Least Squares - Motivation

Science is replete with curve fitting problems. In linear regression one has data points \{(x_n,y_n)\} we wish to fit to a linear model

\[ y = mx + b \]

The question is, what is a good choice of slope \(m\) and intercept \(b\). A perfect fit would have

\[
\begin{align*}
   b + mx_1 &= y_1 \\
   b + mx_2 &= y_2 \\
   &\vdots \\
   b + mx_n &= y_n
\end{align*}
\]

which can be written

\[
\begin{bmatrix}
1 & x_1 \\
1 & x_2 \\
\vdots & \vdots \\
1 & x_n
\end{bmatrix}
\begin{bmatrix}
b \\
m
\end{bmatrix} =
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}
\]

This overdetermined system (\(2\) unknowns, \(n\) equations) then becomes a system problem

\[
(1) \quad Ax = b
\]

where \(b \notin \text{col}(A)\). In what is to follow we examine "least squares" solutions of \((1)\).
Least Squares - Geometry

\[ Ax = b \quad b \in \text{col}(A) \]

Since \( b - \text{proj}_W b \in W^\perp \) we expect it to be closest. In other words, if we can find an \( \hat{x} \) such that

\[ A\hat{x} = \text{proj}_W b \]

then \( \hat{x} \) is the "best" soln to \( Ax = b \) since then \( A\hat{x} \) is closest to \( b \).

We now formalized this.
Best Approximation Theorem

Let $W$ be a subspace of inner product space $V$. Then

$$\hat{w} = \text{proj}_W b \quad \hat{w} \in W$$

is the best approximation to $b$ in the sense that

$$\|b - \text{proj}_W b\| < \|b - w\|$$

for all $w \neq \text{proj}_W b$, $w \in W$.

Proof:

$$b - w = (b - \text{proj}_W b) + (\text{proj}_W b - w) \in W^\perp \in W$$

Pythagorus Thm applies, hence

$$\|b - w\|^2 = \|b - \text{proj}_W b\|^2 + \|\text{proj}_W b - w\|^2$$

$$\|b - w\|^2 > \|b - \text{proj}_W b\|^2$$

Pythagorus for general inner product spaces

$$z = x + y \quad \langle x, y \rangle = 0$$

then

$$\|z\|^2 = \langle x + y, x + y \rangle = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2$$

$$\|z\|^2 = \|x\|^2 + \|y\|^2$$
Least Squares - Normal Eqns

The least squares solution $x$ of $Ax = b$ must satisfy

(1) $Ax = \text{proj}_W b$

Rather than solve this for $x$, consider:

$b - Ax = b - \text{proj}_W b \in \text{col}(A)^\perp = N(A^T)^*$

Hence

$A^T(b - Ax) = A^T(b - \text{proj}_W b) = 0^*$

Since the right side vanishes, $x$ is a soln of:

(2) $A^TAx = A^Tb$

normal eqns

To find $x$, use normal eqns (2).

Theorem

Given $Ax = b$ then the normal eqns

(3) $A^TAx = A^Tb$

is consistent. Any soln of (3) is a least squares soln of $Ax = b$ and has

$Ax = \text{proj}_W b$
**Example**  Find the least squares solution of

\[
\begin{bmatrix}
1 & 2 \\
1 & 1 \\
1 & -2 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix} =
\begin{bmatrix}
13 \\
-13 \\
0 \\
\end{bmatrix}
\]

After some calculations the normal equations

\[ A^T A \mathbf{x} = A^T \mathbf{b} \]

become

\[
(1) \quad \begin{bmatrix}
3 & 1 \\
1 & 9 \\
\end{bmatrix} \mathbf{x} = \begin{bmatrix}
0 \\
13 \\
\end{bmatrix}
\]

In this (typical) example \( A^T A \) is invertible.

Solving (1)

\[ \mathbf{x} = \begin{pmatrix}
-\frac{1}{2} \\
\frac{3}{2} \\
\end{pmatrix} \]

Relation to linear regression \( (x_1, x_2) \) = (intercept, slope)

\[ y = \frac{3}{2} x - \frac{1}{2} \]
Least Squares Calculus Derivation.

\[ F(x) = \|Ax - b\|^2 \]

Seek \( x \) s.t. \( F(x) \) is minimized. Define the residual \( r = Ax - b \). In component form

\[ r_i = \sum_{j=1}^{n} A_{ij} x_j - b_i \]

hence

\[ F(x) = \sum_{i=1}^{m} r_i^2 \]

To minimize \( F(x) \) we must set

(1) \[ \frac{\partial F}{\partial x_k} = \sum_{i=1}^{m} 2 r_i \frac{\partial r_i}{\partial x_k} = 0 \quad \forall k. \]

From the definition of \( r_i \)

(2) \[ \frac{\partial r_i}{\partial x_k} = A_{ik} \]

Use (2) in (1) to get

\[ \frac{\partial F}{\partial x_k} = \sum_{i=1}^{m} 2 \left( \sum_{j=1}^{n} A_{ij} x_j - b_i \right) A_{ik} = 0 \]

\[ \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ik} A_{ij} x_j = \sum_{i=1}^{m} A_{ik} b_i \]

\[ A^T A x = A^T b \]