Part 1: Multiple-Choice Questions (17 questions, 1 point each)

1. Select the TRUE statement concerning normalization.
   a. Performance is increased.  
   b. Data consistency is improved.  
   c. Redundancy is increased.  
   d. The number of tables is reduced.  
   e. Functional dependencies increase.

2. A lack of normalization can lead to which one of the following problems:
   a. Deadlock  
   b. Lost Updates  
   c. Insertion problems  
   d. Deferred updates  
   e. Deletion of data

3. Each of the following is an argument which might be used to support the use of relations which are not fully normalized. Select the weakest argument.
   a. A fully normalized database may have too many tables  
   b. Full normalization may compromise existing applications/systems  
   c. Full normalization may make some queries too complicated  
   d. A fully normalized database may result in tables which are too large  
   e. A fully normalized database may perform too slowly

4. If a non-key attribute of a table can be null, that table automatically violates which normal form (choose the lowest one):
   NONE  1NF  2NF  3NF  BCNF  4NF  5NF

5. If an attribute of a table can have multiple values, that table automatically violates which normal form (choose the lowest one):
   NONE  1NF  2NF  3NF  BCNF  4NF  5NF

6. Given the following relation and dependences, state which normal form the relation is in.
   \[ R(p,q,r,s,t) \]
   \[ p,q \rightarrow r,s,t \]
   \[ r,s \rightarrow p,q,t \]
   \[ t \rightarrow s \]
   a. 1NF  
   b. 2NF  
   c. 3NF  
   d. BCNF  
   e. Unnormalized
7. Which of the following is the highest normal form by which the \( R \) relation can be classified?

\[ R(\text{patient, consultant, hospital, address, date, time}) \]

Given

\[ \text{patient, consultant} \rightarrow \text{hospital, address, date, time} \]
\[ \text{hospital} \rightarrow \text{address} \]

a. 1NF   b. Unnormalized   c. 2NF   d. BCNF   e. 3NF

8. Assume the relation \( R(A, B, C, D, E) \) is in at least 3NF. Which of the following functional dependencies must be FALSE?

a. \( A, B \rightarrow C \)   b. \( A, C \rightarrow E \)   c. \( A, B \rightarrow D \)   d. \( C, D \rightarrow E \)   e. None of the above

9. Consider the relational schema \( R(A, B, C, D, E) \) with non-key functional dependencies \( C, D \rightarrow E \) and \( B \rightarrow C \).

Select the strongest statement that can be made about the schema \( R \)

a. \( R \) is in third normal form   b. \( R \) is in second normal form   c. \( R \) is in BCNF normal form   d. \( R \) is in first normal form   e. None of the above

10. Consider the following functional dependencies:

\[ a, b \rightarrow c, d \] \[ e \rightarrow c \] \[ b \rightarrow e, f \]

Given the same functional dependencies as shown above, which option shows the relations normalized to 3NF of: \( R(a, b, c, d, e, f) \)

a. \( R(a, b, c, d, e, f) \) \( R(e, c) \) \( R(b, e, f) \)  
b. \( R(a, b, c, d) \) \( R(c, e) \) \( R(e, f, b) \)  
c. \( R(a, b, c, d) \) \( R(c, e) \) \( R(b, e, f) \)  
d. \( R(a, b, d) \) \( R(e, c) \) \( R(b, e, f) \)  
e. \( R(a, b, c, d, e, f) \)
11. Given the following relation and dependencies, select the option that is the result of fully normalizing the relation to BCNF.

\[ R(\overline{a},b,c,d,e) \]
\[ a \rightarrow c \quad d \rightarrow c,e \]

a. \( R1(\overline{a},c) \quad R2(d,e) \quad R(\overline{a},b,d) \)
b. \( R1(\overline{a},c) \quad R2(d,c,e) \quad R(\overline{a},b,d) \)
c. \( R1(d,c,e) \quad R(\overline{a},b,d) \)
d. \( R(\overline{a},b,c,d,e) \)
e. None of the above

