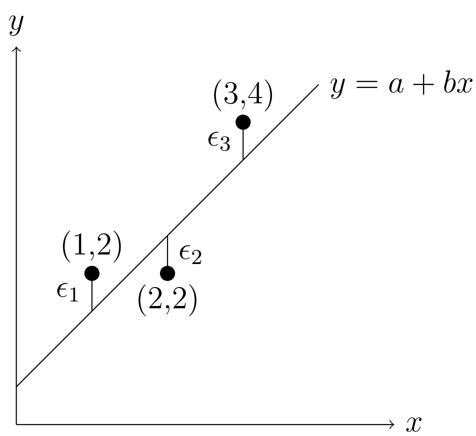
Advanced Linear Algebra

Alvin Lin

January 2019 - May 2019

Least Squares Approximation



Suppose we want to find the line that best fits the three points. We would want to minimize the errors $\epsilon_1, \epsilon_2, \epsilon_3$. There is no direct solution for the line since attempting to solve for it yields the following:

$$A\vec{x} = \vec{b}$$
$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

This is an inconsistent system with no solution. We are more interested in minimizing the error vector $\vec{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_1 \\ \epsilon_3 \end{bmatrix} = \vec{b} - A\vec{x}$. To do this, we can minimize the "least squares error":

$$\|\vec{\epsilon}\| = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2$$

In general, for n data points:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

The least squares solution is the $\vec{x}$ in $\mathbb{R}^2$ that satisfies:

$$\|b - A\bar{x} \leq \|b - Ax\| \forall x \in \mathbb{R}^2$$

We wish to find a and b to minimize $\|\epsilon\|^2$. As x varies over $\mathbb{R}^n$, then Ax varies over $col(A)$, so we want the closest vector in $col(A)$ to $\vec{b}$. Since $\vec{b}$ is not on the subspace, the closest vector to $\vec{b}$ is the orthogonal projection to the subspace.

$$\begin{aligned} A\bar{x} &= \text{proj}_{col(A)}\vec{b} \\ \vec{b} - A\bar{x} &= \vec{b} - \text{proj}_{col(A)}\vec{b} \\ \vec{v} - A\bar{x} &= \text{perp}_{col(A)}\vec{b} \end{aligned}$$

$\vec{b} - A\bar{x}$ is in $(col(A))^T$ or $null(A^T)$. Therefore, we can get the equations of the normal as:

$$\begin{aligned} A^T(\vec{b} - A\bar{x}) = \vec{0} \\ A^T A\bar{x} = A^T\vec{b} \end{aligned}$$

Example

Find the least squares solution to the inconsistent system $A\bar{x} = \vec{b}$ where

$$A = \begin{bmatrix} 1 & 5 \\ 2 & -2 \\ -1 & 1 \end{bmatrix} \quad \vec{b} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$$

$$
\begin{aligned}
A^T A &= \begin{bmatrix} 1 & 2 & -1 \\ 5 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 2 & -2 \\ -1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 6 & 0 \\ 0 & 30 \end{bmatrix} \\
A^T \vec{b} &= \begin{bmatrix} 1 & 2 & -1 \\ 5 & -2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ 16 \end{bmatrix} \\
A^T A \bar{x} &= A^T \vec{b} \\
\begin{bmatrix} 6 & 0 \\ 0 & 30 \end{bmatrix} \bar{x} &= \begin{bmatrix} 2 \\ 16 \end{bmatrix} \\
\bar{x} &= \begin{bmatrix} \frac{1}{3} \\ \frac{8}{15} \end{bmatrix}
\end{aligned}
$$

Example

Find the least squares approximating line $y = a + bx$ for the data points (1,2), (2,2), (3,4).

$$
\begin{aligned}
A\bar{x} &= \vec{b} \\
\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \\
A^T A &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \\
A^T \vec{b} &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 8 \\ 18 \end{bmatrix} \\
A^T A \bar{x} &= A^T \vec{b} \\
\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} 8 \\ 18 \end{bmatrix} \\
\begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} \frac{2}{3} \\ 1 \end{bmatrix} \\
y &= \frac{2}{3} + 1x
\end{aligned}
$$

Moore-Penrose Pseudoinverse

From our normal equations, we can generally approximately a solution to an inconsistent solution, and phrase it as follows:

$$\begin{aligned} A^T A\bar{x} &= A^T\vec{v} \\ \bar{x} &= (A^T A)^{-1} A^T\vec{v} \\ A^+ &= (A^T A)^{-1} A^T \end{aligned}$$

This term A^+ is known as the Moore-Penrose Pseudoinverse.

Example

Find the parabola that gives the best least squares approximation to the points (-1,1), (0,-1), (1,0), (2,2).
In general, parabolas are of the form $y = a + bx + cx^2$, so we can form an inconsistent system of equations to represent this problem.

$$\begin{aligned} a + b(-1) + c(-1)^2 &= 1 \\ a + b(0) + c(0)^2 &= -1 \\ a + b(1) + c(1)^2 &= 0 \\ a + b(2) + c(2)^2 &= 2 \\ \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} &= \begin{bmatrix} 1 \\ -1 \\ 0 \\ 2 \end{bmatrix} \end{aligned}$$

This is an inconsistent system with no exact solution, but we can find a least squares solution by solving the normal equations.

$$A^TA = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 2 \\ 3 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} -\frac{7}{10} \\ -\frac{3}{5} \\ 1 \end{bmatrix}$$

Example

Note that if you have a subspace $W = col(A)$, then the following is true:

$$\text{proj}_W\vec{v} = A\bar{x} = A(A^TA)^{-1}A^T\vec{v}$$

This term is the matrix form of the orthogonal projection.

Find the orthogonal projection of $\vec{v} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$ onto W in $\mathbb{R}^3$ given by $x - y + 2z = 0$.

Traditionally, we define the subspace with two basis vectors and we use the definition of a projection to solve for $\text{proj}_W\vec{v}$.

$$\vec{u_1} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \vec{u_2} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \quad B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

If we want to solve it with the matrix form, we can do the following:

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} \frac{3}{6} & 0 \\ 0 & \frac{2}{6} \end{bmatrix}$$

$$A(A^T A)^{-1} A^T = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{5}{6} & \frac{1}{6} & -\frac{1}{3} \\ \frac{1}{6} & \frac{5}{6} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

This is the projection matrix, which will project any vector onto the column space of A.

$$\text{proj}_W \vec{v} = \begin{bmatrix} \frac{5}{6} & \frac{1}{6} & -\frac{1}{3} \\ \frac{1}{6} & \frac{5}{6} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{5}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix}$$

What makes this method particularly convenient is that it does not require an orthogonal basis for A, it only requires that the column space of A spans the subspace W.

You can find all my notes at `http://omgimanerd.tech/notes`. If you have any questions, comments, or concerns, please contact me at alvin@omgimanerd.tech