Computational Science:
Computational Methods in Engineering

Linear Regression

https://empossible.net/academics/emp4301_5301/

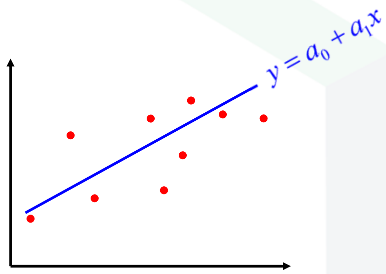
Outline

- Linear regression – algebraic approach
- Linear regression – matrix approach
- Visualizing least squares
- Implementing linear regression

- MATLAB Implementation
https://empossible.net/academics/emp4301_5301/

Goal of Linear Regression

The goal of linear regression is to fit a straight line to a set of measured data that has noise.



$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_M, y_M)$$
$$\downarrow$$
$$y = a_0 + a_1 x$$

EMPossible

3

Linear Regression
(Algebraic Approach)

https://empossible.net/academics/emp4301_5301/

4

Statement of Problem

Given a set of measured data points: (x_1,y_1), (x_2,y_2), …, (x_M,y_M), the equation of the line is written for each point.

$$
\begin{aligned}
y_1 &= a_0 + a_1 x_1 + e_1 \\
y_2 &= a_0 + a_1 x_2 + e_2 \\
&\vdots \\
y_M &= a_0 + a_1 x_M + e_M
\end{aligned}
$$

To be completely correct, an error term e is introduced called the *residual*.

It is desired to determine values of a_0 and a_1 such that the residual terms e_m are as small as possible.

EMPossible

5

Criteria for "Best Fit"

A single quantity is needed that measures how "good" the line fits the set of data.

Guess #1 – Sum of Residuals

$$E = \sum_{m=1}^{M} e_m$$

This does not work because negative and positive residuals can cancel and mislead the overall criteria to think there is no error.

Guess #2 – Sum of Magnitude of Residuals

$$E = \sum_{m=1}^{M} \left| e_m \right|$$

This does not work because it does not lead to a unique best fit.

Guess #3 – Sum of Squares of Residuals

$$E = \sum_{m=1}^{M} e_m^2$$

This works and leads to a unique solution.

EMPossible

6

Equation for Criterion

The line equation for the mth sample is

$$y_m = a_0 + a_1 x_m + e_m$$

Solving this for the residual e_m gives

$$e_m = y_m - \left(a_0 + a_1 x_m\right)$$

This is the fit value of y at point x_m.

This is the measured value of y.

From this, the error criterion is written as

$$E = \sum_{m=1}^{M} e_m^2 = \sum_{m=1}^{M} \left(y_{\text{measured},m} - y_{\text{line},m}\right)^2 = \sum_{m=1}^{M} \left(y_m - a_0 - a_1 x_m\right)^2$$

EMPossible 7

Least-Squares Fit

It is desired to minimize the error criterion E.

Minimums can be identified where the first-order derivative is zero.

$$\frac{\partial E}{\partial a_0} = 0 \quad \text{and} \quad \frac{\partial E}{\partial a_1} = 0$$

Values of a_0 and a_1 are sought that satisfy these equations.

This approach is solving the problem by least-squares (i.e. minimizing the squares of the residuals).

EMPossible 8

Derivation of Least-Squares Fit

Step 1 – Differentiate E with respect to each of the unknowns.

$$\begin{aligned}\frac{\partial E}{\partial a_0} &= \frac{\partial}{\partial a_0}\sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right)^2 \\ &= \sum_{m=1}^{M}\frac{\partial}{\partial a_0}\left(y_m - a_0 - a_1 x_m\right)^2 \\ &= \sum_{m=1}^{M}2\left(y_m - a_0 - a_1 x_m\right)(-1) \\ &= -2\sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right)\end{aligned}$$

$$\begin{aligned}\frac{\partial E}{\partial a_1} &= \frac{\partial}{\partial a_1}\sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right)^2 \\ &= \sum_{m=1}^{M}\frac{\partial}{\partial a_1}\left(y_m - a_0 - a_1 x_m\right)^2 \\ &= \sum_{m=1}^{M}2\left(y_m - a_0 - a_1 x_m\right)(-x_m) \\ &= -2\sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right)x_m\end{aligned}$$

