Least Squares: Theory & Examples.

Announcements:
- Extended Section M/W 10-12
- P/NP / Grade option. Select by 5/6.

Least Squares: Goal is to approximate
Solve a system of Eqs
\[ A \hat{x} \approx b \]
known (aka variable).

Least Squares optimization problem:
\[ \min \| A \hat{x} - b \|_2 \]
\[ \hat{x} \]
Assumption: we are working with Euclidean norm & inner product.
Lots of Geometric Intuition!

Last time: started with a simpler problem: (one-dimensional $\vec{a}$)

$$\min_{x \in \mathbb{R}} \| \hat{a} x - \vec{b} \|^2$$

$$\min_{\vec{v} \in \text{Span}(\hat{a})} \| \vec{v} - \vec{b} \|^2$$

Key idea: error $\vec{e} = \vec{b} - \vec{v}^*$ is orthogonal to $\text{Span}(\hat{a})$.

Let's check that if $\vec{e}$ is orthogonal to $\text{Span}(\hat{a})$, then $\vec{v}^*$ must be closest point.
For any other $\vec{v} \in \text{Span}(\vec{a})$, $\vec{v} \in \mathbb{R}(\vec{a})$

\[
\| \vec{v} - \vec{b} \|^2 = \| (\vec{v} - \vec{v}^*) + (\vec{v}^* - \vec{b}) \|^2 \\
= \| \vec{v} - \vec{v}^* \|^2 + \| \vec{v}^* - \vec{b} \|^2 + 2 \langle \vec{v} - \vec{v}^*, \vec{v}^* - \vec{b} \rangle \\
\quad \text{if } \forall \vec{v}^* \in \text{Span}(\vec{a}), \exists \vec{e} \\
= \| \vec{v} - \vec{v}^* \|^2 + \| \vec{v}^* - \vec{b} \|^2 \\
\geq \| \vec{v}^* - \vec{b} \|^2 \quad (\ast)
\]

Note: we used:

\[
\| \vec{x} + \vec{y} \|^2 = \langle \vec{x} + \vec{y}, \vec{x} + \vec{y} \rangle \\
= \langle \vec{x}, \vec{x} + \vec{y} \rangle + \langle \vec{y}, \vec{x} + \vec{y} \rangle \\
= \langle \vec{x}, \vec{x} \rangle + \langle \vec{x}, \vec{y} \rangle + \langle \vec{y}, \vec{x} \rangle + \langle \vec{y}, \vec{y} \rangle \\
= \| \vec{x} \|^2 + \| \vec{y} \|^2 + 2 \langle \vec{x}, \vec{y} \rangle.
\]

Observe similar to: $(x+y)^2 = x^2 + y^2 + 2xy$. 
Conclusion from \((\ast)\) is that if 

\[ \hat{e} = \hat{v}^* - \hat{b} \]

is orthogonal to \(\text{Span}(\hat{a})\), then \(\hat{v}^*\) is solution to

\[ \min_{\hat{v}} \| \hat{v} - \hat{b} \|^2 \]

\[ \hat{v} \in \text{Span}(\hat{a}) \].

Claim: optimal vector

\[ \hat{v}^* = \frac{\langle \hat{b}, \hat{a} \rangle \hat{a}}{\langle \hat{a}, \hat{a} \rangle} \]

\[ \text{proj}_a (\hat{b}) \]

Proof: just check for desired orthogonality.

\[ \langle \hat{e}, \beta \hat{a} \rangle = \beta \langle \hat{e}, \hat{a} \rangle = \beta \langle \hat{b} - \text{proj}_a (\hat{b}), \hat{a} \rangle \]

\[ \text{a vector in } \text{Span}(\hat{a}) \]

\[ \hat{e} \perp \hat{a} \]

\[ \hat{e} = 0 \] did last time.
More generally:

$$\min_{\hat{x}} \| A \hat{x} - \hat{b} \|^2 = \min_{\hat{v} \in \text{Range}(A)} \| \hat{v} - \hat{b} \|^2$$

Note: If $\hat{e}$ is orthogonal to $\text{Range}(A)$, then $\hat{v}^*$ is optimal for LS problem

$$\min_{\hat{v} \in \text{Range}(A)} \| \hat{v} - \hat{b} \|^2$$

Logic is same as before: For any vector $\hat{v} \in \text{Range}(A)$, $\| \hat{v} - \hat{b} \|^2 \geq \| \hat{v}^* - \hat{b} \|^2$. 

\[\hat{v}^* \text{ way}\]
Thm: If $A$ has zero nullspace, then the LS solution to

$$\min_{\hat{x} \in \mathbb{R}^n} \|A\hat{x} - \hat{b}\|^2$$

is equal to $\hat{x}^* = (A^TA)^{-1}A^T\hat{b}$.

