22

Regression

Mohsen Hajsalehi Sichani and Saeed Khalafinejad
Sharif University of Technology
Iran

1. Introduction

In recent years, data mining has been widely used in various areas of science and engineering and solved many serious problems in different areas of science such as electrical power engineering, genetics, medicine and bioinformatics. Data Mining is used to extract information from data. Data mining uses AI and Statistics in its algorithms. Information refers to patterns underlying data, and data refers to recorded facts. However, the captured data need to be converted into information and knowledge to become useful. Data mining is the entire process of applying computer-based methodology, including new techniques for knowledge conversion into data. The following example is a good motivator:
Imagine you are the owner of a big supermarket and you are asking for the convenience of customers and ease of access to stuffs for customers and a high sale. In this case, if you save all of data such as time of shopping, day of shopping, sold stuffs, name of the customers and so on for about 3 to 6 months and then use data mining you might find the following information:

1. The customers who buy cheese, they also buy bread. You can put cheeses and bread near each other.
2. During holidays customers buy more fast foods such as hamburgers, tuna fish. You can put more of these foods at your supermarket on holidays.
3. Special customers for special occasions order special kind of stuffs. By sending their desired food you can surprise them (risk is part of everything!!!).

Other important usage of data mining can be found at (Hsiang-Chuan Liu, 2008), (Peter C. Austin, 2010).
Some words are important and necessary and you should remember them such as attribute, instance, classification, association, clustering, supervised and unsupervised learning, missing value, overfitting, and target. These terms will be explained shortly during this chapter and also related terms to this chapter will be covered completely.
In this chapter we will focus on linear regression, logistic regression, and neural network (Perceptron) and we will provide sufficient practical examples to make this concept easier to understand. In this way, we will use some free data mining software such as WEKA(www.cs.waikato.ac.nz) and RapidMiner(www.rapidminer.com) which are written by Waikato and Yale University, respectively.
At the end, we will focus on one the important area of regression method which is not well known. This part of the book has a wide variety of use in security such as breaking some patterns of serial numbers, wireless security keys, and so on.

2. Basic concepts of data mining

In data mining, data can be divided into five groups. In other words, attributes can be categorized into five groups: nominal (categorical), numeric (integer, continuous), ordinal, interval and ratio. These terms will be explained in the next paragraphs.
For enlightening, consider weather attribute. The values of the weather attribute can be sunny, rainy, or cloudy. Definitely these values are not comparable or multipliable or not appropriate for mathematical operations. These values are nominal. But, the length attribute can be assigned any numeric value within the range of Natural numbers.
Numeric attributes measure numbers, whether integer or real. Nominal attributes have values that are distinct or can be considered just a label or name. Nominal is the Latin word for name. Consider these two values: hot and cold, you can arrange them but you cannot define any instances. For example you can say, hot is warmer than cold but you do not know how much the difference in degree is. These kinds of attributes are ordinal. The comparison is logical but subtract or add is not acceptable. It might be a little hard sometimes to distinct nominal and ordinal quantities. It depends on user.
Consider the year, for example 2010 and 2012. You cannot add them or subtract them because it does not make sense. You can say 2012 is 2 years greater than 2010, but you cannot say 1.0009 times the year 2010 because year 0 is totally arbitrary and historians chose it. These kinds of attributes are interval.
But if you consider the distance between the object and itself, that is zero, thus distance is a ratio quantity. Mathematical operation is logical and for example it makes sense to multiply 3.14 times a distance to get an circle's area. Instances make dataset. Every single piece of data is an instance. Instances some times are called examples. Each instance is useful and is a part of learning.
Instances are categorized based on the values of features; attributes; that measure different aspects of instances.
Target is an attribute that the instances want to be classified into.
If the target be one and after doing data mining we got rules such as this:
If weather be sunny then the temperature is around 40.
Then this is classification. In other words, classification predicts the value of a given attribute. If these rules are used to predict the value of any attribute then it is association rules. In other words, an association predicts the value of arbitrary an attribute(s).
If temperature = cool then humidity = normal
If temperature = high and temperature $\geq$ *60 then humidity = high*
In Clustering, the groups of examples that belong together are sought.
If the input(s) are assigned to at least one output, and the learning uses the outputs, then this is supervised learning. The unsupervised learning is totally opposite.
If there is no output(s) or the output(s) does not used during learning, then it is unsupervised learning. Please be aware of this matter that the output during the supervised learning is the same as target. Simply, if there is a target and that target is used for learning, then it is supervised learning, else it is unsupervised learning. Classification learning sometimes is called supervised learning because the attributes or the target acts as an input. Missing values are missed values! If you are collecting data, it might be impossible for you to find some data, and then these data are missing data and they will be replace by a question mark "?" like below.
Overfitting is a concept that will occur on following condition:

Weather	*Sunny*	*Rainy*
Temperature	30	?
Play	No	Yes

Table 1. Missing values.

Overfitting might happen when training data are finite and if the learning model cover all of the data. In the following figure, fig 1, the concept of overfitting is totally obvious. Although for the training data the error is minimum, for the testing data, the error will be very high (Kantardzic, 2003), (Witten & Frank, 2005).

