Mathematics for Computer Science
Linear Algebra

Lecture 17: Least squares

Andrei Krokhin

March 14, 2021

Reminder from the last lecture

- Orthonormal = consisting of unit vectors, which are pairwise orthogonal
- Every finite-dimensional inner product space V has an orthonormal basis, which can be constructed from any basis of V via the Gram-Schmidt process.
- For a subspace W of an inner product space V, its orthogonal complement is

$$W^{\perp} = \{\mathbf{x} \in V \mid \langle \mathbf{u}, \mathbf{x} \rangle = 0 \text{ for all } \mathbf{u} \in W\}$$

- If V (or even only W) is finite-dimensional then every vector $\mathbf{u} \in V$ can be uniquely expressed as $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$ where $\mathbf{w}_1 \in W$ and $\mathbf{w}_2 \in W^{\perp}$. Then $\mathbf{w}_1 = \mathrm{proj}_W \mathbf{u}$ is the orthogonal projection of $\mathbf{u}$ onto W.
- If matrix A has linearly independent columns then A can be decomposed as $A = QR$ where Q has orthonormal columns and R is invertible upper triangular.

Contents for today's lecture

- Least squares solutions of inconsistent linear systems
- Using QR-decomposition to find such solutions
- Application: Least squares fitting to data

Reminder: Column space of a matrix

Let $A = [\mathbf{u}_1|\mathbf{u}_2|\ldots|\mathbf{u}_n]$ be an $m \times n$ matrix with column vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

The column space of A is a subspace of $\mathbb{R}^m$ defined as

$$span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \{x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \ldots + x_n\mathbf{u}_n \mid x_1, \ldots, x_n \in \mathbb{R}\}$$

which we can re-write in matrix notation as

$$span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \{[\mathbf{u}_1|\mathbf{u}_2|\ldots|\mathbf{u}_n] \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mid \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n\}$$

and hence also as

$$span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}.$$

Solving inconsistent linear systems

- Assume that we have an inconsistent linear system $A\mathbf{x} = \mathbf{b}$.
- Can we find a vector that comes as close as possible to a being a solution?

Definition (Least Squares Problem)

Given a linear system $A\mathbf{x} = \mathbf{b}$ with m equations and n variables, find a vector $\mathbf{x}$ that minimises $||\mathbf{b} - A\mathbf{x}||$ (w.r.t. the Euclidean inner product on $\mathbb{R}^m$).

We call such a vector $\mathbf{x}$ a least squares solution to the system, the vector $\mathbf{b} - A\mathbf{x}$ is the least squares error vector, and the number $||\mathbf{b} - A\mathbf{x}||$ is the least squares error.

"Least squares" - because the norm is the (square root of the) sum of squares:
if $A\mathbf{x} = \mathbf{a}$ then $||\mathbf{b} - A\mathbf{x}|| = ||\mathbf{b} - \mathbf{a}|| = \sqrt{(b_1 - a_1)^2 + \ldots + (b_m - a_m)^2}$.

If we trust different measurements/equations differently, we can use the weighted Euclidean inner product to compute the norm and get the weighted least squares.

Best approximation theorem

Theorem

If W is a finite-dimensional subspace in an inner product space V and $\mathbf{b} \in V$ then $\mathrm{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W in the sense that

$$||\mathbf{b} - \mathrm{proj}_W \mathbf{b}|| \leq ||\mathbf{b} - \mathbf{w}||$$

for each vector $\mathbf{w} \in W$, and the inequality is strict for all $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$.

Proof.

For any vector $\mathbf{w} \in W$, write

$$\mathbf{b} - \mathbf{w} = \underbrace{(\mathbf{b} - \mathrm{proj}_W \mathbf{b})}_{\text{in } W^\perp} + \underbrace{(\mathrm{proj}_W \mathbf{b} - \mathbf{w})}_{\text{in } W}.$$

By Pythagoras' theorem (if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ then $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$), we have

$$||\mathbf{b} - \mathbf{w}||^2 = ||\mathbf{b} - \mathrm{proj}_W \mathbf{b}||^2 + ||\mathrm{proj}_W \mathbf{b} - \mathbf{w}||^2 \geq ||\mathbf{b} - \mathrm{proj}_W \mathbf{b}||^2.$$

Moreover, the inequality is strict whenever $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$. □

Least squares solutions of linear systems

- Let $W = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$ be the column space of A.
- Since $\mathrm{proj}_W\mathbf{b}$ is the best approximation to $\mathbf{b}$ from W, least squares solutions to $A\mathbf{x} = \mathbf{b}$ (i.e. vectors $\mathbf{x}$ minimising $||\mathbf{b} - A\mathbf{x}||$) are exactly solutions to

$$A\mathbf{x} = \mathrm{proj}_W\mathbf{b}.$$

- We can compute $\mathrm{proj}_W\mathbf{b}$ and solve the system, but there's a more useful way.
- The representation $\mathbf{b} = \mathbf{w}_1 + \mathbf{w}_2$ where $\mathbf{w}_1 \in W$ and $\mathbf{w}_2 \in W^\perp$ is unique, so the equation $A\mathbf{x} = \mathrm{proj}_W\mathbf{b}$ is equivalent to the condition $\mathbf{b} - A\mathbf{x} \in W^\perp$.
- The columns of A are the rows of A^T, so the condition $\mathbf{b} - A\mathbf{x} \in W^\perp$ is equivalent to $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$, which we can re-write as

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

- This is the normal equation (or normal system) associated with $A\mathbf{x} = \mathbf{b}$.

