

Least squares approximation

Lets take some data: the points $(0, 0)$, $(1, 1)$, $(2, 4)$. What is the best straight line to approximate this data? We are trying to find the "best" solution to systems without solutions.

A line through this data has the form $y = c + mx$ where we want to compute m and c .

$$\begin{array}{l} c + 2m = 4 \\ c + 1m = 1 \\ c + 0m = 0 \end{array} \implies \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c \\ m \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix}$$

Trying to solve yields

$$\left(\begin{array}{cc|c} 1 & 2 & 4 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{array} \right) \rightarrow \left(\begin{array}{cc|c} 1 & 2 & 4 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{array} \right) \rightarrow \left(\begin{array}{cc|c} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 1 & 1 \end{array} \right)$$

This doesn't have a solution, but we can still ask what has the best line.

We want to minimize the distance between each point and the line, aka the least squares error.

We will minimize the norm of errors between these points

$$\sqrt{e_1^2 + e_2^2 + e_3^2}$$

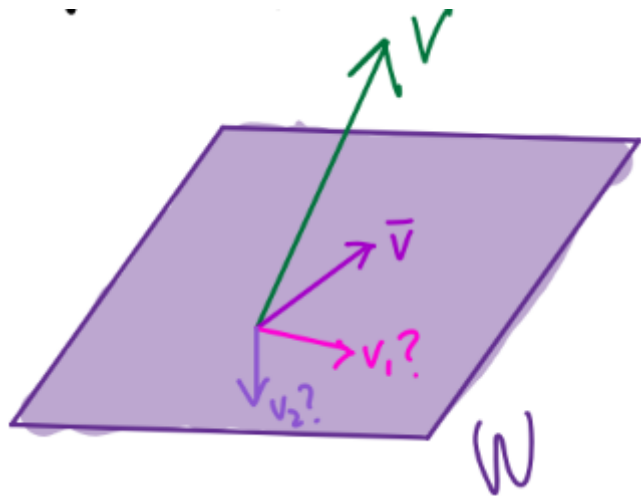
By best approximation, we mean

$$|v - \bar{v}| \leq \langle v - w \rangle \forall w \in W \text{ where } w \neq \bar{v}$$

If $Ax = b$ has a solution, then $e_1 = e_2 = e_3 = 0$. That is, $|b - Ax| = 0$, b and Ax are the same vector. If there is no solution, then what is the

vector $\bar{y} = A\bar{x}$ in $\text{col}(A)$ that is the best approximation to b ?

That is, the vector in W that has the shortest distance to v .



Claim: $\bar{v} = \text{proj}_W(v)$.

Proof: By the Pythagorean theorem, $|v - \bar{v}|^2 + |\bar{v} - w|^2 = |v - w|^2$ since $v - \bar{v}$ is orthogonal to $\bar{v} - w \in W$. Thus, $|v - \bar{v}|^2 \leq |v - w|^2$ for any $w \neq \bar{v}, w \in W$.

We want $x = \bar{x}$ that gives the orthogonal projection of b onto $\text{col}(A)$.

$$A\bar{x} = \text{proj}_{\text{col}(A)}(b)$$

$$b - A\bar{x} = b - \text{proj}_{\text{col}(A)}(b) \text{ (in the null}(A^T))$$

$$A^T(b - A\bar{x}) = 0 \implies A^T A\bar{x} = A^T b.$$

Theorem: Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then a least squares solution \bar{x} of $Ax = b$ satisfies "the normal equations for \bar{x} ":

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

Lets solve this with our example. Remember, we have the points $(0, 0), (1, 1), (2, 4)$.

Since we can't solve the system, we will look for a least squares solution \bar{x} :

$$A^T A \bar{x} = A^T b$$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}$$

so we have

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 3 & | & 5 \\ 3 & 5 & | & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 3 & | & 5 \\ 0 & 2 & | & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 0 & | & -1 \\ 0 & 1 & | & 2 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & 0 & | & -\frac{1}{3} \\ 0 & 1 & | & 2 \end{bmatrix}$$

because $x_1 = c, x_2 = m$, the line we seek is

$$y = 2x - \frac{1}{3}$$

We can calculate the least squares error:

$$\begin{aligned}
|e| &= |b - A\bar{x}| \\
&= \left| \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix} - \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} -\frac{1}{3} \\ 2 \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} \frac{11}{3} \\ \frac{5}{3} \\ -\frac{1}{3} \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \end{pmatrix} \right| \\
&= \frac{1}{3} \left| \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix} \right| \\
&= \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix} \\
&= \frac{\sqrt{6}}{3}
\end{aligned}$$

However, small numerical errors in Gaussian elimination create large errors in the least squares solution (a computer struggles with $\frac{1}{3}$, they approximate it). IE. exploding rockets etc in the real world because of how massively these rounding errors stack up.

QR factorizations are more reliable and stable in this case.

Suppose A is a $m \times n$ matrix with linearly independent columns, $b \in \mathbb{R}^m$.

We can use QR factorization to find the least squares solution \bar{x} for $Ax = b$.

Since A has linearly independent columns, we can find a QR factorization

$A = QR$ where Q is orthogonal and R is upper triangular and invertible. With some manipulation, we can find a much better system of equations.

$$\begin{aligned}A^T A \bar{x} &= A^T b \\ \implies (AR)^T (AR) \bar{x} &= (QR)^T b \\ \implies R^T Q^T Q R \bar{x} &= R^T Q^T b \\ \implies R^T R \bar{x} &= R^T Q^T b \\ \implies R \bar{x} &= Q^T b\end{aligned}$$

The only math we have to do is multiplication and subtraction when solving for R . This is a much nicer system.