3. Regression concept

If you start with regression, you might find it a little confusing. So it is better to forget the meaning of regression in your literature readings.
In statistics, regression analysis is the concept of understanding the relation between independent and dependent variables. Precisely, it tries to understand how the value of dependent variable changes while one of the independent variable is varying when the other independent variables are fixed.
One of the main job of regression is forecasting and predicting. Another job is helping to find out which of independent variables has the most or less (or no) effect on the dependent variable.
There are lots of developed algorithms and functions that are for regression analysis such as linear regression or logistic regression. In the following pages, we make you familiar with linear regression,multilayer perceptron,logistic regression. Then, two of data mining tools will be introduced and two practical examples will be shown. At last, we will focus on one of the most important but less famous usage of data mining which is security. And we provide some useful example of using regression analysis in cracking and breaking serial numbers (Kantardzic, 2003), (Witten & Frank, 2005).

4. Linear regression

Linear regression analyses the relationship between two variables (X, Y) and tries to model the relationship by fitting a linear equation to the observed data. These two variables should be numeric. The linear regression line as a standard curve tries to find new values of X from

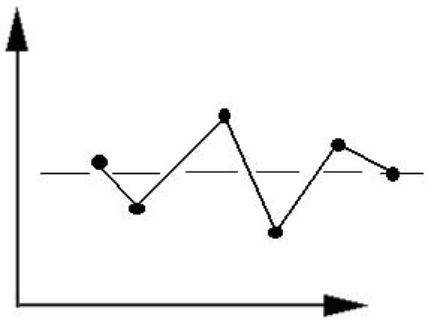

Fig. 1. Overfitting.

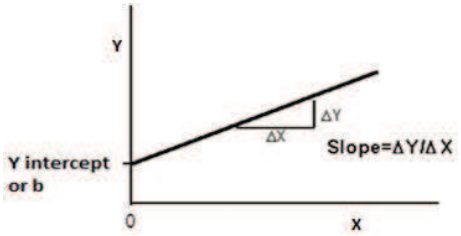

Fig. 2. Intercept and slope.

Y, or Y from X. A linear regression line has an equation like $Y = a + bX$, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when $x = 0$.)
In data mining form, expressing the class as a linear combination of the attributes, with predetermined weights is linear regression.

$$x = w_0 + w_1 a_1 + w_2 a_2 + \ldots + w_k a_k \tag{1}$$

x is the class; $a_1, a_2, \ldots$ are the attribute values; and $w_0, w_1, \ldots$ are weights.
$x^{(1)}$ is the class of the first instance and the superscript above the attribute values denotes that it is the first example.

$a_1^{(1)}$,
$a_2^{(1)}$,
$\vdots$
$a_k^{(1)}$,

$$w_0 a_0^{(1)} + w_1 a_1^{(1)} + w_2 a_2^{(1)} + + w_k a_k^{(1)} = \sum_{j=0}^{k} w_j a_j^{(1)} \tag{2}$$

The next part is choosing the coefficients w_j—there are $k + 1$ of them-to minimize the sum of the squares of these differences over all the training instances. n is number of training instances. Then the sum of the squares of the differences is shown in the following formula (Witten & Frank, 2005).

$$\sum_{i=1}^{n} (x^{(i)} - \sum_{j=0}^{k} w_j a_j^{(i)}) \tag{3}$$

The expression inside the parentheses is the difference between the ith instance's actual class and its predicted class.
The most common method for finding the regression line is the least-squares.
This method calculates the best-fitting line for the observed data by minimizing the sum of the squares. This method is shown in the following example.
The mathematical form of least square is summarized as follows:

$$b = (\sum y - m \sum x)/n \tag{4}$$

$$r = (n\sum(xy) - \sum x \sum y)/(\sqrt{([n\sum x^2 - \sum x)^2][n\sum(y^2) - (\sum y)^2]} \quad (5)$$

$$m = n\sum(xy) - \sum x \sum y / n\sum(x^2) - (\sum x)^2 \quad (6)$$

"m" is slope, "b" is intercept and "r" is correlation coefficient. Linear correlation coefficient, measures the strength and the direction of a linear relationship between two variables. Look at the following example:

$Xvalues$	$Yvalues$
40	4
41	6
42	5
43	8
44	7

Table 2. Finding linear regression between two variables.

Now, we will find slope and intercept. Afterward, we use them to form regression equation.

1. Find the number of values N=5
2. Find XY, X^2 as below

$Xvalues$	$Yvalues$	XY	X^2
40	4	160	1600
41	6	246	1681
42	5	210	1764
43	8	344	1649
44	7	308	1936

Table 3. Find the linear regression between two variables.

3. Find $\sum X$,$\sum Y$,$\sum XY$, $\sum X^2$.
   $\sum X = 210$
   $\sum Y = 30$
   $\sum XY = 1286$
   $\sum X^2 = 8830$
4. Substitute in (6),Slope will be 0.8.
5. Substitute in (4),intercept will be -27.6.
6. Substitute these values in regression equation formula
   Regression Equation: $y = a + bx$, $y = -27.6 + 0.8x$

Suppose, we want to know the approximate y value for the variable $x = 10$. So, we can substitute the value in the above equation.The result is:

Regression Equation:

$$y = a + bx$$
$$y = -27.6 + 0.8 * x$$
$$y = -27.6 + 0.8 * 10 = -19.6$$

5. Neural network

Neural Network (NN) is a simulated neural cell by hardware or software. In this section, terms like neuron, learning, and experience are referring to the concepts of neural networking in a computer system.
Neural networks have the ability to learn by examples. We will discuss neurons, NNs in general, Multilayer Perceptron, and Back Propagation networks. Multilayer Perceptron networks are popular types of network that can be trained to recognize different patterns including images, signals, and texts (M.K. Alsmadi, 2009), (Nirkhi, 2010), (Peter Auer, 2008).

