MatLab Programming Fundamentals

guarantor: Maroš Tunák
tel.: 3465
e-mail: *maros.tunak@tul.cz*

Course objectives

The aim of the course is to acquire basics knowledge and skills of students the MatLab program. At the end of the course students will be able to use MatLab for their own work and will be ready to deepen their programming skills in MatLab.

MatLab Programming Fundamentals

time requirements:	0p+2c
credits:	4
exercises:	Monday 10:40-12:15; 12:30-14:05 (B-PC2, Tunák M.)
	Tuesday 08:50-10:25; 10:40-12:15 (B-PC2, Tunák M.)
consultation:	Wednesday 10:40-12:15 (E-KHT)

Requirements on student/graded credit

1. participation in exercises (max. 3 absences)
2. elaboration of semester work (after approval of the semester work, you can attend a practical demonstration)
3. practical demonstration of acquired skills (there will be 1-2 examples to solve; elaboration time 1 hour; you can use any materials ...)

Content

IS/STAG Syllabus

1. Getting started with Matlab. Working environment, windows, paths, basic commands, variables. Loading, saving and information about variables. Help.
2. Mathematics with vectors and matrices. Creating vectors and matrices. Indexing. Special matrices. Matrix operations. Element by element operations. Relational operations, logical operations, examples and tricks.
3. Control flow. Loops, conditional statements, examples.
4. Script m-files, Function m-files.
5. Visualisation. Two-dimensional graphics. Three-dimensional graphics.
6. Graphical user interface.

7.-10. Statistics and Machine Learning Toolbox. Basics of statistical data processing, exploratory data analysis, descriptive statistics, data visualisation, hypothesis testing, confidence intervals, regression analysis, control charts.

11.-13. Solution of practical problems in textile and industrial engineering.

Literature

Recommended

MathWorks. *Getting Started with MATLAB*. [Online]. Dostupné z:
https://www.mathworks.com/help/matlab/getting-started-with-matlab.html

Study materials

http://elearning.tul.cz

Installation

http://liane.tul.cz/cz/software/MATLAB

Statistics and Machine Learning Toolbox

Basics of statistical data processing, exploratory data analysis, descriptive statistics, data visualisation, hypothesis testing, confidence intervals, regression analysis, control charts.

Linear regression

Simple linear regression

The simplest form of regression is simple linear regression, which considers a linear dependence between two quantities, the input independent variable x (regressor, predictor) and the output dependent variable Y (response). Assume that the actual relationship between Y and x is linear and the observation of Y for each x is a random variable. The expected value of Y for each x is

$$E(Y|x) = \beta_0 + \beta_1 x,$$

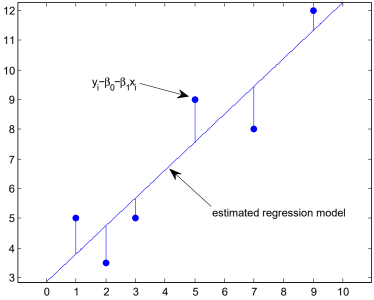
where intercept β_0 and slope β_1 are unknown regression coefficients. Assume that every observation of **Y** can be described by a model

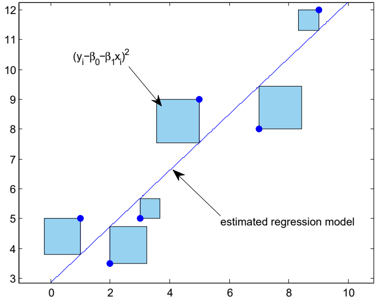
$$Y = \beta_0 + \beta_1 x + \epsilon, \qquad (1)$$

where ϵ is a random error with zero mean and unknown variance σ^2. We assume that the random errors corresponding to the individual observations are uncorrelated random variables.

Linear regression

We consider n observation pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. The left figure shows the scatter plot of the observed data, the candidate for estimating the regression model in the form of a line, and the deviation of the observation from the estimated regression model. The parameters β_0 and $\beta_1$represent estimates of regression coefficients, which according to a certain "criterion" represent the best estimate for the data. Karl Gauss proposed to minimize the sum of squares of vertical deviations in the equation 1 to estimate the parameters β_0 and β_1 (see the right figure). This criterion for estimating regression coefficients is called **least squares method**.





