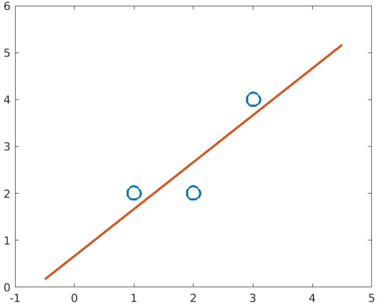
Annotated slides from Friday.

Week 11: Best Approximation and Least Squares

Dr Niall Madden

15 and 18 November, 2022



These slides are adapted (slightly) from ones by Tobias Rossmann.

Outline

+ R Example.

For more details,

- Section 6.3 (Best Approximation) and 6.6 (Least Squares) in Lay et al: https://nuigalway-primo.hosted.exlibrisgroup.com/permalink/f/1pmb9lf/353GAL_ALMA_DS5192067630003626
- Chapters 10 and 11 of *Linear Algebra for Data Science* https://shainarace.github.io/LinearAlgebra/leastsquares.html

1. ▶ Suppose $\hat{b} = \mathrm{proj}_{\mathrm{Col}\,A}(b)$, and then that $\hat{x}$ solves $A\hat{x} = \hat{b}$.
   ▶ Since $\hat{b}$ is the orthogonal projection of b onto Col A, we know that $(b - \hat{b}) \perp \mathrm{Col}\,A$
   ▶ That gives $A^\top(b - \hat{b}) = 0$.
   ▶ So now, $A^\top b - A^\top \hat{b} = 0$
   ▶ But $\hat{b} = A\hat{x}$, so $A^\top b - A^\top(A\hat{x}) = 0$.
   ▶ Thus $\hat{x}$ is the solution to $(A^\top A)\hat{x} = A^\top b$.

   Normal Equations.

We'll see presntly that $A^T A$ is a square matrix

2. ▶ If $\hat{x}$ solves $A^\top A\hat{x} = A^\top b$, then $A^\top(A\hat{x} - b) = 0$.
   ▶ So $(A\hat{x} - b) \perp \mathrm{Col}\,A$.
   ▶ So b can be decomposed as $b = \underbrace{(A\hat{x})}_{\in \mathrm{Col}\,A} + \underbrace{\big(b - A\hat{x}\big)}_{\perp \mathrm{Col}\,A}$.

   ▶ We see that $A\hat{x}$ is the best approximation for b in Col A.

Think of $A^T A \hat{x} = A^T b$ as $\underbrace{(A^T A)}_{\text{matrix}} \hat{x} = A^T b$

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

Algorithm: NORMAL EQUATIONS

Given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, to solve the Least-Squares Problem,

$Ax = b$,

- Form the $n \times n$ matrix $A^T A$, and the vector $A^\top b \in \mathbb{R}^m$.
- Solve the **normal equation**

$$(A^\top A)\hat{x} = A^\top b. \qquad (1)$$

(e.g., using Gaussian elimination.)

Theorem

The always exists a least-squares solution of "$Ax = b$".
The least-squares solutions of "$Ax = b$" are precisely the exact solutions of the **normal equation** (1)

Example

Find a least-squares solution of the system $Ax = b$ when

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}.$$

Also compute the length of the residual (=approximation error).

Note A & b come from the earlier example: find the line that best fits the points $(1, 2)$, $(2, 2)$ and $(3, 4)$.

The line is $u_1 + u_2 x = y$. 3 equations

$u_1 + (1)\, u_2 = 2$

$u_1 + (2)\, u_2 = 2$

$u_1 + (3)\, u_2 = 4$

Given $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ $b = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$.

Normal Eqns are $A^T A \hat{x} = A^T b$.

So compute $A^T A$ and $A^T b$.

$$A^T A = \underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}}_{2\times 3} \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}}_{3\times 2} = \underbrace{\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}}_{2\times 2}.$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 8 \\ 18 \end{bmatrix}$$

So Solve $\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 18 \end{bmatrix}$

So Solve $\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 8 \\ 18 \end{bmatrix}$

That is

$$\begin{aligned} 3\hat{x}_1 + 6\hat{x}_2 &= 8 \\ 6\hat{x}_1 + 14\hat{x}_2 &= 18 \end{aligned} \implies \begin{aligned} 3\hat{x}_1 + 6\hat{x}_2 &= 8 \\ 2\hat{x}_2 &= 2. \end{aligned}$$

So $\hat{x}_2 = 1$. And then we get $\hat{x}_1 = 2/3$.

So

Ans: $\hat{x} = \begin{bmatrix} 2/3 \\ 1 \end{bmatrix}$

Here the Residual is $z = A\hat{x} - b$.

So
$$z = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2/3 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} -1/3 \\ 2/3 \\ -1/3 \end{bmatrix}$$

Lenght of the residual is

$$\|z\| = \sqrt{(-1/3)^2 + (2/3)^2 + (-1/3)^2} = 0.8165$$

In case we don't get to it in class, you should find that

$$A^T A = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \quad A^T b = \begin{bmatrix} 8 \\ 18 \end{bmatrix}.$$

Solving $A^\top A\hat{x} = A^T b$ should give $\hat{x} = \begin{bmatrix} 2/3 \\ 1 \end{bmatrix}$. Then the length of the residual is

$$\|A\hat{x} - b\| = \left\| \begin{bmatrix} -1/3 \\ 2/3 \\ -1/3 \end{bmatrix} \right\| \quad A^T b = \begin{bmatrix} 8 \\ 18 \end{bmatrix} = \sqrt{2/3} = 0.8165.$$

For simplicity, the example above, and exercises below, focus on solving Least Squares Problems with three equations and two unknowns. However, the approach is much more general, and exactly the same approach works with more equations and unknowns. The only significant difference is in how one would understand what the unknowns represent – for example one could seek a function of the form $u_1 + u_2 x + u_3 x^2$ that fits the data.

Also:

- How do we know that $A^\top A$ is invertible?
- Normal matrices are so important, there are special algorithms for solving the associated systems.
- The history is fascinating too!

We can also derive the normal equations using calculus. This is because we are solving a minimization problem. So solving where the derivative of the objective function, $\|A\hat{x} - b\|$ is zero gives the equation. See the end of Chapter 10 of Linear Algebra for Data Science for more details:
https://shainarace.github.io/LinearAlgebra/leastsquares.html

At this point we switched to slides with R code in a Jupyter Notebook.