5.1 History

The history of some of the NN algorithms is summarized as follows:

- 1943 McCulloch-Pitts neuron model
- 1949 Hebbian Network
- 1958 Single Layer Perceptron
- 1982 Hopfield Network
- 1982 Kohonen Self Organization Map(SOM)
- 1986 Back Propagation(BP)
- 1990's Radial Basis Function Network
- 2000's Support Vector Machine(SVM)

5.2 Important functions of NNs

There are four main functions in NNs that are shown below.

1. Identity (Linear) Function
2. Binary Step Function With Threshold θ(Heaviside)[threshold OR hard limit if $\theta = 0$]
3. Bipolar Step Function With Threshold θ [Sign OR symmetrical hard limit if $\theta = 0$]
4. Sigmoid Function (S-shaped Curves)
   a. Binary Sigmoid(Logistic OR Log-Sigmoid)
   b. Bipolar Sigmoid
   c. Hyperbolic Tangent
   d. ArcTan

Fig 3 is linear function, Fig. 4 is Binary Step Function and the two equations under it are its equations, Fig. 5 is Bipolar Step Function and the two equations under it are its equations, and at last Fig. 6 is Binary Sigmoid Function.
The function which is shown in Fig. 6 is Sigmoid function. The coefficient "a" is a number constant and can be chosen between 0.5 and 2.
σ stepness usually $\sigma > 0$
$F(x) = 1/(1 + exp(-\sigma x)) = 1/(1 + e^{(-\sigma x)})$
$f'(x) = dx/dy = \sigma f(x)[1 - f(x)]$

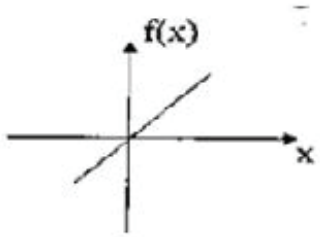

Fig. 3. Identity (Linear) Function $f(x) = x, forall x$

5.3 Neuron

The neuron can be thought as a program, or process that has one or more inputs and produces an output. The inputs simulate what a neuron gets, while the output is what a neuron generates. The following figures can clarify this concept more, fig 7.

5.4 Neural networks definition

A neural network is a group of neurons connected together. Connecting neurons together to form a neural net can be done in different ways such as SOM or Multilayer Perceptron.

5.5 Multilayer pereceptron

Multilayer perceptron (MLP) is a function that learns through back propagation algorithm. Back propagation pseudo-code $(http : //scialert.net/fulltext/?doi = ajsr.2008.146.152 \&org = 11)$. is explained below

The following steps show a Back Propagation NN:

Step 0. Initialize weights and biases.

Step 1. While stopping condition is false, do steps 2-9.

Step 2. For each training pair, do steps 3-8.
Feedforward:

Step 3. Each input unit $(X_i, i = 1, \ldots, n)$ receives input signal x_i

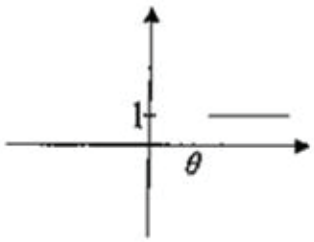

Fig. 4. Binary Step Function with Threshold θ.

$f(x) = 1$ if $x => \theta$
$f(x) = 0$ if $x < \theta$

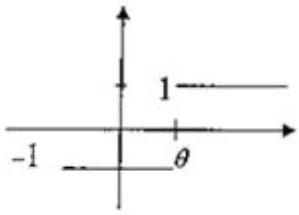

Fig. 5. Bipolar Step Function with Thresholdθ.

$f(x) = 1$ if $x => \theta$
$f(x) = -1$ if $x < \theta$

and broadcasts this signal to all units in hidden layer.

Step 4. Each hidden unit $(Z_j, j = 1,\ldots,p)$ sums its weighted input signals,
$Z_{inj} = v_{0j} + \sum_{i=1}^{n} x_i v_i j$
And applies its activation function to compute its output signal,
$Z_j = f(Z_{inj})$
And sends this signal to all units in the output layer.

Step 5. Each output unit $(Y_k, K = 1,\ldots,m)$ sums its weighted input signals,
$y_{ink} = w_{0k} + \sum_{j=1}^{n} z_j w_{jk}$
And applies its activation function to compute its output signal,
$y_k = f(y_{ink})$
Backprpagation of error:

Step 6. Each output unit $(Y_k, K = 1,\ldots,m)$ receives a target
pattern corresponding to input training pattern, computes its error information term,
$\delta_K = (t_k - y_k) f'(y_{ink})$
Calculates its weight correction term,
$\Delta w_{jk} = \alpha \delta_k z_j$
And calculate its bias correction term,
$\Delta w_{0k} = \alpha \delta_k$
And sends δ_K to units in hidden layer.

Step 7. Each hidden unit $(Z_j, j = 1,\ldots,p)$ sums its delta inputs

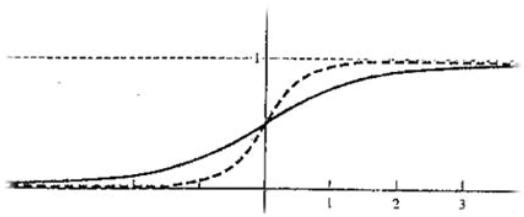

Fig. 6. Binary Sigmoid Function.