Linear regression

Using the equation 1 we can express the n observation in the selection as

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, 2, ..., n \tag{2}$$

and the sum of the squares of the deviations of the observations from the real regression line is given

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{3}$$

Estimates of β_0 and β_1, denote $\hat{\beta}_0$ a $\hat{\beta}_1$, must meet (minimum sum of squares of deviations)

$$\left.\frac{\delta L}{\delta \beta_0}\right|_{\hat{\beta}_0 \hat{\beta}_1} = 2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0$$

$$\left.\frac{\delta L}{\delta \beta_1}\right|_{\hat{\beta}_0 \hat{\beta}_1} = 2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0. \tag{4}$$

Linear regression

By simplification we get

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{5}$$

The system of equations 5 is the so-called **system of normal equations**. The solution of the system of normal equations is the estimates of the regression coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$. After adjusting for least squares estimates, we get

$$\hat{\beta}_0 = \overline{y} + \hat{\beta}_1 \overline{x} \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \dfrac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}. \tag{6}$$

where $\overline{y} = (1/n) \sum_{i=1}^{n} y_i$ a $\overline{x} = (1/n) \sum_{i=1}^{n} x_i$.

Linear regression

The estimated regression line is then given

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{7}$$

All pairs of observations satisfy the relationship

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i, \qquad i = 1, 2, ..., n.$$

where $e_i = y_i - \hat{y}_i$ are called **residues**. The residues describe the 'fit' error of the i -th observation model y_i. Residues provide information on the suitability of the model. Sometimes special symbols for the numerator and denominator are used in the equation 9

$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}, \tag{8}$$

$$S_{xy} = \sum_{i=1}^{n} y_i (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}. \tag{9}$$

Linear regression

- **Example:** The function is given by a table. Approximate the points by the method least squares in the form $f(x) = \beta_0 + \beta_1 x$.

x_i	-1	0	1	2	3	4	5	6
y_i	10	9	7	5	4	3	0	-1

Solution:

$$n = 8, \qquad \sum_{i=1}^{8} x_i = 20, \qquad \sum_{i=1}^{8} y_i = 37, \qquad \overline{x} = 2.5, \qquad \overline{y} = 4.625,$$

$$\sum_{i=1}^{8} x_i^2 = 92, \qquad \sum_{i=1}^{8} y_i^2 = 281, \qquad \sum_{i=1}^{8} x_i y_i = 25,$$

$$S_{xx} = \sum_{i=1}^{8} x_i^2 - \frac{\left(\sum_{i=1}^{8} x_i\right)^2}{8} = 92 - \frac{20^2}{8} = 42,$$

$$S_{xy} = \sum_{i=1}^{8} x_i y_i - \frac{\left(\sum_{i=1}^{8} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{8} = 25 - \frac{(20)(37)}{8} = -67.5.$$

Linear regression

Estimates for the slope, intercept, and simple linear regression model using the least squares method are obtained

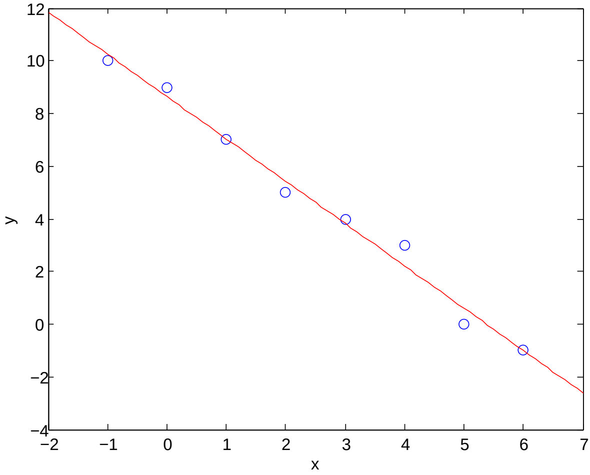
$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-67.5}{42} = -1.6071,$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = 4.625 - (-1.6071)2.5 = 8.6429,$$

$$\hat{y} = 8.6429 - 1.6071x.$$

The model and experimental points are plotted in the graph in the figure.