12. If a relation schema contains two different multi-valued dependencies, that table automatically violates which normal form (choose the lowest one):

<table>
<thead>
<tr>
<th>NONE</th>
<th>1NF</th>
<th>2NF</th>
<th>3NF</th>
<th>BCNF</th>
<th>4NF</th>
<th>5NF</th>
</tr>
</thead>
</table>

13. In spatial databases, what does MBR refer to?

a. Minimum Box Rectangle  
b. Minimum Box Region  
c. Minimum Bounding Region  
d. Minimum Bounding Rectangle  
e. Minimum Bounding Box

14. The two steps (in order) that are performed when answering a spatial query using MBR are

a. filtering and refinement steps  
b. refinement and filtering steps  
c. filtering and screening steps  
d. screening and filtering steps

15. The first step(s) in the knowledge discovery process is/are

a. data selection  
b. visualization  
c. data mining  
d. data transformation  
e. data cleaning and integration

16. The measure of interestingness used by association rule mining that is based on how frequently an itemset appears in data is called

a. correlation  
b. confidence  
c. support  
d. importance  
e. none of the above

17. Association rule mining is an example of

a. Descriptive data mining  
b. Unsupervised learning  
c. Predictive data mining  
d. Learning by example  
e. NOA
Part 2: Essay Questions (83 points all)

1. Consider a relation Book which collects information about book identification numbers ("ISBN numbers"), authors, titles, publishers, and years of publication. Hence the relation is:

   Book = (ISBN, Author, Title, Publisher, Year)

   For convenience, we will henceforth abbreviate these attributes as:

   Book = (I, A, T, P, Y)

   Suppose the following set F of functional dependencies are asserted to hold:

   \[
   \begin{align*}
   TP & \rightarrow I \\
   AP & \rightarrow T \\
   I & \rightarrow ATP
   \end{align*}
   \]

   a. Is the following decomposition of Book lossless? Why or why not? 4

      R1(A, T, P), R2(I, P, Y), R3(I, T)

   b. Does the decomposition in a preserve all functional dependencies in F'\(^+\)? If so, simply state Yes. If not, give a functional dependency that is not preserved. 4

   c. Consider replacing F by an alternative set of functional dependencies G:

      AP \rightarrow IT \\
      I \rightarrow ATP

      Is the effect of G the same as the effect of F? In other words, does G'\(^+\) = F'\(^+\)? Why or why not?
2.

Suppose we have a database for an investment firm, consisting of the following attributes:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Broker</td>
</tr>
<tr>
<td>O</td>
<td>Office of a broker</td>
</tr>
<tr>
<td>I</td>
<td>Investor</td>
</tr>
<tr>
<td>S</td>
<td>Stock</td>
</tr>
<tr>
<td>Q</td>
<td>Quantity of stock owned by an investor</td>
</tr>
<tr>
<td>D</td>
<td>Dividend paid by a stock</td>
</tr>
</tbody>
</table>

Hence the overall schema is \( R(B, O, I, S, Q, D) \). Assume the following functional dependencies are required to hold on this database:

- \( I \rightarrow B \)
- \( IS \rightarrow Q \)
- \( B \rightarrow O \)
- \( S \rightarrow D \)

a. Consider the following database instance \( D_1 \) of \( R \):