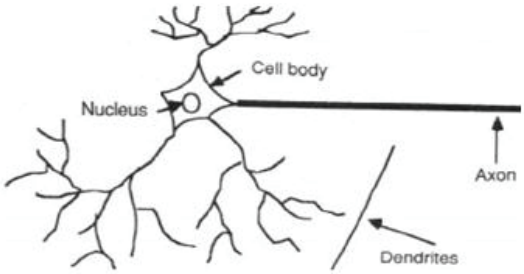

Fig. 7. Natural neuron.

from units in the output layer,
$\delta_{inj} = \sum_{(k=1)}^{m} \delta_K w_{jk}$
And multiplies by derivative of its activation function to calculate its error information term,
$\delta_j = \delta_{inj} f'(z_{inj})$
Calculates its weight correction term,
$\Delta v_{ij} = \alpha \delta_i x_i$
And calculates its bias correction term,
$\Delta v_{0j} = \alpha \delta_j$
Updates weights and biases:

Step 8. Each output unit $(Y_k, K = 1,\ldots,m)$ updates its weights and bias (j=0,...,p):
$W_{jk}(new) = W_{jk}(old) + \Delta w_{jk}$
Each hidden unit $(Z_j, j = 1,\ldots,p)$ updates its weights and bias (i=0,...,n):
$V_{ij}(new) = v_{ij}(old) + \Delta v_{ij}$
Step 9. Test stopping condition.

Two of the most important functions of MLP are Bipolar Sigmoid and Binary Sigmoid. Please consider the next example:
input vector is (0,1)
target is 1
learning rate (α) is 0.25
n=2
p=2
activity function is Binary Sigmoid and slope (m) is 1
$\sigma = 1$

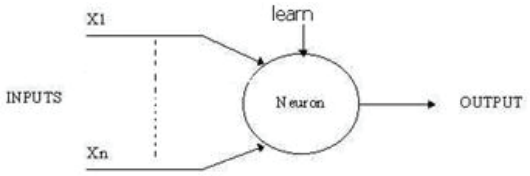

Fig. 8. Computer neuron (simulated).

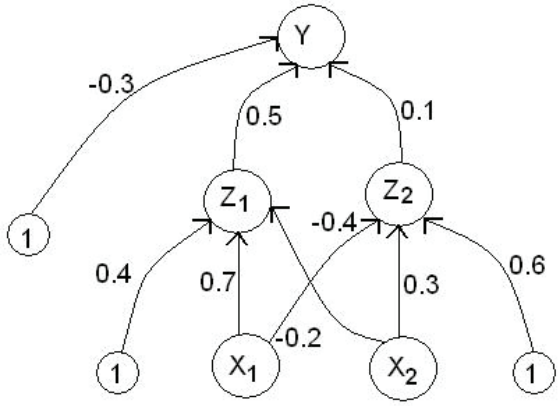

Fig. 9. MLP.

find weights and biases for MLP with above information, and continue until you reach the floating-point with three digits.
$f(x) = 1/1 + e^{-x}$
$f'(x) = f(x)[1 - f(x)]$

Step 0.Initialize weights and biases

Step 1.Begin training:

Step 2.For input vector $X = (0,1)$ with $t_1 = 1$,do steps 3-8.

Feedforward:

Step 3.$x_1 = 0$,$x_2 = 1$

Step 4.For j=1, 2:

$Z_{inj} = v_{0j} + \sum_{i=1}^{n} x_i v_i j$

$z_{in1} = 0.4 + 0 * 0.7 + 1 * (-0.2) = 0.2$

$z_{in2} = 0.6 + 0 * (-0.4) + 1 * 0.3 = 0.9$

$z_j = f(z_{inj})$

$z_1 = 0.550$

$z_2 = 0.711$

Step 5.For k=1:

$y_{ink} = w_{0k} + \sum_{j=1}^{p} z_j w_{jk}$

$y_{in1} = -0.3 + 0.550 * 0.5 + 0.711 * 0.1 = 0.046$

$y_k = f(y_{ink})$

$y_1 = 0.512$

Backpropagation of error

Step 6.For k=1:

$\delta_K = (t_k - y_k) f'(y_{ink})$

$\delta_{k=1} = (1 - 0.512) * f'(0.046) = 0.122$

and for j=1,2:

$\Delta w_{jk} = \alpha \delta_k z_j$

$\Delta w_{11} = 0.25 * 0.122 * 0.550 = 0.017$

$\Delta w_{21} = 0.25 * 0.122 * 0.711 = 0.022$

$\Delta w_{0k} = \alpha \delta_k$

$\Delta w_{01} = 0.25 * 0.122 = 0.031$

Step 7.For j=1,2:

$\delta_{inj} = \sum_{k=1}^{m} \delta_K w_{jk}$

$\delta_{in1} = 0.122 * 0.5 = 0.061$

$\delta_{in2} = 0.122 * 0.1 = 0.012$

$\delta_j = \delta_{inj} f'(z_{inj})$

$\delta_{j=1} = 0.061 * f'(0.2) = 0.015$

$\delta_{j=2} = 0.012 * f'(0.9) = 0.002$

and for i=1,2:

$\Delta v_{ij} = \alpha \delta_i x_i$

$\Delta v_{11} = 0.25 * 0.015 * 0 = 0.000$

$\Delta v_{21} = 0.25 * 0.015 * 1 = 0.004$

$\Delta v_{12} = 0.25 * 0.002 * 0 = 0.000$

$\Delta v_{22} = 0.25 * 0.002 * 1 = 0.001$

$\Delta v_{0j} = \alpha \delta_j$

$\Delta v_{01} = 0.25 * 0.015 = 0.004$

$\Delta v_{02} = 0.25 * 0.002 = 0.001$

Update weights and biases

Step 8. For k=1 and j=0,1,2:

$W_{jk}(new) = W_{jk}(old) + \Delta w_{jk}$

$W_{11}(new) = 0.517$

$W_{21}(new) = 0.122$

$W_{01}(new) = -0.269$

for j=1,2 and i=0,1,2:

$V_{ij}(new) = v_{ij}(old) + \Delta v_{ij}$

$V_{11}(new) = 0.700$

$V_{21}(new) = -0.196$

$V_{12}(new) = -0.400$

$V_{22}(new) = 0.301$

$V_{01}(new) = 0.404$

$V_{02}(new) = 0.601$

Step 9.Test stopping condition.