EMPossible 9

Derivation of Least-Squares Fit

Step 2 – Set the derivatives to zero to locate the minimum of E.

$$\begin{aligned}0 &= \frac{\partial E}{\partial a_0} \\ &= -2\sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right) \\ &= \sum_{m=1}^{M}y_m - \sum_{m=1}^{M}a_0 - \sum_{m=1}^{M}a_1 x_m \\ &= \sum_{m=1}^{M}y_m - Ma_0 - \sum_{m=1}^{M}a_1 x_m\end{aligned}$$

$$\begin{aligned}0 &= \frac{\partial E}{\partial a_1} \\ &= -2\sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right)x_m \\ &= \sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right)x_m \\ &= \sum_{m=1}^{M}y_m x_m - a_0\sum_{m=1}^{M}x_m - \sum_{m=1}^{M}a_1 x_m^2\end{aligned}$$

EMPossible 10

Derivation of Least-Squares Fit

Step 3 – Write these as two simultaneous equations. These are called the *normal equations*.

$$\begin{aligned}0 &= \frac{\partial E}{\partial a_0} \\ &= -2\sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right) \\ &= \sum_{m=1}^{M} y_m - \sum_{m=1}^{M} a_0 - \sum_{m=1}^{M} a_1 x_m \\ &= \sum_{m=1}^{M} y_m - M a_0 - a_1 \sum_{m=1}^{M} x_m\end{aligned}$$

$$\downarrow$$

$$M a_0 + a_1 \sum_{m=1}^{M} x_m = \sum_{m=1}^{M} y_m$$

$$\begin{aligned}0 &= \frac{\partial E}{\partial a_1} \\ &= -2\sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right) x_m \\ &= \sum_{m=1}^{M}\left(y_m - a_0 - a_1 x_m\right) x_m \\ &= \sum_{m=1}^{M} y_m x_m - a_0 \sum_{m=1}^{M} x_m - \sum_{m=1}^{M} a_1 x_m^2\end{aligned}$$

$$\downarrow$$

$$a_0 \sum_{m=1}^{M} x_m + a_1 \sum_{m=1}^{M} x_m^2 = \sum_{m=1}^{M} y_m x_m$$

EMPossible

11

Derivation of Least-Squares Fit

Step 4 – The normal equations are solved simultaneously and the solution is

$$a_0 = y_{\text{avg}} - a_1 x_{\text{avg}}$$

$$x_{\text{avg}} = \frac{1}{M}\sum_{m=1}^{M} x_m$$

$$y_{\text{avg}} = \frac{1}{M}\sum_{m=1}^{M} y_m$$

$$a_1 = \frac{M\sum_{m=1}^{M} x_m y_m - \sum_{m=1}^{M} x_m \sum_{m=1}^{M} y_m}{M\sum_{m=1}^{M} x_m^2 - \left(\sum_{m=1}^{M} x_m\right)^2}$$

Yikes!
There has to be an easier way!

EMPossible

12

Linear Regression
(Matrix Approach)

https://empossible.net/academics/emp4301_5301/

Statement of Problem

It is desired to fit a set of M measured data points to a curve containing $N+1$ terms:

$$f = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_N z_N$$

$f \equiv$ measured value
$z_n \equiv$ parameters from which f is evaluated
$a_n \equiv$ coefficients for the curve fit

EMPossible

Formulation of Matrix Equation

Start by writing the function f for each of the M measurements. The residual terms are also incorporated.

$$
\begin{aligned}
f_1 &= a_0 z_{0,1} + a_1 z_{1,1} + \cdots + a_N z_{N,1} + e_1 \\
f_2 &= a_0 z_{0,2} + a_1 z_{1,2} + \cdots + a_N z_{N,2} + e_2 \\
&\vdots \\
f_M &= a_0 z_{0,M} + a_1 z_{1,M} + \cdots + a_N z_{N,M} + e_M
\end{aligned}
$$

This large set of equations is put into matrix form.

