ESC103 Unit 15

Aspen Erlandsson

November 15, 2022

Abstract

1 Solving Least Squares Problem

$A\overrightarrow{x} \neq \overrightarrow{b}$
We can still get as close as possible:

$A\overrightarrow{x} \approx \overrightarrow{b}$

Example: Let's say we have 3 (x, y) data points (-1, 6), (2, 0), (3, 2)
We want to fit a quadratic in the form below to these three points:

$$y = a + bx + cx^2$$

$$A\overrightarrow{x} = \overrightarrow{b}$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Rewriting with the values we know:

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1|0 \\ 1 & 2 & 4|0 \\ 1 & 3 & 9|2 \end{bmatrix}$$

$= \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 1 \end{array}\right)$

Square $(m = n = 3)$, $r = 3$ $A\overrightarrow{x} = \overrightarrow{b}$ has one solution

$$a = 2b = -3c = 1$$

therefore:

$$y = 2 - 3x + x^2$$

Now let's try to fit a linear function to go through the 3 data points:

$$y = a + bx$$

$$A\overrightarrow{x} = \overrightarrow{b}$$

$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$

$= \left(\begin{array}{cc|c} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{array}\right) = \left(\begin{array}{cc|c} 1 & -1 & 6 \\ 1 & 2 & 0 \\ 1 & 3 & 2 \end{array}\right) = \left(\begin{array}{cc|c} 1 & 0 & 4 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{array}\right)$ This system is tall and thin $(m = 3 > n = 2)$ Rank $= 2$ (two leading ones)

There are no solutions, because notice the bottom equation is

$$00 = 1$$

implying that

$$0 = 1$$

so it cannot be true.

A bit more linear algebra:

Transpose of a matrix:

If matrix A is $m \cdot n$ then the transpose of matrix A, denoted as A^T is the $n \cdot m$matrix here the rows of A^T are the columns of matrix A written in the same order.

A $= \begin{bmatrix} 1 & 2 & -1 \\ 0 & 5 & 6 \end{bmatrix}$ $A^T = \begin{pmatrix} 1 & 0 \\ 2 & 5 \\ -1 & 6 \end{pmatrix}$

$$P1 : (A^T)^T = A$$

$$P2 : (CA)^T = CA^T$$
$$P3 : (A + B)^T = A^T + B^T$$
$$P4 : (AB)^T = B^T A^T$$

$$A\overrightarrow{x} = \overrightarrow{b}$$
$$A\overrightarrow{x} = x_1 \begin{bmatrix} ... \\ a_1 \\ ... \end{bmatrix} + x_2 \begin{bmatrix} ... \\ a_2 \\ ... \end{bmatrix} + x_n \begin{bmatrix} ... \\ a_n \\ ... \end{bmatrix}$$
$$A\overrightarrow{x} = \overrightarrow{b}$$

There *is* a solution to $A\overrightarrow{x} = \overrightarrow{b}$ if $\overrightarrow{b}$ lies in the column space of matrix A.

There is *no* solution to $A\overrightarrow{x} = \overrightarrow{b}$ if $\overrightarrow{b}$ does *not* lie in the column space of matrix A.

To solve the latter problem (the one immediately above, in case you are illiterate), we want to find a vector $\overrightarrow{x}$ such that the vector $A\overrightarrow{x}$ is closest to vector $\overrightarrow{b}$

Let's begin by defining an error vector $\overrightarrow{e}x$ where:

$\overrightarrow{e} = \overrightarrow{b} - A\overrightarrow{x} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$

When we project $\overrightarrow{b}$ onto the column space of Matrix A, the error vector will be orthogonal to every column vector of Matrix A.

$\therefore \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = 0$

We are going to express these dot products in a different way:

$[a_1, a_2, a_3] \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = 0$

...