6. Logistic regression

Logistic regression is part of regression model called generalized linear models (Kantardzic, 2003), (Witten & Frank, 2005), (Handan Ankarali Camdeviren, 2007), (Hsiang-Chuan Liu,

2008). A logistic regression example is shown in the Fig. 10.
The Fig. 10 can be written as the following formula:

$$f(z) = e^z / e^z + 1 = 1/1 + e^{-z} \tag{7}$$

The most important thing about the logistic regression is that the input value can be any value from negative infinity to positive infinity. But the output value only can be between zero and one. The variable z is usually defined as

$$z = B_0 + B_1 x_1 + B_2 x_2 + \ldots + B_k x_k \tag{8}$$

where B_0 is called the intercept and B_1, B_2, B_3, and so on, are called the regression coefficients of x_1, x_2, x_3 respectively.
The two main formulas in statistics which are used in logistic regression are shown below, more information available at (*http* : *//luna.cas.usf.edu/ mbrannic/files/regression/Logistic.html*):

$$Odds(x) = Pr(x)/[1 - Pr(x)] \tag{9}$$

$$Prob = Odds/(1 + Odds) \tag{10}$$

The application of logistic regression may be illustrated by using a fictitious example of death from diabet disease. This simplified model uses only three risk factors (age, sex, and blood Glucose level) to predict the 20-year risk of death from diabet disease. This is the model:

B_0 = - 7.0 (the intercept)
B_1 = + 2.2
B_2 = - 2.0
B_3 = + 1.2
x_1 = age in years, less than 50
x_2 = sex, where 0 is male and 1 is female
x_3 = Glucose level, in $mmol/L$ above 200
Which means the model is
risk of death is:$= 1/1 + e^{-z}$, where $z = -7 + 2.2x_1 - 2x_2 + 1.2x_3$

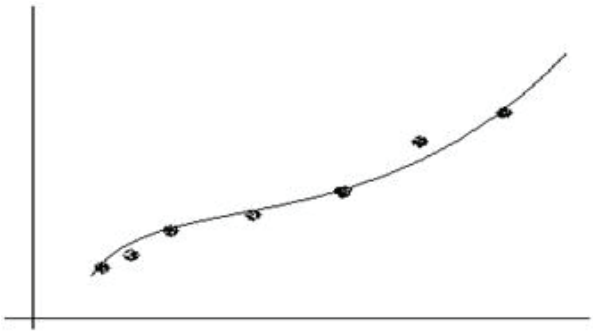

Fig. 10. Logistic regression.

In this model, increasing age is associated with an increase in risk of death from diabet disease (z goes up by 2.2 for every year over the age of 50), female sex is associated with a decrease in risk of death from diabet disease (z goes down by 2.0 if the patient is female), and increasing Glucose is associated with an increase risk of death (z goes up by 1.2 for each 1 $mmol/L$ increase in Glucose above 200). This model will be used to predict Mohsen's risk of death from diabet disease: he is 50 years old and his glucose level is 205. Mohsen's risk of death is therefore
$1/1+e^{-z}$,where $z=-7+2.2*(50-50)-2*(0)+1.2*(205-200)$
This means that by this model, Mohsen's risk of dying from diabet disease in the next 20 years is 0.26.

7. Practical example

Now let's us start some practical examples. The first one will be done by WEKA and the second one by RapidMiner. First of all, we need a data set. Data set is a collection of recorded data in a specific format that you will be familiar with in the next few lines. Our data set name is cmc and its extension is "arff". If you search "cmc.arff" in google you can find and download it easily. When you download it, right click on it and chose "open with" and then open it with "notepad". Better software is "Notepad++" which is free to download and can be easily found through the web. As soon as you open it you will see some things like this:

%1.*Title* : *ContraceptiveMethodChoice*
%2.*Sources* :
%(*a*)*Origin* : *Thisdatasetisasubsetofthe*1987*NationalIndonesia*
%*ContraceptivePrevalenceSurvey*
%.....

@*relationcmc*
@*attribute* $Wifes-age$ $INTEGER$
@*attribute* $Wifes-education$ 1,2,3,4
@*attribute* $Husbands-education$ 1,2,3,4
@*attribute* $Number-of-children-ever-born$ $INTEGER$
@.....

@*data*

24,2,3,3,1,1,2,3,0,1
45,1,3,10,1,1,3,4,0,1
....
As you can see it is composed of 5 groups.
First group is: %. Whatever line started with this is a comment for user.
Second group is: $'$@*relationcmc*$'$. This is the name of dataset.
Third group is: $'$@*attributeWifes*$-$*ageINTEGER*$'$. This line says that Wifes-age is an attribute and its type is Integer. Integer and Real belongs to Numeric types. The next line of this group says that $Wifes-education$ is an attribute which it has just four values as 1,2,3,4. These numbers can be interpreted as labels. By reading the **first group** you can find out that these number are referring to what. For example 1 means low education$(1=low,2,3,4=high)$.
Fourth group is: $'$@*data*$'$. This means that the data is started from the next line.