$$
\begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_{M-1} \\ f_M \end{bmatrix} = \begin{bmatrix} z_{0,1} & z_{1,1} & \cdots & z_{N,1} \\ z_{0,2} & z_{1,2} & \cdots & z_{N,2} \\ z_{0,3} & z_{1,3} & \vdots & z_{N,3} \\ \vdots & \vdots & & \vdots \\ z_{0,M-1} & z_{1,M-1} & \cdots & z_{N,M-1} \\ z_{0,M} & z_{1,M} & \cdots & z_{N,M} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_{M-1} \\ e_M \end{bmatrix} \Rightarrow \begin{matrix} [f] = [Z][a] + [e] \\ \text{or} \\ \mathbf{f} = \mathbf{Za} + \mathbf{e} \end{matrix}
$$

EMPossible

15

Formulation of Solution by Least-Squares (1 of 4)

Step 1 – Solve matrix equation for **e**.

$$\mathbf{f} = \mathbf{Za} + \mathbf{e} \quad \rightarrow \quad \mathbf{e} = \mathbf{f} - \mathbf{Za}$$

Step 2 – Calculate the error criterion E from **e**.

$$
E = \sum_{m=1}^{M} e_m^2 = \begin{bmatrix} e_1 & e_2 & e_3 & \cdots & e_{M-1} & e_M \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_{M-1} \\ e_M \end{bmatrix} = \mathbf{e}^T \mathbf{e}
$$

Step 3 – Substitute the equation for **e** from Step 1 into the equation for E from Step 2.

$$E = \mathbf{e}^T \mathbf{e} = (\mathbf{f} - \mathbf{Za})^T (\mathbf{f} - \mathbf{Za})$$

EMPossible

16

Formulation of Solution by Least-Squares (2 of 4)

Step 4 – The new matrix equation is algebraically manipulated as follows in order to make it easier to find its first-order derivative.

$$E = (\mathbf{f} - \mathbf{Za})^T (\mathbf{f} - \mathbf{Za}) \quad \text{original equation}$$

$$= (\mathbf{f}^T - \mathbf{a}^T \mathbf{Z}^T)(\mathbf{f} - \mathbf{Za}) \quad \text{distribute the transpose}$$

$$= \mathbf{f}^T \mathbf{f} - \underbrace{\mathbf{f}^T \mathbf{Za} - \mathbf{a}^T \mathbf{Z}^T \mathbf{f}}_{\text{These are scalars and transposes of each other so they are equal.}} + \mathbf{a}^T \mathbf{Z}^T \mathbf{Za} \quad \text{expand equation}$$

$$= \mathbf{f}^T \mathbf{f} - 2\mathbf{a}^T \mathbf{Z}^T \mathbf{f} + \mathbf{a}^T \mathbf{Z}^T \mathbf{Za} \quad \text{combine terms}$$

EMPossible 17

Formulation of Solution by Least-Squares (3 of 4)

Step 5 – Differentiate E with respect to $\mathbf{a}$.

It is desired to determine $\mathbf{a}$ that minimizes E.
This can be accomplished using the first-derivative rule.

$$E = \mathbf{f}^T \mathbf{f} - 2\mathbf{a}^T \mathbf{Z}^T \mathbf{f} + \mathbf{a}^T \mathbf{Z}^T \mathbf{Za}$$

$$\frac{\partial E}{\partial \mathbf{a}} = \frac{\partial}{\partial \mathbf{a}} \left(\mathbf{f}^T \mathbf{f} - 2\mathbf{a}^T \mathbf{Z}^T \mathbf{f} + \mathbf{a}^T \mathbf{Z}^T \mathbf{Za} \right) \quad \text{substitute in expression for } E$$

$$= \cancel{\frac{\partial}{\partial \mathbf{a}} \mathbf{f}^T \mathbf{f}} - 2\frac{\partial}{\partial \mathbf{a}} \mathbf{a}^T \mathbf{Z}^T \mathbf{f} + \frac{\partial}{\partial \mathbf{a}} \mathbf{a}^T \mathbf{Z}^T \mathbf{Za} \quad \mathbf{f} \text{ is not a function of } \mathbf{a}$$

$$= -2\mathbf{Z}^T \mathbf{f} + 2\mathbf{Z}^T \mathbf{Za} \quad \text{finish differentiation}$$