...

$$[a_n, a_n, a_n] \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = 0$$

Let's put these n equations together

$$\begin{bmatrix} \leftarrow \overrightarrow{a_1} \rightarrow \\ \leftarrow \overrightarrow{a_2} \rightarrow \\ \leftarrow \overrightarrow{a_3} \rightarrow \\ \leftarrow \ldots \rightarrow \\ \leftarrow \overrightarrow{a_n} \rightarrow \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \ldots \\ e_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

n x m, m x 1, n x 1

What we have is:

$$A^T \overrightarrow{e} = A^T(\overrightarrow{b} - A\overrightarrow{x_{LS}}) = \overrightarrow{0}$$

We want to solve for $\overrightarrow{x_{LS}}$

$$A^T \overrightarrow{b} - A^T A \overrightarrow{x_{LS}} = \overrightarrow{0}$$

Normal Equations

$$\therefore A^T A \overrightarrow{x_{LS}} = A^T \overrightarrow{b}$$

Initial matrix A was n x m, meaning the transposed matrix is m x n, therefore $A^T A$ is n x n. $\overrightarrow{x_{LS}}$ is nx1. $A^T \overrightarrow{b}$ is n x 1.

Let's define:

$$A^* = A^T A$$
$$\overrightarrow{b^*} = A^T \overrightarrow{b}$$
$$A^* \overrightarrow{x_{LS}} = \overrightarrow{b^*}$$

This is just another $A\overrightarrow{x} = \overrightarrow{b}$ type of problem :D

The least squares system is definitely square: (n x n)

If rank of $A^* = n$ then there is a unique solution for $\overrightarrow{x_{LS}}$ (the least squares solution).

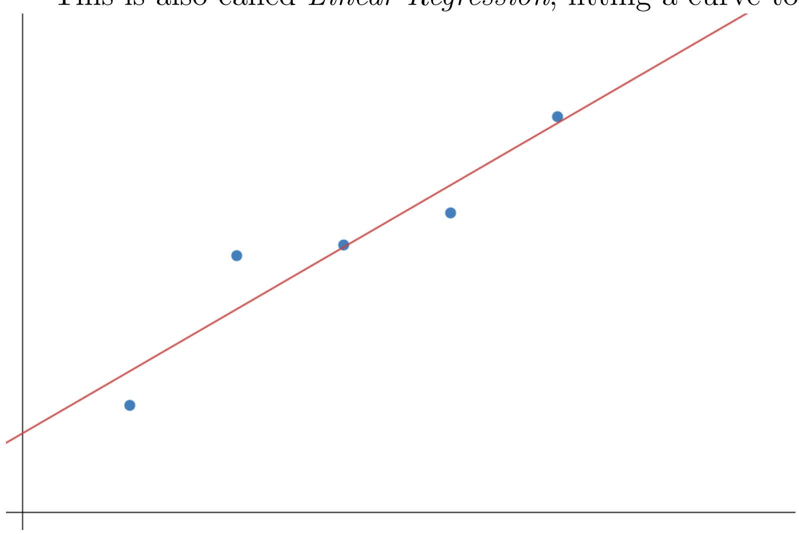
Why do we call this least squares?:
When we project $\overrightarrow{b}$ onto the column space of Matrix A, we are finding the shortest vector $\overrightarrow{e}$.

$$||\overrightarrow{e}|| = \sqrt{e_1^2 + e_e^2 + e_3^2 + e_m^2}$$

It turns out that minimizing the $||\overrightarrow{e}||$ is equivalent to minimizing $||\overrightarrow{e}||^2$. Now we can conveniently remove the square root we had in the above magnitude calculation.

$$||\overrightarrow{e}||^2 = e_1^2 + e_2^2 + e_3^2 ... + e_m^2$$

This is also called *Linear Regression*, fitting a curve to some data:



We have some points:

$$(x_1, y_2), (x_2, y_2), (x_3, y_3....)$$

We can construct our data fitting matrix:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

Now that won't be perfect, so we can define our error vector as the difference between the true data points, and the points predicted by our best fit.

$$\overrightarrow{e} = \overrightarrow{b} - A\overrightarrow{x_{LS}}$$

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \begin{bmatrix} a + bx_1 \\ a + bx_2 \\ a + bx_3 \\ a + bx_4 \\ a + bx_5 \end{bmatrix}$$

$a + bx_1$ is the *predicted* height of the point, while y_1 is the *true* position of the point. For example, the first point on the graph above is below the predicted line, so therefore e_1 would be negative in this particular case as a result of the subtraction.