Part A

Question 1

Supervised Learning

In supervised learning, labelled data is used to create and train the model. It means some data is already marked as correct answers. So the supervised learning can be compared to the learning which is held in the presence of the supervisor.Different Algorithms in Supervised Learning :

- Decision Tree
- Naive Baise
- K- Nearest Neighbor
- Neural Network
- Support Vector Machine (SVM)

Unsupervised Learning

In Unsupervised learning will use unlabeled data to train the model. These unsupervised learning does not need any supervision. We can permit the model to work on its own and classify data. It will sort data according to the patterns and similarities without any prior training to the data sets.
Algorithms:

- Clustering
- Principle Component Analysis (PCA)

Reinforcement Learning

Reinforcement Learning is more alike how we learn as humans. When it comes to Reinforcement Learning, it will continuously, learn from its environment while interacting with it. It rewards positive or negative based on its actions.

Reinforcement learning mostly used in self-driving cars and alpha go Chess games.

Question 2

Classification

1. Classification is supervised learning Algorithm.
2. Classification deals with labeled data.
3. Classification involves in the prediction of the input variables based on the build model.
4. Classification process involves two stages which is training stage and testing stage.

Clustering

1. Clustering is an unsupervised learning Algorithm.
2. Clustering deals with unlabeled data.
3. Data points grouped as cluster based on their similarities.
4. Clustering is generally used to analyze data and interferences from it for better decision making.
5. Clustering involves only in the process of grouping data.

Question 3

Clustering Strengths

1. Clustering models are simple to implement.
2. Can easily use for wide range of data sets.
3. Easy to make decisions based on clustering.

Clustering Weaknesses

1. It can create the model with scallings.
2. We should manually give k.

Question 5

= (input variables + hidden layers) x nodes +
(output layer + hidden layers) x nodes
= (5+1) x 12 + (12+ 1) x 3
= (6 x 12) + (13+3)
=72 + 39
= 111

111 network parameters are required to be trained.

Question 6

Association Rule Mining

For the given set of transaction data set, Association Rule mining helps to find the rules which can predict the occurrences of the given dataset based on the other occurrences of the items on the given transation.

Apriori Principle

Apriori Principle reduce the number of items in an item sets that we need to examine. In simple terms Apriori Principle says,
"If an item is infrequent, then all its superset must be infrequent."
For example, if {cake} was found to be infrequent. We can expect that {cake, Pizza} must be equal or even more infrequent. So when we consolidating the list of popular itemsets. We don't need to consider {cake, Pizza} nor any other configuration that contains {cake}.

Question 7

We can't gather personal information as data without owners' permission.

Question 8

B is not considered a support vector.
Because A, C, D , all these three are near the hyperplane. So we considered them as support vectors.

The goal of SVM is to find the best line or the best decision boundary that can segregate n-dimentional space into classes. So that we can easily put new data points in the correct category in the future.We call that most suitable decision boundary are SVM hyperplane.

Question 9

Advantages of PCA

1. Remove correlated features.
2. Improve algorithm performance.
3. Reduce data dimensionality.
4. Reduce overfitting.
5. Improve data visualization.

Disadvantages of PCA

1. Information Loss.
2. Independent Variables becomes less interpretation.

Question 10

Overall Classification Accuracy = (20 + 1 + 16) / 3 = 12.3%

Part B

Question B-1

Find Frequent 1 – size itemsets

a = 0.5
b= 0.7
c = 0.5
d = 0.9
e = 0.6

All the above itemsets are above 0.3, so it means above itemsets can procedure the next step.

Find Frequent 2-Size itemsets

ab = 0.3
ac = 0.2 rejected because it's < 0.3
ad = 0.4
ae = 0.4
bc = 0.3
bd = 0.6
be = 0.4
cd = 0.4
ce = 0.2 rejected because it's < 0.3
de = 0.6

Above 2 two-Size itemset are infrequent and all the other 8 two-size item-sets can procedure next step.

Find Frequent 3-Size itemsets

abc = rejected because ac is infrequent

abd = 0.2 rejected because it's < 0.3

abe = 0.4
acd = rejected because ac is infrequent.
ace = rejected because ce is infrequent.
ade = 0.4
bcd = 0.2 rejected because it's < 0.3
bce = rejected because ce is infrequent.
bde = 0.4

Above 3 three-Size itemsets frequent.

Question B-2

Frequency Table			
Weather class	good	bad	
Overcast	2	2	4/16 = 0.25
Rain	2	4	6/16 = 0.375
Sunny	4	2	6/16 = 0.375
Total	8	8	
	8/16 = 0.5	8/16 = 0.5	

Likelihood Table – P(weather|mood)		
Weather class	Good	Bad
Overcast		
Rain		
Sunny		