<table>
<thead>
<tr>
<th>B</th>
<th>O</th>
<th>I</th>
<th>S</th>
<th>Q</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merrill</td>
<td>SLC</td>
<td>Greene</td>
<td>IBM</td>
<td>100</td>
<td>$1.50</td>
</tr>
<tr>
<td>Schwab</td>
<td>Provo</td>
<td>Hatch</td>
<td>Unisys</td>
<td>200</td>
<td>$0.70</td>
</tr>
<tr>
<td>Edwards</td>
<td>Loa</td>
<td>Orton</td>
<td>Novell</td>
<td>300</td>
<td>$0.05</td>
</tr>
<tr>
<td>Carl</td>
<td>Bicknell</td>
<td>Hansen</td>
<td>Borland</td>
<td>400</td>
<td>$2.00</td>
</tr>
<tr>
<td>Schwab</td>
<td>Provo</td>
<td>Hatch</td>
<td>Novell</td>
<td>500</td>
<td>$0.10</td>
</tr>
</tbody>
</table>

Is \( D_1 \) consistent with the dependencies specified above? Why or why not?

b. Give a lossless decomposition of \( R \) into Boyce-Codd Normal Form.

c. Does your answer to Question b preserve all given and implied functional dependencies? If No, state which dependencies that are not preserved.
3. Consider the following function dependencies for a relation \( R(X, Y, Z, W) \):
   \[
   \begin{align*}
   WX & \rightarrow Y \\
   X & \rightarrow Z \\
   Z & \rightarrow WY
   \end{align*}
   \]
   Give a derivation of \( X \rightarrow Y \) from the given functional dependencies above. Justify your steps with Armstrong’s axioms.

4. Consider a relation \( R(A, B, C, D, E, F, G, H) \) with Functional Dependencies
   \[
   \begin{align*}
   AB & \rightarrow E \\
   C & \rightarrow D \\
   F & \rightarrow GH \\
   B & \rightarrow F
   \end{align*}
   \]
   (a) Consider \( R_1(A, B, C, D) \) with the above functional dependencies. What would be a candidate key for \( R_1 \)?
   (b) Is \( R_1(A, B, C, D) \) in second normal form? If not, say why.
   (c) Is \( R_1(A, B, C, D) \) and \( R_2(A, B, E) \) a lossless join decomposition of \( R'(A, B, C, D, E) \)? Explain why or why not.
   (d) Is \( R_1(A, B, C, D), R_2(A, B, E), R_3(F, G, H) \) and \( R_4(B, F) \) a dependency preserving decomposition? Explain why or why not.
   (e) Now start with these functional dependencies:
   \[
   \begin{align*}
   AB & \rightarrow E \\
   C & \rightarrow D \\
   F & \rightarrow GH \\
   FG & \rightarrow GH \\
   B & \rightarrow FG
   \end{align*}
   \]
   Find a minimal covering of these functional dependencies. Then use it to synthesize a set of dependency preserving, 3NF relations with a lossless join. I.e. use algorithm 11.4 from the text book.
5. List three spatial data types.
   a. 
   b. 
   c. 

6. What are the three spatial operators? Give an example of each operator.
   a. 
   b. 
   c. 

7. List two types of spatial queries. Give an example of each one.
   a. 
   b. 
8. Any member of R-Tree family has a set of characteristics. List three of these characteristics.
   a. 
   b. 
   c. 

9. Draw the R-tree for the following set of rectangles.
10. Consider the following set of spatial objects and the corresponding R-tee. How many pages (nodes, I/O) are read in order to search for object #5?

11. What is the main difference between classification and clustering?

12. List three factors that affect classification based on decision trees.
   a. 
   b. 
   c. 
13. List three methods used to represent the distance between clusters
   a. 
   b. 
   c. 

14. Consider the following set of transactions

<table>
<thead>
<tr>
<th>Transaction</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>Bread, Jelly, PeanutButter</td>
</tr>
<tr>
<td>$t_2$</td>
<td>Bread, PeanutButter</td>
</tr>
<tr>
<td>$t_3$</td>
<td>Bread, Milk, PeanutButter</td>
</tr>
<tr>
<td>$t_4$</td>
<td>Beer, Bread</td>
</tr>
<tr>
<td>$t_5$</td>
<td>Beer, Milk</td>
</tr>
</tbody>
</table>

What is the support of PeanutButter & Jelly?