CSC338 Winter 2022

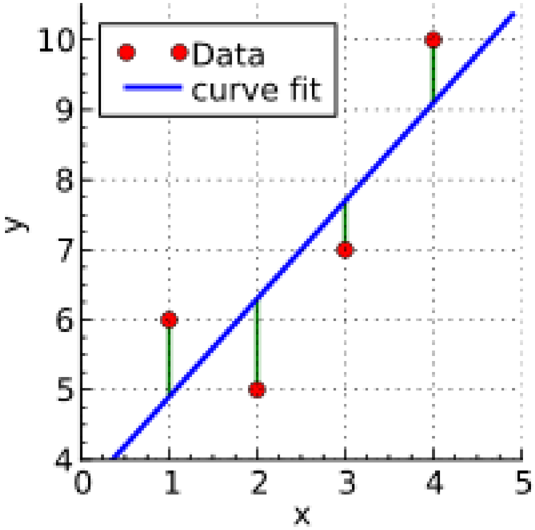
Week 6 - Linear Least Squares Part 2

Ilir Dema

University of Toronto

Feb 18, 2022

CURVE FITTING



Derivation of normal equations - Recap

- $\phi(\boldsymbol{x}) = (\boldsymbol{b} - \boldsymbol{Ax})^T(\boldsymbol{b} - \boldsymbol{Ax})$
- Require $\nabla\phi = 0$.
- $\nabla\phi = 2(\boldsymbol{A}^T\boldsymbol{Ax} - \boldsymbol{A}^T\boldsymbol{b}) = 0$.
- $H(x) = 2\boldsymbol{A}^T\boldsymbol{A}$.
- Therefore we need two conditions:
  - $\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{A}^T\boldsymbol{b}$ (normal equations)
  - $\boldsymbol{A}^T\boldsymbol{A}$ be positive definite.
  - The last condition can be guaranteed if $\boldsymbol{A}$ has full rank (that is n, assuming $\boldsymbol{A}$ is $m \times n, m > n$.
- Since $\boldsymbol{A}^T\boldsymbol{A}$ is positive definite, using Cholesky factorization, $\boldsymbol{L}^T\boldsymbol{Lx} = \boldsymbol{A}^T\boldsymbol{b}$ we can solve the system.
- If $rank(\boldsymbol{A}) < n$ the solution may not be unique.

Shortcomings of Normal Equations

Information can be lost in forming $\boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{A}^T\boldsymbol{b}$

For example, take

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix},$$

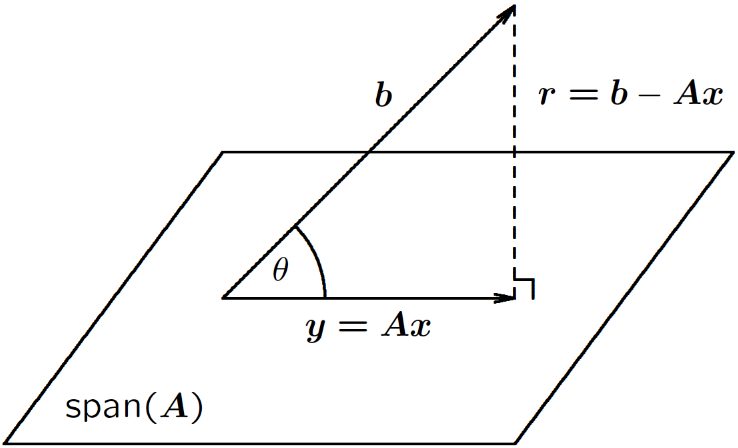
where ϵ is positive number smaller than $\sqrt{\epsilon_{\text{mach}}}$

Then in floating-point arithmetic

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 1+\epsilon^2 & 1 \\ 1 & 1+\epsilon^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

which is singular

Geometrical viewpoint



Orthogonality

- Two vectors $\boldsymbol{u}, \boldsymbol{v}$ are orthogonal (perpendicular) iff $\boldsymbol{u} \cdot \boldsymbol{v} = 0$. Write: $\boldsymbol{u} \perp \boldsymbol{v}$.
- Let U be a subspace of some finite dimensional vector space, not equal to the whole space. *Orthogonal Complement* of U is called the subspace $U^{\perp}$ with the property: for all $\boldsymbol{u} \in U, \forall \boldsymbol{w} \in U^{\perp}, \boldsymbol{u} \perp \boldsymbol{w}$.
- It is a fact that for any proper subspace V of a finite dimensional vector space W, any vector $\boldsymbol{w} \in W$ can be uniquely written as follows: $\boldsymbol{w} = \boldsymbol{v} + \boldsymbol{v}^{\boldsymbol{T}}$ where $\boldsymbol{v} \in V$ and $\boldsymbol{v}^{\boldsymbol{T}} \in V^{T}$.

Orthogonal Projectors

- Recall we want "the closest" vector to $\boldsymbol{b}$ among $\boldsymbol{Ax}$ which is nothin but the span of columns of $\boldsymbol{A}$.
- The closest vector to $\boldsymbol{b}$ in the span of $\boldsymbol{A}$ is its projection in it. **We want to know the projection operator.**

Definition

A matrix $\boldsymbol{P}$ is called *orthogonal projector* if it is idempotent: $\boldsymbol{P}^2 = \boldsymbol{P}$ and symmetric: $\boldsymbol{P}^T = \boldsymbol{P}$.

- Orthogonal projector onto orthogonal complement $span(\boldsymbol{P})^\perp$ is given by $\boldsymbol{P}^\perp = \boldsymbol{I} - \boldsymbol{P}$.
- As an immediate consequence, for any vector $\boldsymbol{v}$, we have that $\boldsymbol{v} = \boldsymbol{Pv} + (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{v}$.