EMPossible 18

Formulation of Solution by Least-Squares (4 of 4)

Step 6 – Find the value of $\mathbf{a}$ that makes the derivative equal to zero.

$$\frac{\partial E}{\partial \mathbf{a}} = -2\mathbf{Z}^T\mathbf{f} + 2\mathbf{Z}^T\mathbf{Z}\mathbf{a} = 0$$

$$-2\mathbf{Z}^T\mathbf{f} + 2\mathbf{Z}^T\mathbf{Z}\mathbf{a} = 0$$

$$2\mathbf{Z}^T\mathbf{Z}\mathbf{a} = 2\mathbf{Z}^T\mathbf{f}$$

$$\mathbf{Z}^T\mathbf{Z}\mathbf{a} = \mathbf{Z}^T\mathbf{f}$$

Observe that this is the original equation $\mathbf{Za} = \mathbf{f}$ premultiplied by $\mathbf{Z}^T$.

$$\boxed{\mathbf{a} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{f}}$$

EMPossible 19

DO NOT SIMPLIFY FURTHER!

If the least-squares equation was simplified further, it would give

$$\begin{aligned}\mathbf{a} &= \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{f} \\ &= \mathbf{Z}^{-1}\underbrace{\left(\mathbf{Z}^T\right)^{-1}\mathbf{Z}^T}_{\mathbf{I}}\mathbf{f} \\ &= \mathbf{Z}^{-1}\mathbf{f}\end{aligned}$$

This is just the original equation again ($\mathbf{f} = \mathbf{Za}$) without the least-squares approach incorporated.

EMPossible 20

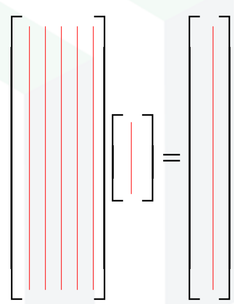
Visualizing Least-Squares

https://empossible.net/academics/emp4301_5301/

21

Visualizing Least-Squares (1 of 3)

Initially, a matrix equation was given that had more equations than unknowns.

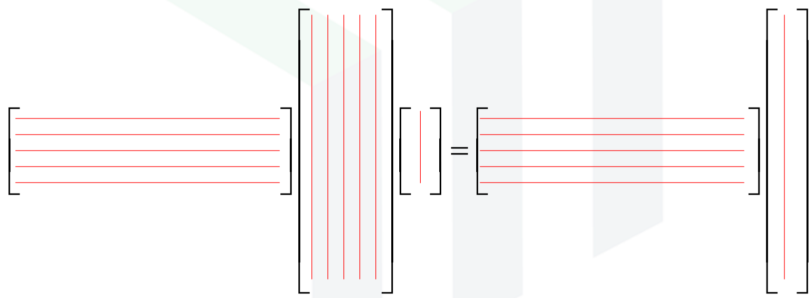


EMPossible

22

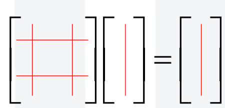
Visualizing Least-Squares (2 of 3)

The equation was premultiplied by the transpose of $\mathbf{A}$.



Visualizing Least-Squares (3 of 3)

The matrix equation reduced to the same number of equations as unknowns, which is solvable by many standard algorithms.



Implementing Linear Regression

https://empossible.net/academics/emp4301_5301/

Least-Squares Algorithm

Step 1 – Construct matrices. $\mathbf{Z}$ is essentially just a matrix of the coordinates of the data points. $\mathbf{f}$ is a column vector of the measurements.

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_{M-1} \\ f_M \end{bmatrix} \qquad \mathbf{Z} = \begin{bmatrix} z_{0,1} & z_{1,1} & \cdots & z_{N,1} \\ z_{0,2} & z_{1,2} & \cdots & z_{N,2} \\ z_{0,3} & z_{1,3} & \cdots & z_{N,3} \\ \vdots & \vdots & & \vdots \\ z_{0,M-1} & z_{1,M-1} & & z_{N,M-1} \\ z_{0,M} & z_{1,M} & & z_{N,M} \end{bmatrix}$$

Step 2 – Solve for the unknown coefficients $\mathbf{a}$.

$$\mathbf{a} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{f}$$

Step 3 – Extract the coefficients from $\mathbf{a}$.

