How to identify a linear model with the least square method.

4TTV313U - introduction to simulation of dynamic process

2023

We have a physical system. We suppose the dynamic of that system follows a model that is linear to it's parameters. That means there exists a list of functions $\phi_0(t)$, $\phi_1(t)$, $\phi_2(t)$, \cdots , $\phi_k(t)$ depending of the time t and a list of reals $\alpha_1, \alpha_2, \cdots, \alpha_k$ such that we have the following equality :

$$\phi_0(t) = \alpha_1 \times \phi_1(t) + \alpha_2 \times \phi_2(t) + \alpha_3 \times \phi_3(t) + \dots + \alpha_k \times \phi_k(t). \tag{1}$$

Suppose now, we can make some experiments to obtain, for different times t the values of $\phi_0(t)$, $\phi_1(t), \phi_2(t), \dots, \phi_k(t)$.

In this document, we learn to compute, from our data, the parameters $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ that fit our data. We means, by fitting our data, that the error we make between our model and the experiments are the smallest possible according to some criteria we will define later.

For example, suppose that w_1 , w_2 , w_3 are linear velocities. A linear model according to the parameters α_1 , α_2 and α_3 could be :

$$\sqrt{w_3} \cdot \exp t = \alpha_1 \times t \cdot w_1^2 + \alpha_2 \times w_1 \cdot w_2 + \alpha_3 \times \cos(3 \cdot w_2 \cdot t).$$

In that example, we can define $\phi_0(t) = \sqrt{w_3} \cdot \exp t$, $\phi_1(t) = t \cdot w_1^2$, $\phi_2(t) = w_1 \cdot w_2$ and $\phi_3(t) = \cos(3 \cdot w_2 \cdot t)$ to obtain

$$\phi_0(t) = \alpha_1 \times \phi_1(t) + \alpha_2 \times \phi_2(t) + \alpha_3 \times \phi_3(t)$$

Suppose, we obtain the data set of the Table 1:

id	No. exp.	date	t	w_1	w_2	w_3
0	1	01/01	0	0.1	0.5	0.3658
1	1	01/01	1	0.4	0.2	1.1444
2	1	01/01	2	0.8	0.6	10.1610
3	1	01/01	3	0.5	0.4	0.3676
4	2	02/01	0	0.2	0.6	4.8159
5	2	02/01	1	0.5	0.3	4.0255
6	2	02/01	2	0.7	0.7	7.7276
$\overline{7}$	2	02/01	3	0.4	0.3	0.1518

Table 1: The measures of two experiments.

We can compute the different values of ϕ_0 , ϕ_1 , ϕ_2 and ϕ_3 for our data set :

id	$\phi_0(t) = \sqrt{w_3} \exp t$	$\phi_1(t) = tw_1^2$	$\phi_2(t) = w_1 w_2$	$\phi_3(t) = \cos(3w_2t)$
0	0.60481402	0.	0.05	1.
1	2.90792637	0.16	0.08	0.82533561
2	23.55359426	1.28	0.48	-0.89675842
3	12.17786617	0.75	0.2	-0.89675842
4	2.19451589	0.	0.12	1.
5	5.45386517	0.25	0.15	0.62160997
6	20.54051275	0.98	0.49	-0.49026082
7	7.82563038	0.48	0.12	-0.90407214

Table 2: The values of ϕ_0 , ϕ_1 , ϕ_2 and ϕ_3

We want now to estimate the better parameter $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$ in order to fit the model to the data set.

The idea is to write equation 1 by using the matrix formalism :

$$\begin{pmatrix} \phi_0(t) \end{pmatrix} = \begin{pmatrix} \phi_1(t) & \phi_2(t) & \cdots & \phi_k(t) \end{pmatrix} \times \begin{pmatrix} \alpha_1 \\ \alpha_1 \\ \cdots \\ \alpha_k \end{pmatrix}.$$

In our example, we obtain :

$$\left(\sqrt{w_3} \cdot \exp t \right) = \left(t \cdot w_1^2 \quad w_1 \cdot w_2 \quad \cos(3 \cdot w_2 \cdot t) \right) \times \left(\begin{array}{c} \alpha_1 \\ \alpha_1 \\ \alpha_3 \end{array} \right)$$

If we define by $b(t) = (\sqrt{w_3} \cdot \exp t)$ and by $a(t) = (t \cdot w_1^2 - w_1 \cdot w_2 - \cos(3 \cdot w_2 \cdot t))$, we obtain the following equation:

$$b(t) = a(t) \times \alpha. \tag{2}$$

In fact [b(t), a(t)] is one row of Table 2. So we can concatenate each row of our data set to obtain the following matrix definitions :

$$B = \begin{pmatrix} b(t_1) \\ b(t_2) \\ b(t_3) \\ \cdots \\ b(t_N) \end{pmatrix} \qquad A = \begin{pmatrix} a(t_1) \\ a(t_2) \\ a(t_3) \\ \cdots \\ a(t_N) \end{pmatrix}.$$

For example, in the previous example, the matrix B is the first column of the Table 2, and the matrix A contains the 3 last columns :

$$B = \begin{pmatrix} 0.60481402\\ 2.90792637\\ 23.55359426\\ 12.17786617\\ 2.19451589\\ 5.45386517\\ 20.54051275\\ 7.82563038 \end{pmatrix} A = \begin{pmatrix} 0. & 0.05 & 1. \\ 0.16 & 0.08 & 0.82533561\\ 1.28 & 0.48 & -0.89675842\\ 0.75 & 0.2 & -0.89675842\\ 0. & 0.12 & 1. \\ 0.25 & 0.15 & 0.62160997\\ 0.98 & 0.49 & -0.49026082\\ 0.48 & 0.12 & -0.90407214 \end{pmatrix}$$

Now, from equation 2 we obtain the following general identity :

$$B = A \times \alpha.$$

In fact, the model is not perfect, and we made some errors during the experiments. So we have

$$B = A \times \alpha + \epsilon. \tag{3}$$

where ϵ is the error vector.