Orthogonality and Least Squares Problem

Proposition

If $rank(\boldsymbol{A}) = n$ then

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$$

is the projection operator on $span(\boldsymbol{A})$ and

$$\boldsymbol{b} = \boldsymbol{P}\boldsymbol{b} + \boldsymbol{P}^{\perp}\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{y} + \boldsymbol{r}.$$

Pseudoinverse and conditioning number

Definitions

- Let $\boldsymbol{A}$ be $m \times n$ matrix. The matrix $\boldsymbol{A}^{+} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is called *pseudoinverse* of $\boldsymbol{A}$.
- Please check that $\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}^{+}$ is an orthogonal projector onto $span(\boldsymbol{A})$.
- The least squares solution $\boldsymbol{A}\boldsymbol{x} \approx \boldsymbol{b}$ is given by $\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{b}$.
- If $m > n$ for $\boldsymbol{A} \in M_{m\times n}$, $cond(\boldsymbol{A}) = \|\boldsymbol{A}\|_2\|\boldsymbol{A}^{+}\|_2$ if $rank(\boldsymbol{A}) = n$. Otherwise, $cond(\boldsymbol{A}) = \infty$.

Observe that conditioning number of $\boldsymbol{A}^T\boldsymbol{A}$ would be $\boldsymbol{(cond(A))^2}$.

Sensitivity of least linear squares

Sensitivity of $\boldsymbol{Ax} \approx \boldsymbol{b}$ dependos on both $\boldsymbol{A}, \boldsymbol{b}$.

Definition

$$\cos\theta = \frac{\|\boldsymbol{Ax}\|_2}{\|\boldsymbol{b}\|_2}$$

Proposition

$$\frac{\|\boldsymbol{\Delta x}\|_2}{\|\boldsymbol{x}\|_2} \leq \frac{\boldsymbol{cond(A)}}{\cos\Theta} \frac{\|\boldsymbol{\Delta b}\|_2}{\|\boldsymbol{b}\|_2}$$

$$\frac{\|\boldsymbol{\Delta x}\|_2}{\|\boldsymbol{x}\|_2} \leq ((\boldsymbol{cond(A)})^2 \tan\theta + \boldsymbol{cond(A)}) \frac{\|\boldsymbol{\Delta A}\|_2}{\|\boldsymbol{A}\|_2}$$

How to avoid numerical difficulties of normal equations

- Need alternative method that avoids numerical difficulties of normal equations
- Need numerically robust transformation that produces easier problem
- Observation: linear least squares is pased on orthogonal projections, dot products, and generally concepts of Euclidean norm.
- A set of transformations that preserve Euclidean distance and angles, is the set of *orthogonal transformations*.
- A square matrix $\boldsymbol{Q}$ is orthogonal if $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$.
- Multiplying both sides of least squares problem by orthogonal matrix does not change its solution.

Triangular Least Squares

- The idea is to transform $\boldsymbol{Ax} \approx \boldsymbol{b}$ to $\begin{bmatrix} \boldsymbol{R} \\ \boldsymbol{O} \end{bmatrix} \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \end{bmatrix} \approx \begin{bmatrix} \boldsymbol{b_1} \\ \boldsymbol{b_2} \end{bmatrix}$ so that $\boldsymbol{R}$ is upper triangualr and the new system has same solution as $\boldsymbol{Ax} \approx \boldsymbol{b}$.
- Then we can use backward substitution to compute $\boldsymbol{x_1}$, which means $\boldsymbol{Rx_1} = \boldsymbol{b_1}$ can be solved exactly, whereas $\boldsymbol{Ox_2} = \boldsymbol{b_2}$ generally has no solution.
- Then $\min \|\boldsymbol{Ax} - \boldsymbol{b}\|_2 = \|\boldsymbol{b_2}\|_2$.

QR Factorization

- We aim to find an orthogonal matrix $\boldsymbol{Q}$ such that $\boldsymbol{A} = \boldsymbol{Q}\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{bmatrix}$.
- Linear least squares problem $\boldsymbol{Ax} \approx \boldsymbol{b}$ transforms to $\boldsymbol{Q}^T\boldsymbol{Ax} \approx \boldsymbol{Q}^T\boldsymbol{b} = \begin{bmatrix}\boldsymbol{c}_1\\ \boldsymbol{c}_2\end{bmatrix}$
- Both problems have same solution because $\|\boldsymbol{r}\|_2^2 = \|\boldsymbol{b} - \boldsymbol{Ax}\|_2^2 = \|b - Q\begin{bmatrix}R\\ O\end{bmatrix}x\|_2^2 = \|c_1 - Rx\|_2^2 + \|c_2\|_2^2$
- Why is the above correct? Can you prove it?

Householder Transformation

- Householder transformation enables annihilation of subdiagonal entries of successive columns of $\boldsymbol{A}$, eventually reaching triangular form.
- The final QR factorization can be expressed as a composition of Householder transformations.
- Fact: product of orthogonal matrices is orthogonal (prove it!).
- Householder transform has this form:

  $$\boldsymbol{H} = \boldsymbol{I} - 2\frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}$$

  for nonzero vectors $\boldsymbol{v}$.
- Execise: show that $\boldsymbol{H}$ is orthogonal.
- How to apply it: give a vector $\boldsymbol{a}$, chose $\boldsymbol{v} = \boldsymbol{a} - \alpha\boldsymbol{e}_1$ where $\alpha = \pm\|a\|_2$ with the sign chosen appropriately to avoid cancellation.

References

Michael T. Heath
Scientific Computing (Revised Second Edition)
SIAM