Linear regression



Linear regression

```
xi=[-1 0 1 2 3 4 5 6];
yi=[10 9 7 5 4 3 0 -1];
n=length(xi);
sx=sum(xi); sy=sum(yi);
xm=mean(xi); ym=mean(yi);
sx2=sum(xi.^2); sy2=sum(yi.^2);
sxy=sum(xi.*yi);
Sxx=sx2-(sx^2)/n; Syy=sxy-sx*sy/n;
Beta1=Syy/Sxx; Beta0=ym-Beta1*xm;
x=-2:0.1:7;
yhat=Beta0+Beta1*x;
plot(xi,yi,'or',x,yhat)
xlabel('x'),ylabel('y')
```

Linear regression

Polynomial regression

Polynomial regression is an approximation of the values of observations of functions in the form of a polynomial n -th degree

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n + \epsilon. \tag{10}$$

in the sense of the least squares method, where the estimates of the coefficients are obtained by minimizing the sum of the squares of the deviations of the observations from the resulting polynomial. Polynomial regression is a linear regression because the resulting polynomial is expressed as a linear combination of coefficients and simple functions

$$Y = \sum_{j=0}^{n} \beta_j x^j. \tag{11}$$

Linear regression

- **Example:** The function is given by a table. Approximate the points by the method least squares in the form $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$.

x_i	-3	0	2	4
y_i	3	1	1	3

Solution: The sum of the squares of the deviations of the observations from the polynomial of the second degree is given

$$L = \sum_{i=1}^{n} \epsilon_i^2 = (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2.$$

Linear regression

Estimates $\hat{\beta}_0$, $\hat{\beta}_1$ $\hat{\beta}_2$, must meet the condition of the minimum sum of squares of deviations

$$\left.\frac{\delta L}{\delta \beta_0}\right|_{\hat{\beta}_0 \hat{\beta}_1 \hat{\beta}_1} = 2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2)(-1) = 0$$

$$\left.\frac{\delta L}{\delta \beta_1}\right|_{\hat{\beta}_0 \hat{\beta}_1 \hat{\beta}_2} = 2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2)(-x_i) = 0.$$

$$\left.\frac{\delta L}{\delta \beta_2}\right|_{\hat{\beta}_0 \hat{\beta}_1 \hat{\beta}_2} = 2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2)(-x_i^2) = 0. \qquad (12)$$

By simplification we get

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i + \hat{\beta}_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i^2 + \hat{\beta}_1 \sum_{i=1}^{n} x_i^3 + \hat{\beta}_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i \qquad (13)$$

Linear regression

By adding the sums we get

$$
\begin{aligned}
4\hat{\beta}_0 + 3\hat{\beta}_1 + 29\hat{\beta}_2 &= 8 \\
3\hat{\beta}_0 + 29\hat{\beta}_1 45 + \hat{\beta}_2 &= 5 \\
29\hat{\beta}_0 + 45\hat{\beta}_1 + 353\hat{\beta}_2 &= 79 \qquad (14)
\end{aligned}
$$

By solving the system of normal equations, we obtain estimates of regression coefficients $\hat{\beta}_0$, $\hat{\beta}_1$ $\hat{\beta}_2$. The resulting searched polynomial then has the form (see figure)

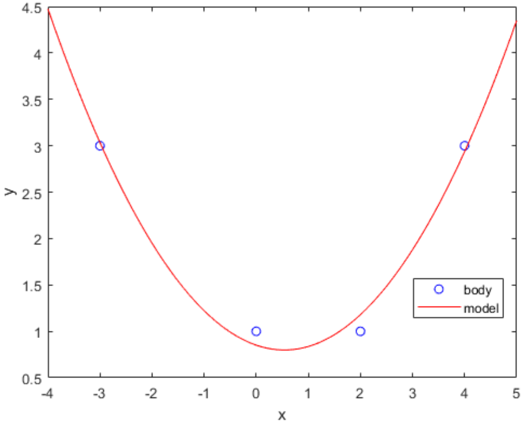
$$\hat{y} = 0.8505 - 0.1925x + 0.1785x^2.$$

In MatLab for polynomial regression, we can use the function `polyfit`, which finds the coefficients of the polynomial $\beta(x)$ of degree n, which fits the data in the sense of the least squares method. The result is a line vector of length $n+1$ containing the coefficients of the polynomial with decreasing power.

$$\beta(x) = \beta_1 x^n + \beta_2 x^{n-1} + \ldots + \beta_{n+1}.$$

Function `polyval` returns the values of polynomial of degree n calculated in points x.