:

So now, the problem consists in finding the parameter α that minimizes the error vector ϵ . More precisely, we want to find the parameter α that minimizes the square error $||\epsilon||^2 = \sum_i \epsilon_i^2$. We will prove, at the end of that document, the solution of that problem is :

$$\alpha = \left(A^t \cdot A\right)^{-1} \cdot A^t \cdot B. \tag{4}$$

where A^t in the transposition of the matrix A.

For example, in our problem :

$$A^{T} = \begin{pmatrix} 0. & 0.16 & 1.28 & 0.75 & 0. & 0.25 & 0.98 & 0.48 \\ 0.05 & 0.08 & 0.48 & 0.2 & 0.12 & 0.15 & 0.49 & 0.12 \\ 1. & 0.825 & -0.896 & -0.896 & 1. & 0.621 & -0.490 & -0.904 \end{pmatrix},$$

$$A^{t} \cdot A = \begin{pmatrix} 3.4798 & 1.3525 & -2.44737363 \\ 1.3525 & 0.5707 & -0.62924384 \\ 2.44737363 & -0.62924384 & 5.73363126 \end{pmatrix}$$
$$(A^{t} \cdot A)^{-1} = \begin{pmatrix} 10.73850093 & -23.20290054 & 2.03725027 \\ -23.20290054 & 52.12843577 & -4.18315528 \\ 2.03725027 & -4.18315528 & 0.58491517 \end{pmatrix}$$

and

$$\alpha = (9.90646281 \ 21.72126782 \ -0.46834528)^t$$

The error me make with the model is :

$$\epsilon^{t} = (A \cdot \alpha - B)^{t} = (0.012 \quad 0.028 \quad -0.027 \quad 0.016 \quad -0.056 \quad -0.010 \quad 0.040 \quad -0.040)$$

and the square root error is : $||\epsilon|| = 0.0924$.

We will prove now that, the solution of our problem is :

$$\alpha = \left(A^t \cdot A\right)^{-1} \cdot A^t \cdot B.$$

We want to minimize the squared error :

$$L = ||\epsilon||^2 = \sum_i \epsilon_i^2.$$

As $L \ge 0$ there exists a minimum. This minimum is reached when, the derivative of L according to the parameters α_i are equal to 0 :

$$\frac{\partial L}{\partial \alpha_i} = 0$$

Let $\nabla_{\alpha} L$ be the gradient of L defined by :

$$\nabla_{\alpha}L = \begin{pmatrix} \frac{\partial L}{\partial \alpha_1} \\ \frac{\partial L}{\partial \alpha_2} \\ \cdots \\ \frac{\partial L}{\partial \alpha_k} \end{pmatrix}$$

If L is minimum then its gradient is null :

$$\exists \alpha_{\min}, \forall \alpha, \ L(\alpha) \ge L(\alpha_{\min}) \qquad \Longrightarrow \qquad \nabla_{\alpha} L(\alpha_{\min}) = 0.$$

We will now determine α_{\min} the solution of our problem by computing $\nabla_{\alpha}L$:

$$\nabla_{\alpha}L = \nabla_{\alpha}\epsilon^{t}\epsilon = \nabla_{\alpha}(B - A\alpha)^{t}(B - A\alpha) = \nabla_{\alpha}(B^{t} - \alpha^{t}A^{t})(B - A\alpha);$$

= $\nabla_{\alpha}B^{t}B - \nabla_{\alpha}B^{t}A\alpha - \nabla_{\alpha}\alpha^{t}A^{t}B + \nabla_{\alpha}\alpha^{t}A^{t}A\alpha.$

As $B^t B$ is a constant, we have $\nabla_{\alpha} B^t B = 0$.

Let's define the vector e_i in \mathbb{R}^k as a vector where each row contains a zero, except for row *i*, which contains a 1. Now, we can compute $\frac{\partial B^t A \alpha}{\partial \alpha_i} = B^t A e_i$ and deduce that : $\nabla_{\alpha} B^t A \alpha = (B^t A)^t = A^t B$. Similarly, we can express, $\frac{\partial \alpha^t A^t B}{\partial \alpha_i} = e_i^t A^t B$ and obtain : $\nabla_{\alpha} \alpha^t A^t B = A^t B$. Finally, by using the Leibniz rule,

$$\nabla_{\alpha}\alpha^{t}A^{t}A\alpha = \nabla_{\beta}\beta^{t}A^{t}A\alpha + \nabla_{\beta}\alpha^{t}A^{t}A\beta \mid_{\beta \to \alpha} = A^{t}A\alpha + (\alpha^{t}A^{t}A)^{t} = 2A^{t}A\alpha,$$

and we obtain that

$$\nabla_{\alpha}L(\alpha_{\min}) = -2A^{t}B + 2A^{t}A\alpha_{\min} = 0$$

We conclude that :

$$\alpha_{\min} = \left(A^t A \right)^{-1} A^t B.$$