Least squares solutions of linear systems

On the previous slide, we proved the following.

Theorem

1. *For every linear system* $A\mathbf{x} = \mathbf{b}$*, the associated normal system* $A^T A\mathbf{x} = A^T\mathbf{b}$ *is consistent, and its solutions are exactly least square solutions of* $A\mathbf{x} = \mathbf{b}$*.*
2. *Moreover, if* W *is the column space of* A *and* $\mathbf{x}_0$ *is any least squares solution of* $A\mathbf{x} = \mathbf{b}$ *then* $A\mathbf{x}_0 = \mathrm{proj}_W\mathbf{b}$*.*

Example: Computing least squares solutions

Find least squares solutions for the linear system

$$\begin{array}{rcl} x_1 - x_2 & = & 4 \\ 3x_1 + 2x_2 & = & 1 \\ -2x_1 + 4x_2 & = & 3 \end{array} \qquad A\mathbf{x} = \begin{pmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix} = \mathbf{b}.$$

The associated normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{pmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix}.$$

Computing the matrix products, we get

$$\begin{pmatrix} 14 & -3 \\ -3 & 21 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 10 \end{pmatrix}.$$

Solving this yields a unique least squares solution $x_1 = 17/95$ and $x_2 = 143/285$.

(If needed, can now easily compute the error vector $\mathbf{b} - A\mathbf{x}$ and error $||\mathbf{b} - A\mathbf{x}||$)

When are least squares solutions unique?

Equivalently: when does a normal system $A^TA\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff A^TA is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- A^TA is square, so it is invertible iff $A^TA\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- Each solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^TA\mathbf{x} = \mathbf{0}$.
- Let $\mathbf{x}_0$ be a solution of $A^TA\mathbf{x} = \mathbf{0}$, i.e. $A^TA\mathbf{x}_0 = \mathbf{0}$.

  Then $A\mathbf{x}_0$ is both in the column space of A and in the null space of A^T (which are orthogonal complements of each other). Hence, $A\mathbf{x}_0 = \mathbf{0}$.
- Thus $A\mathbf{x} = \mathbf{0}$ has only the trivial solution iff the same is true for $A^TA\mathbf{x} = \mathbf{0}$.

□

Finding a unique least squares solution

Let A be an $m \times n$ matrix with linearly independent column vectors.

- From the two previous theorems, for every column $\mathbf{b} \in \mathbb{R}^m$, the system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution

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

- If W is the column space of A then

$$\mathrm{proj}_W \mathbf{b} = A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}.$$

- If $A = QR$ is a QR-decomposition (which exists under our assumption), then

$$\mathbf{x} = R^{-1} Q^T \mathbf{b} \quad \text{(or, equivalently, } R\mathbf{x} = Q^T \mathbf{b}\text{)}$$

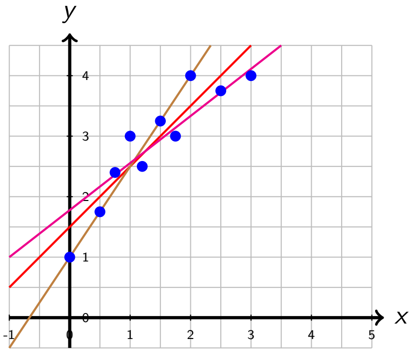
For the last item, substitute $A = QR$ into the first equation above and simplify.

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b} = ((QR)^T (QR))^{-1} (QR)^T \mathbf{b} = R^{-1} Q^T \mathbf{b}.$$

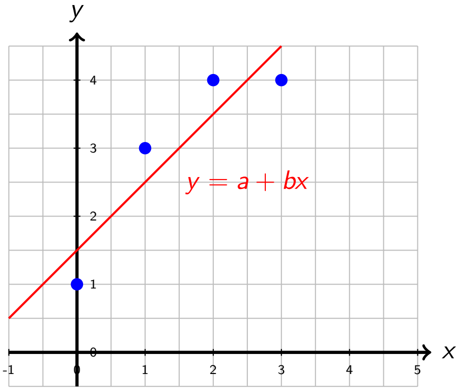
The simplification uses $Q^T Q = I$ — true because Q has orthonormal columns.

Application: Least squares fitting to data

- Assume that you want to determine, possibly approximately, the (quantitative) dependency between two parameters x and y in some process.
- You perform some experiments and get data: (x_1, y_1), $(x_2, y_2), \dots, (x_n, y_n)$.
- You want to fit a curve $y = f(x)$ in the plane to this data (to some accuracy).
- The simpler the curve, the better. So try a straight line $y = a + bx$ first. (The method extends to more than 2 variables and to more complex curves)



Example



$$a + bx_1 = y_1$$
$$a + bx_2 = y_2$$
$$a + bx_3 = y_3$$
$$a + bx_4 = y_4$$

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}$$

Least squares straight line fit

Need to find a least squares solution to the system $A\mathbf{v} = \mathbf{y}$:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}.$$

Matrix A has linearly independent columns, so least squares solution is unique.

From the associated normal system $A^T A\mathbf{v} = A^T\mathbf{y}$, we get

$$\begin{pmatrix} a \\ b \end{pmatrix} = \left(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}.$$

From this, $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}$ so $y = 1.5 + x$ is the least squares straight line fit

(aka the regression line)

What we learnt today

- Least squares solutions for inconsistent linear systems
- How to use QR-decomposition to find least squares solution
- Least squares fitting to data

Next time:

- Orthogonal matrices and spectral decomposition