Machine Learning

5. week

- Linear Regression
- Multiple Linear Regression
  - Least squares method

Umut ORHAN, PhD.

Linear Regression

In linear regression, relation between an input and an output variables is described with a linear equation, and it is assumed that this equation can produce targets with an unpredictable error. This equation is

$$Y = W_0 + W_1 X + E$$

Here, W_i represent weights, X shows input variable, E is unpredictable error, and Y is target variable.

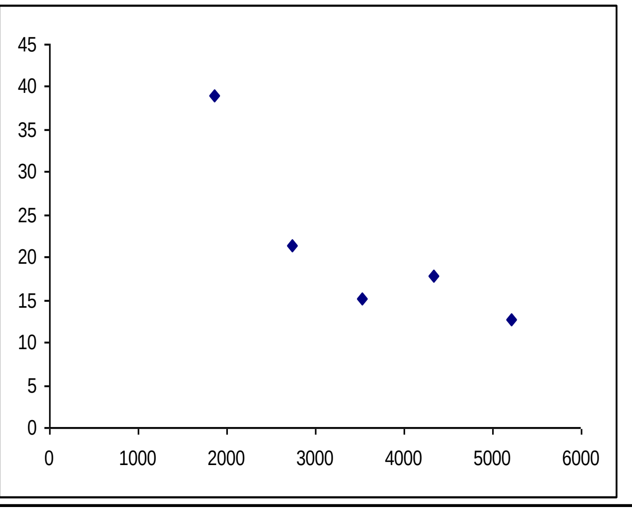
Umut ORHAN, PhD.

Linear Regression

Each data sample is distant with an error (*E*). This error is assumed such a small that it is neglected. By using data with *X* and *Y* vectors, detection of all *W* weight is the main aim of linear regression.

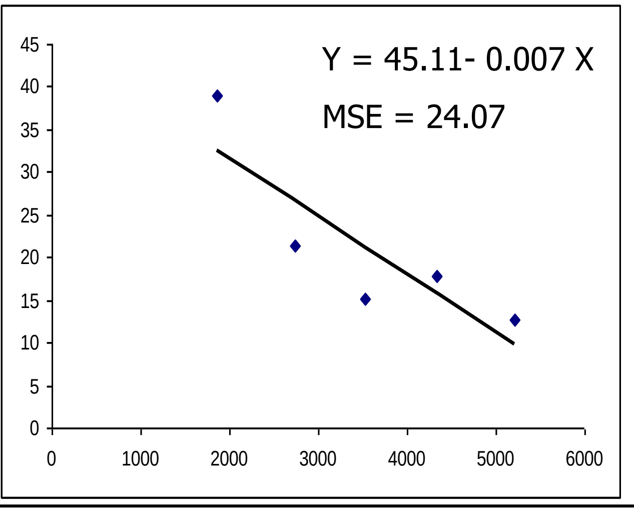
MSE is frequently used to measure and compare of successes of prediction and estimation methods like regression.

Sample



X	D
2743	21,4
3518	15,2
1855	38,9
5214	12,7
4341	17,8

Sample



X	D
2743	21,4
3518	15,2
1855	38,9
5214	12,7
4341	17,8

Multiple Linear Regression

It is a version of linear regression to implement a data with *M* features (variables). It is assumed that summation of weighted values of all input variables produces target with an error.

$$Y = W_0 + \sum_{j=1}^{M} W_j X_j + E$$

W_j represent weights, X_j are variables (or features), E is error, and Y is target variable.

Multiple Linear Regression

Distance of data samples from a M+1 dimensional hyper-plane is negligible. The aim is, as in simple linear regression, to determine W weights.

Solution of both simple and multiple linear regression is based on least squares method.

Least Squares Method

Mathematical solution of least squares method can be expressed as below.

$$W_j = \frac{X_j D}{X_j^2} = \frac{\sum_{i=1}^{N} x_{ij} d_i}{\sum_{i=1}^{N} x_{ij}^2} \qquad \mathbf{W} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{D}$$

Here, N is the total number of data samples, i is index of the data samples, and j is index of the data features.