Proof: Just need to check that

$\hat{e} = \hat{x}^* - \hat{b}$ is orthogonal to $R(A)$, where

$\hat{V} = A\hat{x}^* = A(A^TA)^{-1}A^T\hat{b}$.

First claim that $N(A) = \{\hat{0}\} \Rightarrow (A^TA)$ is invertible. (we'll check this later). Let $\hat{V} \in R(A)$.

$$\langle \hat{e}, \hat{V} \rangle = \langle \hat{x}^* - \hat{b}, \hat{V} \rangle$$

$$= \langle \underbrace{A(A^TA)^{-1}A^T\hat{b} - \hat{b}}, \hat{V} \rangle$$

If $\hat{V} \in R(A) \Rightarrow \hat{V} = A\hat{x}$ for some $\hat{x}$.  

\[
\langle \tilde{a}, \tilde{b} \rangle \quad = \quad \tilde{a}^T \tilde{b} \\
= \quad \tilde{x}^T \tilde{a}.
\]

\[
= \langle A (A^T A)^{-1} A^T \tilde{b}, \tilde{x} \rangle \\
- \langle \tilde{b}, A \tilde{x} \rangle \\
= \left( \tilde{x}^T A (A^T A)^{-1} A^T \tilde{b} \right) \\
- \left( \tilde{x}^T A^T \tilde{b} \right) \\
= \tilde{x}^T (A^T A) (A^T A)^{-1} A^T \tilde{b} \\
- \tilde{x}^T A^T \tilde{b} \\
= \tilde{x}^T A^T \tilde{b} - \tilde{x}^T A^T \tilde{b} = 0 \checkmark
\]

Only thing left to do:

**Technical Lemma:** \( N(A) = N(A^T A) \), i.e., \( A^T A \) is invertible.

**Pf:** First show \( N(A) \subseteq N(A^T A) \).

Take \( \tilde{x} \in N(A) \Rightarrow (A^T A) \tilde{x} = A^T (A \tilde{x}) = 0 \Rightarrow \tilde{x} \in N(A^T A) \)
Second show: \( N(A^T A) \subseteq N(A) \)

Take \( \tilde{x} \in N(A^T A) \Rightarrow A^T A \tilde{x} = 0 \)

Clear idea:

\[
\| A \tilde{x} \|_2^2 = \langle A \tilde{x}, A \tilde{x} \rangle = \langle \tilde{x}, A^T A \tilde{x} \rangle
\]

\[
\tilde{x}^T (A^T A) \tilde{x} = 0.
\]

\[\Rightarrow A \tilde{x} = 0 \Rightarrow \tilde{x} \in N(A)\]

\[
\langle A \tilde{x}, A \tilde{x} \rangle = (A \tilde{x})^T A \tilde{x}
\]

\[
= \tilde{x}^T (A^T A) \tilde{x}
\]

\[
= \langle \tilde{x}, A^T A \tilde{x} \rangle.
\]

**Remark:** If \( A \) has non-trivial nullspace, then LS solutions are set of \( \tilde{x} \) satisfying:

\[
(A^T A) \tilde{x} = A^T \tilde{b}. \quad \text{ (normal equations)}
\]

**Note:** This system of equations is always
consistent (need to show $R(\mathbf{A}^T) = R(\mathbf{A}^\top \mathbf{A}))$.

**Examples:**

**Linear Regression**

![Graph showing linear regression](image)

Want: Find "best" $a, b$ such that

$$ax_i + b \approx y_i \quad \text{for } i = 1 \ldots m.$$  

Answer: set up LS problem

$$\min_{a, b} \left\| \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots \\ x_m & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \right\|^2 = \min_{a, b} \sum_{i=1}^{m} \left( ax_i + b - y_i \right)^2$$
\[ [\theta] = (A^T A)^{-1} A^T \hat{b}. \]

Example: Hubble Space Telescope

Imprecise mirror made image blurry.

Image deblurring can be cast as a LS problem.
real image of Jupiter from HST

After designing a deblurring filter using Least Squares!
(See EE120 for details)