Example: Data Fitting

Have data: \((x_i, y_i)\) and model:

\[ y(x) = a + bx + cx^2 \]

- Find data that best fit model:

\[
\begin{pmatrix}
1 & x_1 & x_1^2 \\
1 & x_2 & x_2^2 \\
1 & x_3 & x_3^2
\end{pmatrix}
\begin{pmatrix}
a \\
b \\
c
\end{pmatrix}
=
\begin{pmatrix}
y_1 \\
y_2 \\
y_3
\end{pmatrix}
\]

\[ A x = b \]

\[ \|Ax - b\|_2^2 \]
Demo: Interactive Polynomial Fitting
Properties of Least-Squares

Consider the least squares problem $Ax \cong b$ and its associated objective function

$$\varphi(x) = \|b - Ax\|_2^2.$$  

- Is there always a solution to a linear least-squares problem?

  $\forall \geq 0 \quad \varepsilon \to 0 \quad (\|x\| \to \infty) \Rightarrow \exists \text{minimum}$

- Is it unique?

  If $A$ has a nullspace, then no.

- Examine the objective function, find its minimum.

  $$\varphi(x) = (b - Ax)^T (b - Ax) = b^T b - 2x^T A^T b + x^T A^T A x$$

  Normal equations

  $$\nabla \varphi(x) = 0 - 2A^T b + 2A^T A x = 0 \Leftrightarrow A^T A x = A^T b$$
In-class activity: Least Squares
Demo: Polynomial fit using the normal equations

- What’s the shape of $A^TA$?

Demo: Issues with the normal equations
Why is $r \perp \text{span}(A)$ a good thing to require?

(‘Pythagoras’)

Phrase that as an equation.

$$A^r b - A^r A x = 0$$
Write that with an orthogonal projection matrix $P$.

$A x = P b$
About Orthogonal Projectors

- What is a projector?
  \[ P^2 = P \]

- What is an orthogonal projector?
  \[(Py - y) \perp P y \rightarrow P \text{ symmetric, i.e. } P^2 = P \]

- How do I make one projecting onto \( \text{span}\{q_1, q_2, ..., q_k\} \)?
  Assume (or make) \( q_1, ..., q_k \) orthonormal. (Gram–Schmidt)
  \[ Q Q^T \times \]

- Check that \( P = A(A^T A)^{-1} A^T \) is an orthogonal projector onto \( \text{colspan}(A) \).
  \[
  P^2 = A(A^T A)^{-1} A^T A \left( A^T A \right)^{-1} A^T = A \left( A^T A \right)^{-1} A^T = P
  \]

- What assumptions do we need to define the \( P \) from the last question?
  \[
  \begin{pmatrix}
  A \\
  \vdots \\
  A^T \\
  \end{pmatrix} \begin{pmatrix}
  \vdots \\
  \end{pmatrix} = \begin{pmatrix}
  \vdots \\
  \end{pmatrix} \quad (A^T A) \text{ is only invertible if } A \text{ has full rank.}
Pseudoinverse

○ What is the pseudoinverse of $A$?

\[ x = (A^\top A)^{-1} A^\top b \]

$$(A^\top A)^{-1} A^\top$$ solves $A x = b$.

○ What can we say about the condition number in the case of a tall-and-skinny, full-rank matrix?

○ What does all this have to do with solving least squares problems?
3.2 Sensitivity and Conditioning
Sensitivity and Conditioning of Least Squares

- How sensitive is it?
- What values of $\theta$ are bad?
- Any comments regarding dependencies?