Linear regression



Linear regression

```
xi=[-3 0 2 4];                 % observations x
yi=[3 1 1 3];                  % observations y
Beta=polyfit(xi,yi,2)          % 2 degree polynomial parameter estimate
x=-4:0.1:5;                    % range x for which the value ...
    function will be calculated
yhat=polyval(Beta,x);          % evaluation of polynomial
plot(xi,yi,'ob',x,yhat,'r')    % plot of observations and model
xlabel('x'),ylabel('y')        % labels
legend('body', 'model','location','best')
```

Linear regression

- total sum of squares

$$S_T = \sum (y_i - \overline{y})^2$$

$$S_T = S_A + S_R$$

- sum of squares regression

$$S_A = \sum (\hat{y}_i - \overline{y})^2$$

- sum of squares error

$$S_R = \sum (y_i - \hat{y}_i)^2$$

- coefficient of determination

$$R^2 = \frac{S_A}{S_T} = 1 - \frac{S_R}{S_T}$$

Linear regression

```
clear,clc,close all

xi=[-3 0 2 4];              % observations x
yi=[3 1 1 3];               % observations y
plot(xi,yi,'o')             % scatter plot
xlabel('x'),ylabel('y')
Beta=polyfit(xi,yi,2)       % 2 degree polynomial parameter estimate
x=-4:0.1:5;                 % range x for which the value ...
    function will be calculated
yhat=polyval(Beta,x);       % evaluation of polynomial
hold on
plot(x,yhat)                % plot of the model

%% coefficient of determination
ST=sum((yi-mean(yi)).^2)  % total sum of squares
yhatxi=polyval(Beta,xi);  % calculation of the functional ...
    values of the model at points xi
SA=sum((yhatxi-mean(yi)).^2) % sum of squares regressino
SR=sum((yi-yhatxi).^2)      % sum of squares error
R2=SA/ST                    % coefficient of determination
```

Lineární regrese

```
Beta =
    0.1785   -0.1925    0.8505
ST =
     4
SA =
    3.9402
SR =
    0.0598
R2 =
    0.9851
```

Linear regression

Graphical user interface

Another way is to use an interactive graphical user interface (*Basic Fitting GUI*, see figure. The GUI contains three panels. The first contains model selection and plotting options. The second panel shows model coefficient estimates and residual norm. Results can be exported The last column is used to calculate and export values (see the following example).

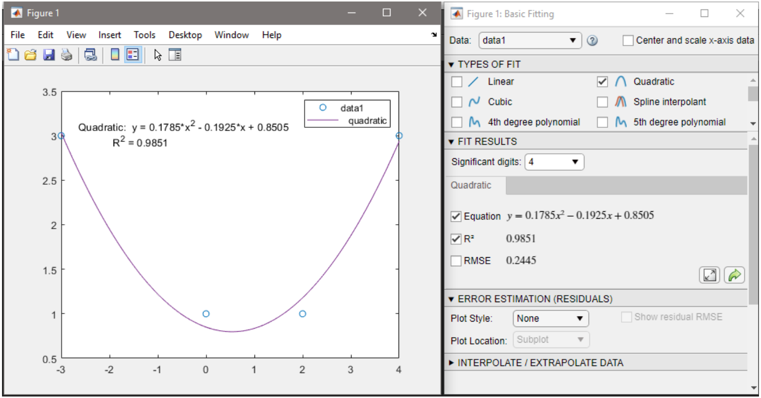
- **Example:** The function is given by a table. Approximate the points by the method least squares in the form $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$.

x_i	-3	0	2	4
y_i	3	1	1	3

```
>> xi=[-3 0 2 4];
>> yi=[3 1 1 3];
>> plot(xi,yi,'o')
```

Figure: Tools → Basic Fitting

Linear regression



List of functions

Command	Description
» polyfit(x,y,n)	polynomial curve fitting of degree n
» polyval(p,x)	returns the values of polynomial p at point in x

Examples for practice

Examples for practice

1. **Example:** In the data file `data.xls` the results of the electromagnetic shielding effectiveness [dB] in two frequencies (1.5 GHz and 0.6 GHz) depending on the number of layers of samples was measured. Describe this dependence by a linear model obtained using the least square method, get parameter estimates and plot the model for both cases. Create m-file (script) named `se.m`.

Solution