Midterm

Part 1: [20 Points] General Information

Fill in the blanks with the following words:

ANN, association rule, C4.5, classification, data mart, data mining, logistic regression, self- organizing map, weights, data warehouse.

Data Mining is the art and science of discovering useful novel patterns from data.

Logistic Regression is a supervised learning method.

Association Rule is an unsupervised learning method.

ANN can be used for both classification and regression.

C4.5 is one type of decision tree algorithm.

Self Organizing Map is one type of ANN.

In ANN, if the desired output does not match the computed output, we update the **weights** of the network.

Logistic Regression works with dependent variables with binary values and effectively becomes decision mode.

Data Warehouse is an organized collection of integrated, subject-oriented databases designed to support decision support functions.

Data Mart is a departmental data warehouse

Part 2: [20 Points] Decision Tree

Construct a decision tree:

```
Attributes    Rule         Error   Total Error
------------------------------------------------------------
Weather     Sunny  -> No    1/3    2/10
            Cloudy -> Yes   0/3
            Rainy  -> No    1/4

Tempera     Hot   -> No     2/4    4/10
            Mild  -> Yes    2/5
            Cool  -> No     0/1

Humidity    Normal -> Yes   2/4    4/10
            High   -> No    3/6

Wind        Week   -> Yes   1/4    3/10
            Strong -> No    2/6
```

At this point we have tree for Cloudy weather

```
[Weather] --- ( = Cloudy) --->  Yes
          --- ( = Sunny ) --->  ???
          --- ( = Rainy ) --->  ???
```

Calculate for Sunny weather

```
Attributes    Rule          Error   Total Error
--------------------------------------------------------------
Temera    Hot   -> No       0/2       0/3
          Mild  -> Yes      0/1

Humidity  Normal  -> Yes    1/2       1/3
          High    -> No     0/1

Wind      Week  -> No       0/1       1/3
          Strong -> Yes     1/2
```

```
[Weather] --- ( = Cloudy) --->  Yes
          --- ( = Sunny ) --->  [Temperature] --- ( = Hot   ) --->  No
                                                --- ( = Mild  ) --->  Yes
          --- ( = Rainy ) --->  ??
```

Calculate for Rainy weather

```
Attributes    Rule          Error   Total Error
--------------------------------------------------------------
Temper    Mild    -> No       1/3     1/4
          Cool    -> No       0/1

Humidity  High    -> No       1/3     1/4
          Normal  -> No       0/1

Wind      Strong  -> No       0/3     0/4
          Week    -> Yes      0/1
```

```
[Weather] --- ( = Cloudy) --->  Yes
          --- ( = Sunny ) --->  [Temperature] --- ( = Hot   ) --->  No
                                                --- ( = Mild  ) --->  Yes
          --- ( = Rainy ) --->  [Wind]        --- ( = Strong) --->  No
                                                --- ( = Week  ) --->  Yes
```

Decision:

1. No

2. Yes

3. No

Part 3: [20 Points] Regression

```
x           y         xi - X       yi - y'          (xi - x')^2
----------------------------------------------------------------------------------------------------------------------
106         115       6            12.135           3.452164
97          104       -3           1.135            51.008164
108         98        8            -4.865           14.884164
96          101       -4           -1.865           66.292164
112         98        12           -4.865           61.748164
111         106       11           3.135            47.032164
99          98        -1           -4.865           26.440164
----------------------------------------------------------------------------------------------------------------------
Σ(x) = 729    Σ(y) = 720    Σ(xi-x) = 29   Σ(yi - y') = -0.055   Σ(xi - x')^2 = 270.857148
x' = 104.142   y' = 102.856

b1 = Σ(xi-x) (yi-y')/Σ(xi-x')^2
   = (29 * -0.055) / 270.857148
```

```
    = - 0.005888712968358

 b0 = y' - b1x'
    = 102.856 - (- 0.005888712968358 X 104.142)
    = 103.469262345950739
```

Part 4: [10 Points] Data Mining

Show which of the following is a data-mining problem by choosing **True** if it is a data mining problem and **False** if it is not.

Find the relation between student's grades and the number of hours they study. **[True]**

Find how the weather will be tomorrow in a specific region. **[True]**

Find the heart rate of a person. **[False]**

Find which items the customers in a mall buy together. **[True]**

Guessing the value when tossing a fair dice. **[False]**

Find the relationship between the DNA sequence and the diseases. **[True]**

Find the output of specific digital circuit for a given input. **[False]**

Find the time the next train will come. **[False]**

Find the world population growth for the coming 5 years. **[True]**

Find the patient records using the social security number. **[False]**

Part 5: [20 Points] Artificial Neural Networks (ANN)

Write an algorithm (Or flowchart) that shows the main stages of supervised training of an ANN:

1. Gather data. Divide into training data and test data. The training data
needs to be further divided into training data and validation data.

2. Select the network architecture, such as Feed Forward Network.

3. Select the algorithm, such as Multi-Layer Perception.

4. Set network parameters.

5. Train the ANN with training data.

6. Validate the model with validation data.

7. Freeze the weights and other parameters.

8. Test the trained network with test data.

9. Deploy the ANN when it achieves good predictive accuracy.

Fine the weights w0, w1 and w2 so that the artificial neural network will emulate a NAND gate:

```
Before passing the input value to the PE we have to find the weighted sum of them.
Weighted sum of all the input values can be find using following formula

x1w1 + x2w2 + x0w0

Lets initialize the weight as follow

w0 = -2, w1 = 1, w2 = 1.

if we use these weight to calculate weighted sum for all the given input sequence we get
1) 0 * 1 + 0 * 1 + 1 * -2 = 0 + 0 - 2 = -2
   if we pass this value to our transfer function we get 1 as t < 0 is 1

2) 0 * 1 + 1 * 1 + 1 * -2 = 0 + 1 - 2 = -1
   if we pass this value to our transser function we get 1 as t < 0 is 1

3) 1 * 1 + 0 * 1 + 1 * -2 = 0 + 1 - 2 = -1
   if we pass this value to our transser function we get 1 as t < 0 is 1

4) 1 * 1 + 1 * 1 + 1 * -2 = 1 + 1 - 2 = 0
   if we pass this value to our transfer function we get 0 as t = 0 is 0


So we can conclude that to emulate a logical NAND w0 = -2, w1 = 1, w2 = 1
```