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{bmatrix}$$

EMPossible

Least-Squares for Solving $\mathbf{Ax} = \mathbf{b}$

Suppose it is desired to solve $\mathbf{Ax} = \mathbf{b}$, but there exists more equations than there are unknowns.

This must be solved as a "best fit" because a perfect fit is impossible in the presence of noise.

Least-squares is implemented simply by premultiplying the matrix equation by $\mathbf{A}^T$.

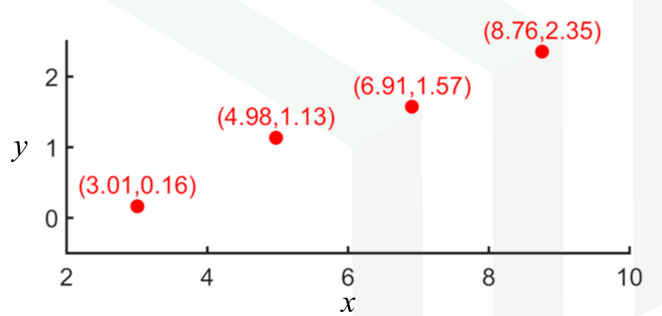
$$\mathbf{Ax} = \mathbf{b} \quad \rightarrow \quad \mathbf{A}'\mathbf{x} = \mathbf{b}'$$
$$\mathbf{A}' = \mathbf{A}^T\mathbf{A}$$
$$\mathbf{b}' = \mathbf{A}^T\mathbf{b}$$

EMPossible

27

Example 1 (1 of 3)

Fit a line to the following set of points.



$$y = mx + b$$

EMPossible

28

Example 1 (2 of 3)

Step 1 – Build matrices

$$\begin{aligned} y_1 &= mx_1 + b \\ y_2 &= mx_2 + b \\ y_3 &= mx_3 + b \\ y_4 &= mx_4 + b \end{aligned} \rightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} \rightarrow \mathbf{f} = \begin{bmatrix} 0.16 \\ 1.13 \\ 1.57 \\ 2.35 \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} 3.01 & 1 \\ 4.98 & 1 \\ 6.91 & 1 \\ 8.76 & 1 \end{bmatrix}$$

With practice, you will be able to write these matrices directly from the measured data.

Step 2 – Solve by least squares.

$$\mathbf{Z}^T\mathbf{Z}\mathbf{x} = \mathbf{Z}^T\mathbf{f} \quad \rightarrow \quad \begin{bmatrix} 3.01 & 4.98 & 6.91 & 8.76 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3.01 & 1 \\ 4.98 & 1 \\ 6.91 & 1 \\ 8.76 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 3.01 & 4.98 & 6.91 & 8.76 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.16 \\ 1.13 \\ 1.57 \\ 2.35 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 158.3462 & 23.6600 \\ 23.6600 & 4.0000 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 37.5437 \\ 5.2100 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 0.3656 \\ -0.8602 \end{bmatrix}$$

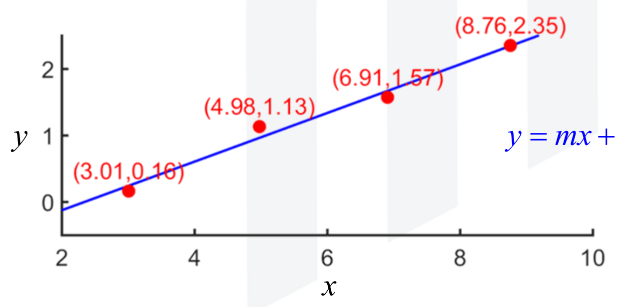
EMPossible 29

Example 1 (3 of 3)

Step 3 – Extract coefficients

$$\begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 0.3656 \\ -0.8602 \end{bmatrix} \rightarrow \begin{aligned} m &= 0.3656 \\ b &= -0.8602 \end{aligned}$$

Step 4 – Plot the result



$$y = mx + b = 0.3656x - 0.8602$$

EMPossible 30