Fifth group is: $'24,2,3,3,1,1,2,3,0,1'$. This line is start of data. This can be interpreted like this: The attribute which is Wifes-age has value 24, the second attribute which is Wifes-education has value 2, and so far. There are some important rules here such as the number of attributes should be the same as number of values in data part. For example, if we have 10 value in each line of data which are separated by $','$ and we should have 10 attributes.
If you read more and do more practice you can find out more rules. One of the best resources is chapter 7 to 14 of (Witten & Frank, 2005)
Let's go and execute linear regression algorithm on this data set. For executing linear regression the target should be numeric and it is better that other attributes be numeric but it is depend on the usage and aim of linear regression. Without any purpose but only making familiar reader with linear regression we change all attributes to numeric by just renaming the type of attributes.
At the end it is like this:
$@attribute\ Wifes-age$ numeric
$@attribute\ Wifes-education$ numeric
$@attribute\ Husbands-education$ numeric
$@attribute\ Number-of-children-ever-born$ numeric
$@attribute\ Wifes-religion$ numeric
$@attribute\ Wifes-now-working$ numeric
$@attribute\ Husbands-occupation$ numeric
$@attribute\ Standard-of-living-index$ numeric
$@attribute\ Media-exposure$ numeric
$@attribute\ Contraceptive-method-used$ numeric

Sometimes for some purposes you can execute filter on your data such as converting numeric data to nominal or removing some attributes. The following figure,fig 11, shows the place of filters in Weka.

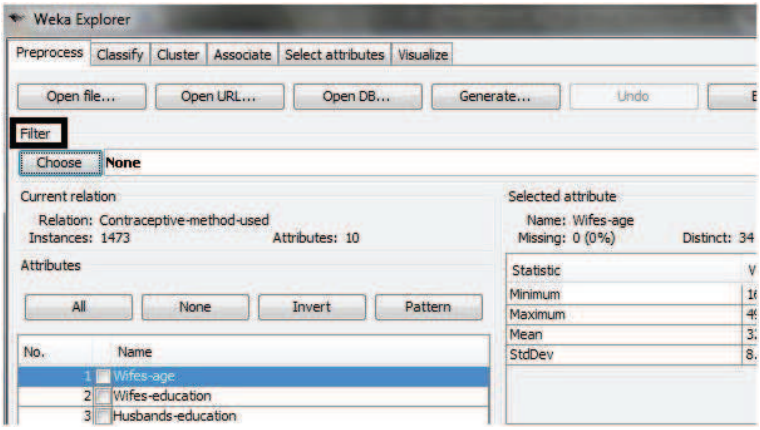

Fig. 11. Filter.

For executing linear regression, we chose "classify" from top tab (as shown in the above picture, fig 12). Then we chose "linear regression" from functions and leave other setting unchanged. Afterwards, we chose the last attribute as the target as shown in the below image, fig 13, and click on start to execute the algorithm.

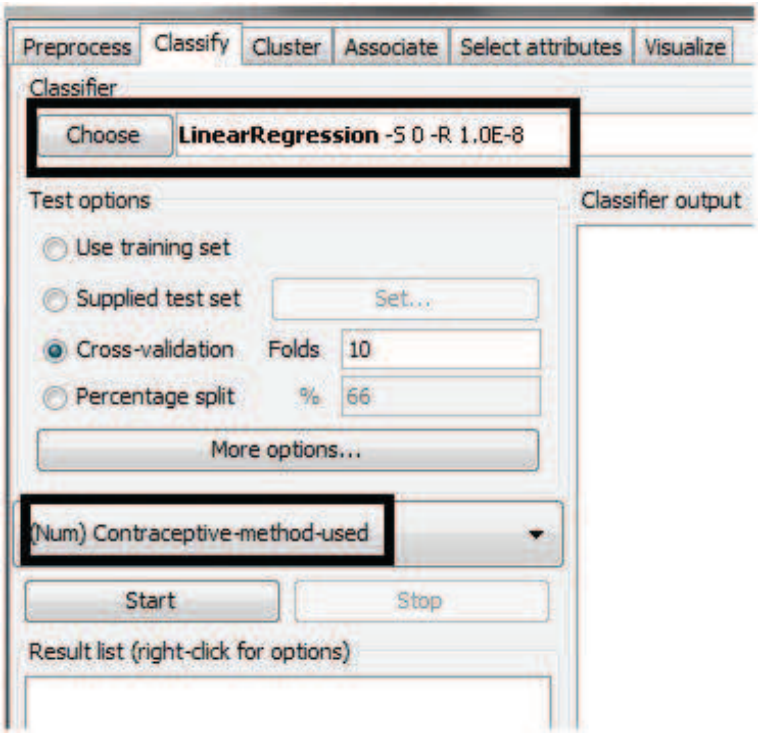

Fig. 12. Classify.

Output is shown in the following image, fig 13.
As you can see in figure 13, the regression equation based on the target ($'contraceptive - method - used'$) is found and also some other values such as correlation coefficient are also found.
Enough is enough. Let's go to a very simple security example. A good example is in (M. Hajsalehi Sichani, 2009). Imagine you are a programmer and you have created software and you have designed a system for entering activation code. Its algorithm is like this:

1. Get the CPU id, like 2300
2. Multiply it by 3 and give it back to user as given-number, $2300 * 3 = 6900$
3. User must call you and tell you his given number (6900) and you put this number in the following equation:
   $3 * x + 5$
   and you give him 23705 $(3 * 6900 + 5 = 20705)$.
4. Now the user enters 20705 in the software as activation code.
5. Your program will substitute the given number in the equation $3 * x + 5$, and if the activating number is equal to the result then it let the user to use your software.

