 - c. knowledge acquisition
 - d. rapid prototyping
7. Which of the following problems is not a good candidate for an expert systems solution?
 - a. bankruptcy prediction
 - b. determining company advertising strategies
 - c. solving the national debt crisis
 - d. providing investment strategies to avoid paying taxes
8. An expert system with an empty knowledge base.
 - a. expert system shell
 - b. inference shell
 - c. state space
 - d. goal tree
9. Knowledge about knowledge is known as
 - a. metaknowledge
 - b. class knowledge
 - c. structured knowledge
 - d. classified knowledge

10. Knowledge acquisition involves one or more experts and a
 - a. systems analyst.
 - b. knowledge engineer.
 - c. knowledge designer.
 - d. systems engineer.
11. Developing a rough version of a system that is suitable for testing.
 - a. validating
 - b. field reporting
 - c. verifying
 - d. prototyping
12. An internal test of an expert system whose purpose is to determine if the system uses the same reasoning process as the expert(s) used to build the system.
 - a. validation
 - b. verification
 - c. reliability
 - d. suitability

(12.4)

13. The knowledge contained in a knowledge-based system
 - a. contains the knowledge of one or more human experts.
 - b. contains the knowledge of one or more intelligent agents.
 - c. contains the knowledge of exactly one human expert.
 - d. contains knowledge extracted from one or several sources.

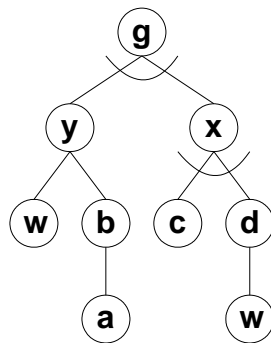
Computational Questions

- Construct a goal tree using the following production rules. Assume the goal is g .

<u>Rule 1</u>	<u>Rule 2</u>	<u>Rule 3</u>
IF x & y	IF z	IF c
THEN z	THEN g	THEN z

<u>Rule 4</u>	<u>Rule 5</u>	<u>Rule 6</u>
IF a & b	IF w	IF d
THEN w	THEN g	THEN c

- Write production rules for the following goal tree.



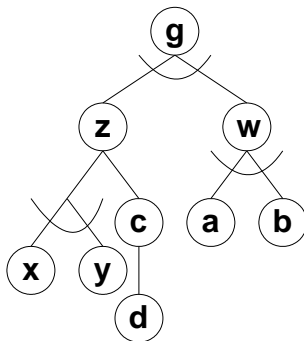
Answers to Chapter 12 Questions

Multiple Choice Questions

1. a
2. d
3. c
4. c
5. c
6. a
7. c
8. a
9. a
10. b
11. d
12. b
13. d

Computational Questions

1. Here is the goal tree



2. Here are the rules.

<u>Rule 1</u>	<u>Rule 2</u>	<u>Rule 3</u>	<u>Rule 4</u>	<u>Rule 5</u>	<u>Rule 6</u>
IF y & x	IF w	IF b	IF a	IF c & d	IF w
THEN g	THEN y	THEN y	THEN b	THEN x	THEN d

Chapter 13

Multiple Choice Questions

(13.1)

1. Computing the probability of picking a heart from a deck of 52 cards can be determined using _____ probability technique.
 - a. an objective
 - b. an experimental
 - c. a subjective
 - d. an inexact
2. Probabilities based on frequency distributions obtained through sampling.
 - a. objective
 - b. experimental
 - c. subjective
 - d. inexact
3. A car mechanic tells you that there is a 75% chance that your car will need major repair work within the next six months. This statement is an example of
 - a. an objective probability.
 - b. an experimental probability.
 - c. a subjective probability.
 - d. a fuzzy probability.

(13.2)

4. A fuzzy set is associated with a
 - a. linguistic variable.
 - b. certainty factor.
 - c. hypothesis to be tested.
 - d. linguistic value.

5. This process converts a crisp numeric value to its corresponding degree of membership within a fuzzy set.
 - a. composition
 - b. inference
 - c. fuzzification
 - d. defuzzification

6. This technique is used to determine the height of a rule consequent membership function as determined by the truth of the rule's antecedent condition.
 - a. fuzzy set union
 - b. fuzzy set intersection
 - c. center of gravity
 - d. clipping

(13.3)

7. Given evidence E and hypothesis H . The likelihood of sufficiency is computed as
 - a. the conditional probability of E being true given H is true divided by the conditional probability of E being true given H is false.
 - b. the conditional probability of E being true given H is true divided by the conditional probability of E being false given H is false.
 - c. the conditional probability of E being false given H is true divided by the conditional probability of E being true given H is false.
 - d. the conditional probability of E being false given H is true divided by the conditional probability of E being false given H is false.

8. With Bayes theorem the probability of hypothesis H — specified by $P(H)$ — is referred to as
 - a. an a priori probability
 - b. a conditional probability
 - c. a posterior probability
 - d. a bidirectional probability

9. For Bayes theorem to be applied, the following relationship between hypothesis H and evidence E must hold.
- $P(H/E) + P(H/\sim E) = 1$
 - $P(H/E) + P(\sim H/E) = 1$
 - $P(H/E) + P(H/\sim E) = 0$
 - $P(H/E) + P(\sim H/E) = 0$

Computational Questions

1. The probability that a person owns a sports car given that they subscribe to at least one automotive magazine is 40%. We also know that 3% of the adult population subscribes to at least one automotive magazine. Finally, the probability of a person owning a sports car given that they don't subscribe to at least one automotive magazine is 30%. Use this information together with Bayes theorem to compute the probability that a person subscribes to at least one automotive magazine given that they own a sports car.

Answers to Chapter 13 Questions

Multiple Choice Questions

- a
- b
- c
- d
- c
- d
- a
- a
- b

Computational Questions

1. Here is the computation.

$$\frac{(0.40)(0.03)}{(0.40)(0.03) + (0.30)(0.97)} \approx 0.03960$$

Chapter 14

Multiple Choice Questions

(14.1)

1. Autonomy is an agent's ability to
 - a. react to a changing environment.
 - b. act without direct intervention from others.
 - c. confer with other agents.
 - d. react to sensory information received from the environment.

(14.2)

2. An agent's ability to choose its actions in the context of other agents.
 - a. autonomy
 - b. cooperation
 - c. adaptivity
 - d. coordination
3. This type of agent resides inside a data warehouse in an attempt to discover changes in business trends.
 - a. semiautonomous agent
 - b. cooperative agent
 - c. data mining agent
 - d. filtering agent

(14.3)

4. An expert system contains _____ knowledge whereas the knowledge processed by an intelligent agent is _____
- a. personal, general
 - b. general, personal
 - c. direct, indirect
 - d. indirect, direct
5. Which of the following best differentiates between a data mining approach to problem-solving and an expert systems approach?
- a. The output of an expert system is a set of rules and the output of a data mining technique is a decision tree.
 - b. A data mining technique builds a model without the aid of a human expert whereas an expert system is built from the knowledge provided by one or more human experts.
 - c. A model built using a data mining technique can explain how decisions are made but an expert system cannot.
 - d. An expert system is built using inductive learning whereas a data mining model is built using one or several deductive techniques.
6. Data mining techniques and intelligent agents are similar in that they
- a. are both used for hypotheses testing.
 - b. both have the ability to learn.
 - c. are goal directed.
 - d. both build models from data.

Answers to Chapter 14 Questions

Multiple Choice Questions

- 1. b
- 2. d
- 3. c
- 4. b
- 5. b
- 6. b

