Practice 7
Least squares method

Contents

1 Least squares method **1**

2 Applications of the least squares method **3**

2.1 Line fitting 3

2.2 Curve fitting 4

1 Least squares method

The method of **least squares** is a standard approach to the approximate solution of overdetermined systems, i.e. sets of equations in which there are more equations than unknowns. In these cases, normally, there is not solution but it is desired to obtain a vector such that $\mathsf{A}\vec{x}$ be as close to $\vec{b}$ as posible.

- A *solution by least squares* of a system of linear equations $\mathsf{A}\vec{x} = \vec{b}$ is a vector $\vec{x}_M$ in $\mathbb{R}^n$ such that $\|\vec{b} - \mathsf{A}\vec{x}_M\| \leq \|\vec{b} - \mathsf{A}\vec{x}\|$ for all $\vec{x} \in \mathbb{R}^n$.

  Theorem 1 (Best approximation). Let W be a linear subspace of $\mathbb{R}^n$ and set $\vec{x} \in \mathbb{R}^n$. The vector $Proj_W(\vec{x})$ is the vector of W which is closer to $\vec{x}$ than the other vectors in W in the following sense:

$$\|\vec{x} - Proj_W(\vec{x})\| \leq \|\vec{x} - \vec{v}\| \qquad \text{for all } \vec{v} \in W.$$

  If we apply this theorem to the subspace $Col(\mathsf{A})$ and we call $\vec{b}_p$ to the orthogonal projection of $\vec{b}$ over $Col(\mathsf{A})$, we know that b_p is the closest to b vector of $Col(\mathsf{A})$. Then, there exists a vector $\vec{x}_M$ that will be a solution of the system $\mathsf{A}\vec{x} = \mathsf{b}_p$. Therefore we have $\mathsf{A}\vec{x}_M = \vec{b}_p$ and the vector $\vec{x}_M$ is a solution by least squares. In fact, $\vec{x}_M$ is the vector of coefficients of a linear combination of $\vec{b}_p$ in terms of the columns of A.

- As we saw in the previous practice (see also Section IV of Lesson 1):

$$\mathsf{A}^t\mathsf{A}\vec{x}_M = \mathsf{A}^t\vec{b}.$$

  This is called **system of normal equations** for $\vec{x}_M$.

- If the columns of A are linearly independent, the matrix $\mathsf{A}^t\mathsf{A}$ is invertible and the solution by least squares $\vec{x}_M = (\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\vec{b}$ **is unique**.

- The *residual error* is $\|\vec{b} - \mathsf{A}\vec{x}_M\|$.

Example 1. We shall determine the solutions by least squares of $\mathsf{A}\vec{x} = \vec{b}$, where

$$\mathsf{A} = \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}.$$

We must solve $\mathsf{A}^t\mathsf{A}\vec{x}_M = \mathsf{A}^t\vec{b}$.

$$\mathsf{A}^t\mathsf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix}, \qquad \mathsf{A}^t\vec{b} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}.$$

The system

$$\begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix} \begin{bmatrix} x_{1M} \\ x_{2M} \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

has, as unique solution,

$$\begin{bmatrix} x_{1M} \\ x_{2M} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The residual error is

$$\|\vec{b} - \mathsf{A}\vec{x}_M\| = \left\| \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \sqrt{6}$$

With Scilab:

The command `x=A\b` computes a least squares solution of the system $\mathsf{A}\vec{x} = \vec{b}$. If there is only one least squares solution, this command computes it. Otherwise, it computes the least squares solution with, at most, $\operatorname{rank}(\mathsf{A})$ nonzero components. In the last case, the provided solution is not, in general, the one with minimum norm. If we want to compute the least squares solution with minimum norm we can use the command `x=lsq(A,b)`.

```
-->A=[1 3;1 -1;1 1];b=[5;1;0];
-->lsq(A,b)
 ans  =


    1.
    1
```

You can check that the command `x=A\b` gives the same solution.

2 Applications of the least squares method

In Science and Engineering, the experiments produce a set of data (x_1, y_1), $(x_2, y_2), \dots,$ (x_n, y_n) with different x-coordinates. The problem consists of finding a function $y = f(x)$ relating the data in the best possible way. To find this solution implies to solve a typical problem by least squares method.

2.1 Line fitting

The simplest relation between two variables x and y is the linear equation $y = \beta_0 + \beta_1 x$.

Often, when we represent graphically the set of experimental data it seems that they are (or they should be) aligned. We want to determine the parameters β_0 y β_1 making the line as "close" to the points as possible.

The *regression line* is the line $y = \beta_0 + \beta_1 x$ minimizing the residual error.

If the points be on the line, the parameters β_0 and β_1 would satisfy the equations

$$\beta_0 + \beta_1 x_i = y_i \qquad \text{para } i \in \{1, 2, \dots, n\}.$$

In matrix form:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}.$$

We call *design matrix* to

$$\mathsf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{bmatrix},$$

vector of parameters to the vector

$$\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

and *observation vector* to

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$$

The *residual vector* is

$$\vec{\varepsilon} = \vec{y} - \mathsf{X}\vec{\beta}.$$

To determine the vector of parameters it is enough to "solve" the system $\mathsf{X}\vec{\beta} = \mathsf{y}$ by the least squares method.