Least Squares Method

The optimization of least squares method is based on minimization of mean squared error (MSE) equation given below.

$$\frac{1}{N}\sum_{i=1}^{N} e_i^2 = \frac{1}{N}\sum_{i=1}^{N}\left(d_i - \sum_{j=1}^{M} w_j x_{ij}\right)^2$$

Here, e_i are unpredictable errors, w_j are weights, d_i are desired targets in data, and x_{ij} are inputs.

Competitive Classification

A prediction method like regression can be used in classification by thresholding.

For data with two-classes, if the labels (desired targets in data) are chosen as 0 and 1, only one threshold is implemented as 0.5.

But if there are more classes, then we should determine output by competition.

Competitive Classification

Depending on the aim, competition is based on to choose minimum or maximum one of many results.

Sample

A part from popular IRIS dataset is given at right.

We have an input matrix with 4 features and an output with 3 classes.

Because classification methods can easily learn data with two-classes, we should transform D labels into three different binary expressions.

X_1	X_2	X_3	X_4	D
5.1	3.5	1.4	0.2	A
4.9	3	1.4	0.2	A
7	3.2	4.7	1.4	B
6.4	3.2	4.5	1.5	B
6.3	3.3	6	2.5	C
5.8	2.7	5.1	1.9	C

Sample

Unlike classification methods, prediction systems like regression can produce continuous values.

For this reason, in systems with one binary output, we should use thresholding to turn continuous values into binary form.

But here, we will use competitive choosing.

D	D_1	D_2	D_3
A	1	0	0
A	1	0	0
B	0	1	0
B	0	1	0
C	0	0	1
C	0	0	1

Sample

Now, we'll solve 3 problems by multiple linear regression,

$$D_1 = 1.2 + 0.29X_1 - 0.16X_2 - 1.02X_3 + 1.43X_4$$
$$D_2 = -3.4 + 0.82X_1 - 0.2X_2 + 0.04X_3 - 0.52X_4$$
$$D_3 = 3.2 - 1.11X_1 + 0.35X_2 + 0.98X_3 - 0.92X_4$$

Computed mean squared errors (MSEs);

for D_1 , MSE=0.030,

for D_2 , MSE=0.301, (this is the worst one)

for D_3 , MSE=0.014

Sample

Here, Y variables show outputs computed by linear regression equations.

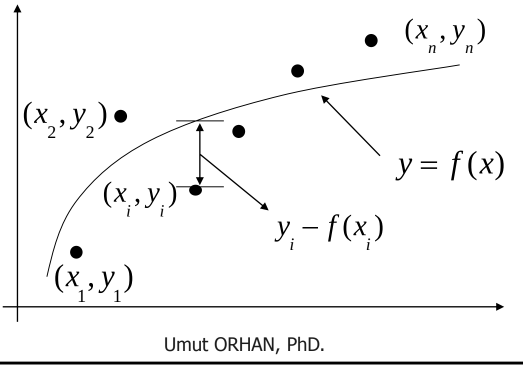
Now, we can **choose** the **maximum value** for each row to determine winning results.

Thus, all samples are classified correctly.

D	Y_1	Y_2	Y_3
A	0.98	0.06	-0.04
A	1.00	-0.01	0
B	-0.06	1.19	-0.13
B	0.11	0.63	0.25
C	-0.02	0.07	0.95
C	-0.02	0.05	0.96

Nonlinear Regression

Nonlinear regression is a form of regression analysis in which observational data are modeled by a nonlinear function.



Nonlinear Regression

There are many kind of nonlinear model. Some of them are:

1. Exponential model: $(y = ae^{bx})$
2. Power model: $(y = ax^{b})$
3. Saturation growth model: $\left(y = \dfrac{ax}{b+x}\right)$
4. Polynomial model: $(y = a_0 + a_1x + ... + a_mx^{m})$

Nonlinear Regression

The aim is the same as in linear regression:
to estimate parameters of the equation.

The data are fitted by several methods, the most known ones are maximum likelihood and gauss-newton methods.

MATLAB Application

```
>edit Regresyon_ornek.m
```

Student should study with this sample by using given datasets.

Presentation Task

You can choose one of two algorithms listed below.

- Maximum Likelihood
- Gauss-Newton