Name: _____
Score: _____ / _____

# BDA Midterm

*Instructions*

1. Allotted exam duration is 2 hours.
2. Closed book/notes.
3. You may use a plain calculator, but no cell phones.
4. No personal items including electronic devices (cell phones, computers, PDAs).
5. Cell phones must be turned in to your proctor before beginning exam.
6. No additional papers are allowed. Sufficient blank paper is included in the exam packet.
7. Exams are copyrighted and may not be copied or transferred.
8. Restroom and other personal breaks are not permitted.

## 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.

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

____ is a supervised learning method.

____ is an unsupervised learning method.

____ can be used for both classification and regression.

____ is one type of decision tree algorithm.

____ is one type of ANN.

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

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

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

____ is a departmental data warehouse

## Part 2: [20 points] Decision Tree

The following dataset will be used to train a decision tree for predicting whether there is a game or not on a given day. The factors that help to decide for the game are weather, temperature, humidity and wind.

| Day | Weather | Temperature | Humidity | Wind | Play? |
|---|---|---|---|---|---|
| 1 | Sunny | Hot | Normal | Weak | No |
| 2 | Cloudy | Hot | High | Weak | Yes |
| 3 | Sunny | Mild | Normal | Strong | Yes |
| 4 | Cloudy | Mild | High | Strong | Yes |
| 5 | Rainy | Mild | High | Strong | No |
| 6 | Rainy | Cool | Normal | Strong | No |
| 7 | Rainy | Mild | High | Weak | Yes |
| 8 | Sunny | Hot | High | Strong | No |
| 9 | Cloudy | Hot | Normal | Weak | Yes |
| 10 | Rainy | Mild | High | Strong | No |

Q2.A) [17 points] Construct a decision tree that is trained by the table above.

Q2.B)[3 points] What is the decision of each of the following cases?

| Weather | Temperature | Humidity | Wind | PLAY? |
|---|---|---|---|---|
| rainy | mild | high | strong | |
| cloudy | mild | normal | strong | |
| sunny | hot | high | strong | |

## Part 3: [20 points] Regression

$$b1=\Sigma(x_i-x)\ (y_i-y')/\Sigma(x_i-x')^2$$
$$b_0=y' - b_1x'$$

The following table shows the relationship between intelligence quotient (IQ ) and the performance in a test.

1- [**15 point**] Write down the Linear Regression model for this problem assuming we wish to predict the performance in this test for a given IQ

2- [**5 points**] What is the prediction of the performance if the IQ is 100

| IQ | Performance |
|---|---|
| 106 | 115 |
| 97 | 104 |
| 108 | 98 |
| 96 | 101 |
| 112 | 98 |
| 111 | 106 |
| 99 | 98 |

## 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
- False

Find how the weather will be tomorrow in a specific region.

- True
- False

Find the heart rate of a person.

- True
- False

Find which items the customers in a mall buy together.

- True
- False

Guessing the value when tossing a fair dice.

- True
- False

Find the relationship between the DNA sequence and the diseases.

- True
- False

Find the output of specific digital circuit for a given input.

- True
- False

Find the time the next train will come.

- True
- False

Find the world population growth for the coming 5 years.

- True
- False

Find the patient records using the social security number.

- True
- False

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

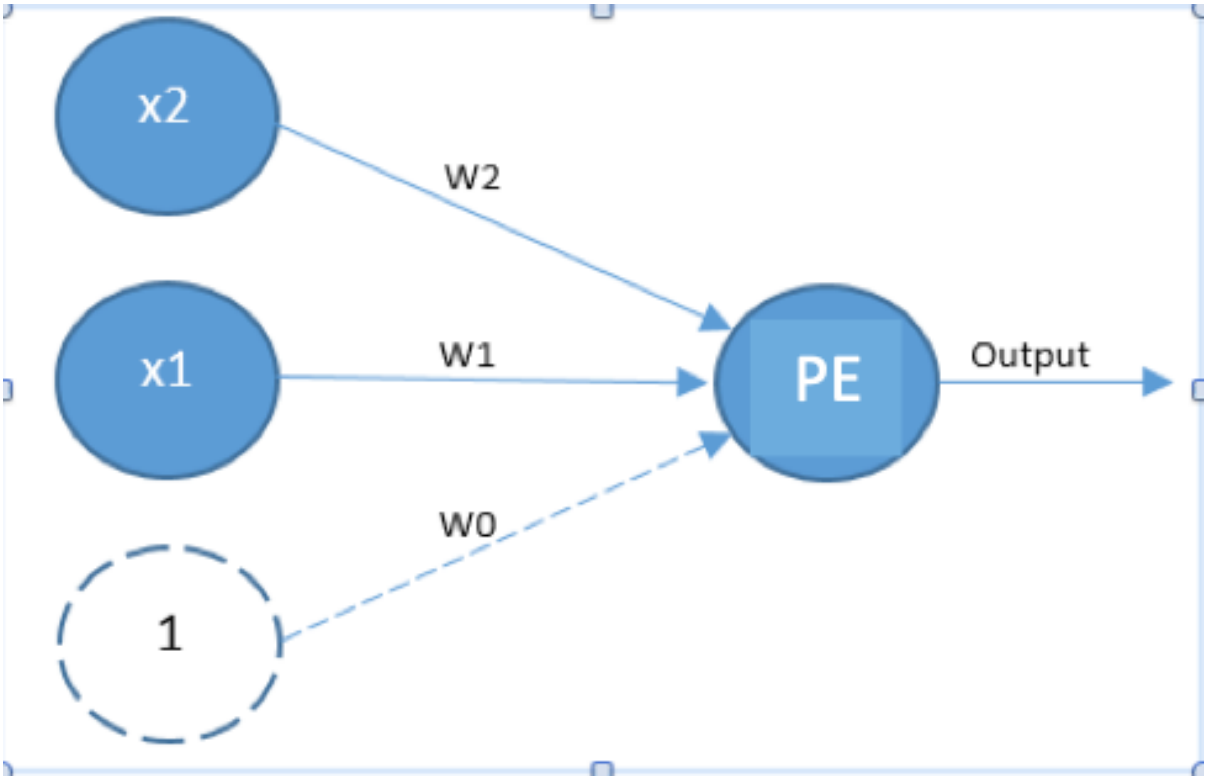
Write an algorithm (Or flowchart) that shows the main stages of supervised training of an ANN

Consider the following artificial neural network that has three inputs, with the third input always fixed to 1. In addition to the weighted sum, the PE contains a transfer function of the form:

$$f(t) = \begin{cases} 1, & t < 0 \\ 0, & t = 0 \\ 0, & t > 0 \end{cases}$$

Use this configuration to fine the weights w0, w1 and w2 so that the artificial neural network will emulate a logical NAND gate:

| X1 | X2 | Output |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |



## Part 6: [10 points] SCI Question

Write a short essay that connects one of the studied topics to the Science of Creative Intelligence (SCI). You can pick any topic.