27. Least Squares

Consider a system of equations

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

which is overdetermined, that is, the number of equations is more than the number of variables, $m > n$. For most $\vec{b}$ there won't be a solution. Often this is because of noise, meaning that the data is not quite correct. In this case we expect that there is a point $\vec{b}_0$ very close to $\vec{b}$ for which we can solve the equations.

How to choose $\vec{b}_0$? Well the set of all vectors $\vec{b}_0$ for which there is a solution is the column space $\operatorname{Col}(A)$. So let's choose the closest point $\vec{b}_0$ to $\vec{b}$ in the column space.

Definition 27.1. *Let A be an $m \times n$ matrix.*

The least squares solution to $A\vec{x} = \vec{b}$ is a vector $\vec{x}_0$ such that

$$\|\vec{b} - A\vec{x}_0\| \leq \|\vec{b} - A\vec{x}\| \qquad \text{for all} \qquad x \in \mathbb{R}^n.$$

We will see two methods to find $\vec{x}_0$.

Method #1

Example 27.2. *Find a least squares solution to the equation $A\vec{x} = \vec{b}$, where*

$$A = \begin{pmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{pmatrix} \qquad \text{and} \qquad \vec{b} = \begin{pmatrix} 4 \\ -2 \\ -3 \end{pmatrix}.$$

Let $\vec{u}_1 = (1, 3, -2)$ and $\vec{u}_2 = (5, 1, 4)$. We want the closest point $\vec{b}_0$ to the column space of A, we want the orthogonal projection of $\vec{b}$ onto the column space W, the span of $\vec{u}_1$ and $\vec{u}_2$, $\operatorname{proj}_W \vec{b}$.

Note that

$$\vec{u}_1 \cdot \vec{u}_2 = (1, 3, -2) \cdot (5, 1, 4) = 0,$$

so that $\vec{u}_1$ and $\vec{u}_2$ are orthogonal. So finding $\operatorname{proj}_W \vec{b}$ is straightforward, if we write

$$\vec{b}_0 = \alpha_1 \vec{u}_1 + \alpha_2 \vec{u}_2,$$

then

$$\alpha_1 = \frac{\vec{b} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} = \frac{(4, -2, -3) \cdot (1, 3, -2)}{(1, 3, -2) \cdot (1, 3, -2)} = \frac{4}{14} = \frac{2}{7},$$

and

$$\alpha_2 = \frac{\vec{b} \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2} = \frac{(4, -2, -3) \cdot (5, 1, 4)}{(5, 1, 4) \cdot (5, 1, 4)} = \frac{6}{42} = \frac{1}{7},$$

Therefore

$$\vec{b}_0 = \frac{2}{7}(1, 3, -2) + \frac{1}{7}(5, 1, 4) = (1, 1, 0).$$

Note we already know how to solve

$$A\vec{x} = \vec{b}_0,$$

The solution is

$$\vec{x}_0 = \frac{1}{7}(2, 1).$$

What do we do if the columns of A are not orthogonal? We could apply Gram-Schmidt but unfortunately this is quite expensive, that is, it takes quite a bit of time to find an orthogonal basis of $\mathrm{Col}(A)$.

Method #2

We know that the vector

$$\vec{b}_1 = \vec{b} - \vec{b}_0,$$

is orthogonal to the column space of A. But we already saw that

$$\mathrm{Col}(A)^T = \mathrm{Nul}(A^T).$$

Hence

$$\vec{b} - \vec{b}_0 \in \mathrm{Nul}(A^T).$$

that is,

$$A^T(\vec{b} - \vec{b}_0) = \vec{0}.$$

Suppose that $\vec{x}_0$ is a solution of

$$A\vec{x} = \vec{b}_0 \qquad \text{so that} \qquad A\vec{x}_0 = \vec{b}_0.$$

Then

$$A^T(\vec{b} - A\vec{x}_0) = \vec{0}.$$

Rearranging we get

Theorem 27.3. *$\vec{x}$ is a least squares solution of $A\vec{x} = \vec{b}$ if and only if $\vec{x}$ is a solution of*

$$A^T A\vec{x} = A^T\vec{b}.$$

Let's solve (27.2) again, using the second method.

Example 27.4. *Find a least squares solution to the equation $A\vec{x} = \vec{b}$, where*

$$A = \begin{pmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{pmatrix} \qquad \text{and} \qquad \vec{b} = \begin{pmatrix} 4 \\ -2 \\ -3 \end{pmatrix}.$$

$$A^T = \begin{pmatrix} 1 & 3 & -2 \\ 5 & 1 & 4 \end{pmatrix}.$$

So

$$A^T A = \begin{pmatrix} 1 & 3 & -2 \\ 5 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 14 & 0 \\ 0 & 42 \end{pmatrix}$$

and

$$A^T \vec{b} = \begin{pmatrix} 1 & 3 & -2 \\ 5 & 1 & 4 \end{pmatrix} \begin{pmatrix} 4 \\ -2 \\ -3 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$$

We are supposed to solve

$$\begin{pmatrix} 14 & 0 \\ 0 & 42 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$$

which has the unique solution

$$(x, y) = (2/7, 1/7)$$

as expected.

Example 27.5. *Find the least squares solutions to the equation $A\vec{x} = \vec{b}$, where*

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \vec{b} = \begin{pmatrix} 1 \\ 3 \\ 8 \\ 2 \end{pmatrix}.$$

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

So

$$A^T A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}$$

and

$$A^T \vec{b} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 8 \\ 2 \end{pmatrix} = \begin{pmatrix} 14 \\ 4 \\ 10 \end{pmatrix}$$

We are supposed to solve

$$\begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 14 \\ 4 \\ 10 \end{pmatrix}$$

We apply Gaussian elimination:

$$\left(\begin{array}{ccc|c} 4 & 2 & 2 & 14 \\ 2 & 2 & 0 & 4 \\ 2 & 0 & 2 & 10 \end{array}\right) \to \left(\begin{array}{ccc|c} 2 & 1 & 1 & 7 \\ 1 & 1 & 0 & 2 \\ 1 & 0 & 1 & 5 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 1 & 1 & 0 & 2 \\ 2 & 1 & 1 & 7 \end{array}\right)$$

so that

$$\to \left(\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 0 & 1 & -1 & -3 \\ 0 & 1 & -1 & -3 \end{array}\right) \to \left(\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 0 & 1 & -1 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

x and y are basic variables and z is a free variable.

$$y - z = -3 \qquad \text{so that} \qquad y = -3 + z.$$

Therefore

$$x + z = 5 \qquad \text{so that} \qquad y = 5 - z.$$

The general solution is

$$(x, y, z) = (5 - z, -3 + z, z) = (5, -3, 0) + z(-1, 1, 1).$$