Example 2. We shall find the line $y = \beta_0 + \beta_1 x$ fitting the points $(0,1)$, $(1,1)$, $(2,2)$ and $(3,2)$. First we compute the design matrix X and we compute $\mathsf{X}^t\mathsf{X}$.

$$\mathsf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \qquad \mathsf{X}^t\mathsf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix}.$$

We obtain

$$\mathsf{X}^t\vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 11 \end{bmatrix}$$

We solve

$$\begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 6 \\ 11 \end{bmatrix}, \quad \text{whose solution is} \quad \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9/10 \\ 2/5 \end{bmatrix}$$

Therefore, the regression line is

$$y = \frac{9}{10} + \frac{2}{5}x.$$

With Scilab:

```
-->X=[1 0;1 1;1 2;1 3],y=[1;1;2;2];
-->X1=X'*X,  y1=X'*y
 X1  =
    4.    6.
    6.    14.
 y1  =
    6.
    11.
-->b=X1\y1
 b  =
    0.9
    0.4
```

We compute the norm of the residual vector $\|\vec{y} - \mathsf{X}\vec{\beta}\|$:

```
-->norm(y-X*b)
 ans  =
    0.4472136
```

2.2 Curve fitting

If we represent graphically the set of data and we observe that they are not (or they do not should be) on a line, we can find the curve that fits better the points:

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x),$$

where f_0, $f_1, \ldots$, f_n are known functions and β_0, $\beta_1, \ldots$, β_n the parameters we need determine. Doing the same as before, we must find a vector of parameters such that the norm of the residual vector is the minimum one. In this case, the matrix expression of the system is:

$$\begin{bmatrix} f_0(x_1) & f_1(x_1) & \ldots & f_k(x_1) \\ f_0(x_2) & f_1(x_2) & \ldots & f_k(x_1) \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ f_0(x_n) & f_1(x_n) & \ldots & f_k(x_n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_k \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ \ldots \\ y_n \end{bmatrix}.$$

The coefficient matrix is the design matrix.

Example 3. With the data given in the following table

1	2	3	4	5
1.8	2.7	3.4	3.8	3.9

we determine the vector of parameters, the residual vector and the regression curve associated to the function $y = \beta_0 + \beta_1 x + \beta_2 x^2$.

1. We must "solve" by the least square method the system of linear equations $\vec{y} = \mathsf{X}\vec{\beta}$:
$$\begin{bmatrix} 1.8 \\ 2.7 \\ 3.4 \\ 3.8 \\ 3.9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$
2. We compute
$$\mathsf{X}^t\mathsf{X} = \begin{bmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix}, \qquad \mathsf{X}^t\vec{y} = \begin{bmatrix} 15.6 \\ 52.1 \\ 201.5 \end{bmatrix}.$$
3. Solving this system (unique solution) $\mathsf{X}^t\mathsf{X}\vec{\beta} = \mathsf{X}^t\vec{y}$ we obtain the vector of parameters minimizing the norm of the residual error:
$$\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0.58 \\ 1.34 \\ -0.136 \end{bmatrix}.$$

Therefore, the best fitting parabola is $y = 0.58 + 1.34x - 0.136x^2$.

If we compute $\vec{\varepsilon} = \vec{y} - \mathsf{X}\vec{\beta}$, we obtain the residual vector

$$\begin{bmatrix} 0.114 \\ -0.026 \\ 0.008 \\ 0.014 \\ 0.008 \end{bmatrix}.$$

With Scilab:

We introduce the matrices X, $\vec{y}$:

```
-->X=[1 1 1;1 2 4;1 3 9;1 4 16;1 5 25], y=[1.8;2.7;3.4;3.8;3.9];
```

and, with the command \, we obtain the solution by least squares method:

```
-->b=X1\y1
 b  =
    0.58
    1.3442857
  - 0.1357143
```

The best fitting parabola has the parameters corresponding to b and its residual error is:

```
-->ER=norm(y-X*b)
 ER =
      0.0338062
```

Example 4. With the data of the above example we compute the "best" parameters corresponding to a function of the type $y = \beta_1 \cos(\pi x/3) + \beta_2 \sin(\pi x/3)$.

$$\mathsf{X} = \begin{bmatrix} \cos\frac{\pi}{3} & \sin\frac{\pi}{3} \\ \cos\frac{2\pi}{3} & \sin\frac{2\pi}{3} \\ \cos\frac{3\pi}{3} & \sin\frac{3\pi}{3} \\ \cos\frac{4\pi}{3} & \sin\frac{4\pi}{3} \\ \cos\frac{5\pi}{3} & \sin\frac{5\pi}{3} \end{bmatrix}$$

$$\mathsf{X}^t\mathsf{X} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \qquad \mathsf{X}^t\vec{y} = \begin{bmatrix} -3.8 \\ -2.8 \end{bmatrix}$$

We solve the system and we obtain the vector of parameters:

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -3.8 \\ -2.8 \end{bmatrix}, \quad \text{whose solution is} \quad \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -1.9 \\ -0.9 \end{bmatrix}.$$

Therefore, the required function is $y = -1.9\cos(\pi x/3) - 0.9\sin(\pi x/3)$, and the residual vector is

$$\vec{\varepsilon} = \vec{y} - \mathsf{X}\vec{\beta} = \begin{bmatrix} 3.5 \\ 2.5 \\ 1.5 \\ 2 \\ 4 \end{bmatrix}.$$

Observe that the residual vector was better when we used the first type of function.

With Scilab:

We introduce the design matrix X (the vector y is already introduced):

```
-->X=[cos(%pi/3),   sin(%pi/3)
-->   cos(%pi*2/3), sin(%pi*2/3)
-->   cos(%pi*3/3), sin(%pi*3/3)
-->   cos(%pi*4/3), sin(%pi*4/3)
-->   cos(%pi*5/3), sin(%pi*5/3)]
 X  =
    0.5     0.8660254
  - 0.5     0.8660254
  - 1.      1.225D-16
  - 0.5   - 0.8660254
    0.5   - 0.8660254
```

We can obtain the parameters directely:

```
-->b=X\y
 b  =
  - 1.9
  - 0.9237604
```

The result has not been the same due to rounding errors with trigonometric functions.