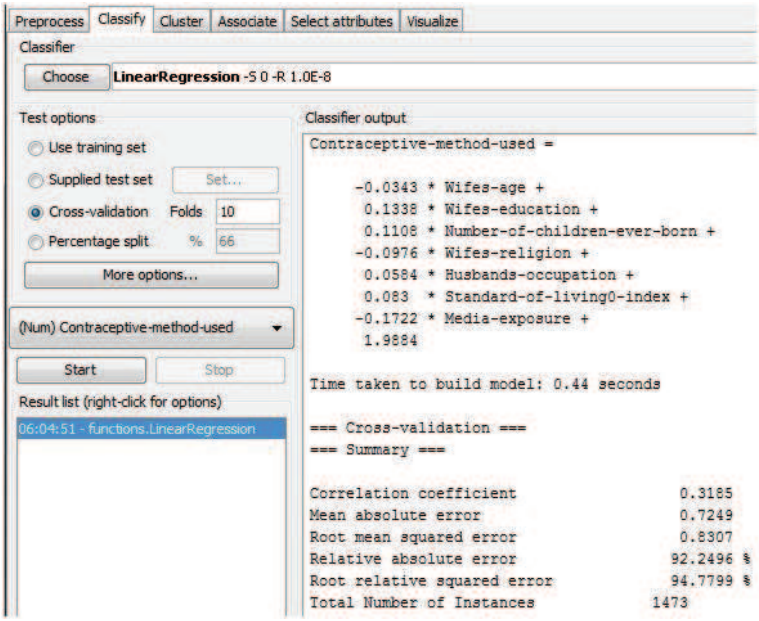

Fig. 13. Output of Weka.

Notice that instead of CPU id you can get his name and convert it to Ascii codes which are also integers number. Remember that in reality these kinds of algorithms are much more complicated than here.
Now as a cracker, Saeed, calls you and wants to activate the following numbers (left column) and you gave him the activation numbers (right column). Then he changes the data to an acceptable format (arff). The following lines are content of arff file.

@relationcrack
@attribute given — number numeric
@attribute activation — code numeric

@data

6900,20705
6903,20714
6906,20723
6909,20732
6912,20741
6915,20750
6918,20759
6921,20768
6927,20786
6930,20759

6936, 20813
Then will start his work with RapidMiner. We will persuade him from now in figure 14 through 18, respectively.

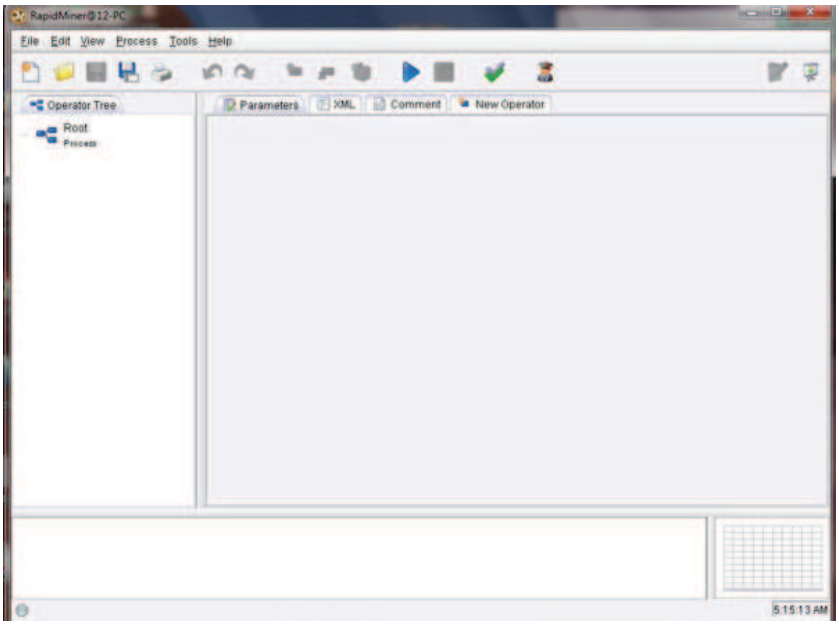

Fig. 14. Rapidminer enviroment.

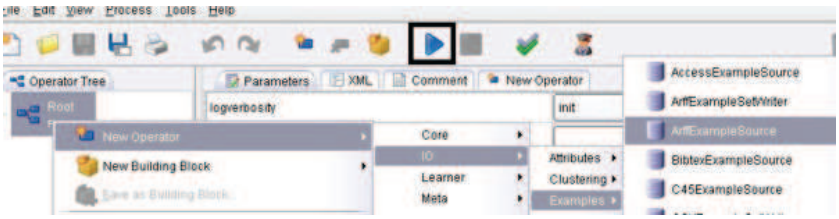

Fig. 15. Rapidminer first step.

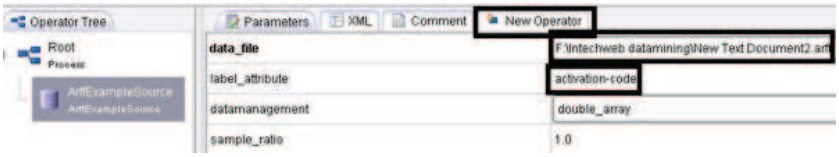

Fig. 16. Rapidminer second step.

As you can see in fig 18, the RapidMiner found the equation and the pattern behind the data.

8. Conclusion

We hope, in this chapter, you became familiar with the basic concept of data mining, linear regression, logistic regression, and neural network.

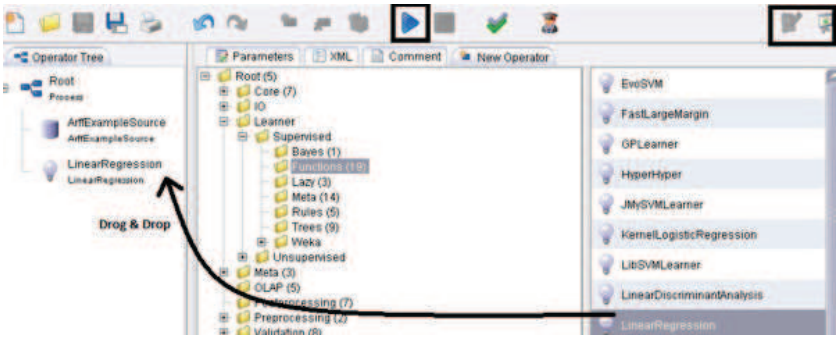

Fig. 17. Rapidminer 3rd step.

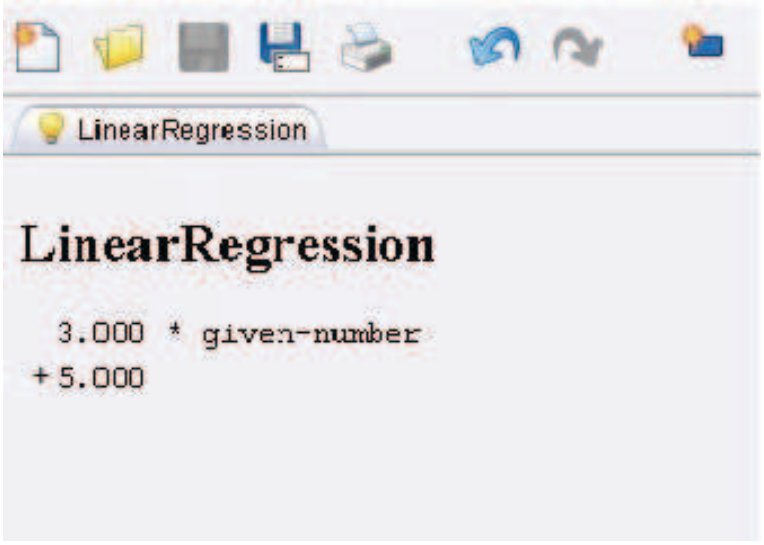

Fig. 18. Rapidminer found equation!.

At the end of this chapter, we focus on two of the data mining tools, Weka and RapidMiner, and show one practical example by each of them, individually. The second practical example was a security example which was a simplified one. Other data mining software are exists but may not be free like SPSS. The similar logic is behind them and if you know how to work with one of them, you can work with the rest of them. Just install them and start working.
At last, we hope you have found this ability to go and study data mining by your-self and use different resources such as google, sciencedirect, and IEEE.
We would announce a great thanks to H. Ghominejad for her technical support and also a great thanks to *Intechweb.org* team for their support.

9. References

Handan Ankarali Camdeviren, Ayse Canan Yazici, Z. A. R. B.-M. A. S. (2007). Comparison of logistic regression model and classification tree: An application to postpartum depression data, *Expert Systems with Applications* vol. 32: 987–994. www.sciencedirect.com.

Hsiang-Chuan Liu, Shin-Wu Liu, P.-C. C. W.-C. H. C.-H. L. (2008). A novel classifier for influenza a viruses based on svm and logistic regression, *International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR '08* Vol. 1: 287–291. www.IEEE.org.

Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons.

M. Hajsalehi Sichani, A. M. (2009). A new analysis of rc4: A data mining approach (j48). www.secrypt.com.

M.K. Alsmadi, K. Bin Omar, S. N.-I. A. (2009). Performance comparison of multi-layer perceptron (back propagation, delta rule and perceptron) algorithms in neural networks, *IEEE International Advance Computing Conference, IACC 2009* pp. 296–299. www.IEEE.org.

Nirkhi, S. (2010). Potential use of artificial neural network in data mining, *The 2nd International Conference on Computer and Automation Engineering (ICCAE)* Vol. 2: 339–343. www.IEEE.org.

Peter Auer, Harald Burgsteiner, W. M. (2008). A learning rule for very simple universal approximators consisting of a single layer of perceptrons, *Neural Networks* vol. 21: 786–795. www.sciencedirect.com.

Peter C. Austin, Jack V. Tu, D. S. L. (2010). Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure, *Journal of Clinical Epidemiology, In Press, Corrected Proof, Available online 21 March 2010* . www.sciencedirect.com.

Witten, I. H. & Frank, E. (2005). *Data Mining : Practical machine learning tools and techniques*, 2nd edn, Morgan Kaufmann series in data management systems, UNITED STATES OF AMERICA.

Knowledge-Oriented Applications in Data Mining
Edited by Prof. Kimito Funatsu

ISBN 978-953-307-154-1
Hard cover, 442 pages
Publisher InTech
Published online 21, January, 2011
Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by 'Data Mining' address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821