Mathematics Methods for Computer Science

Instructor: Xubo Yang

SJTU-SE DALAB

Lecture

Designing and Analyzing Linear Systems

Theorist's Dilemma

"Find a nail for this really interesting hammer."

$$A\vec{x} = \vec{b}$$

Today's Lesson

Linear systems are insanely important.

Linear Regression

$$f(\vec{x}) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = \vec{a}^T \vec{x}$$

Find $\{a_1, \cdots, a_n\}$

n Experiments

$$\vec{x}^{(k)} \mapsto y^{(k)} \equiv f\left(\vec{x}^{(k)}\right)$$

$$\begin{gathered}
y^{(1)} = f\left(\vec{x}^{(1)}\right) = a_1 x_1^{(1)} + a_2 x_2^{(1)} + \cdots + a_n x_n^{(1)} \\
y^{(2)} = f\left(\vec{x}^{(2)}\right) = a_1 x_1^{(2)} + a_2 x_2^{(2)} + \cdots + a_n x_n^{(2)} \\
\vdots
\end{gathered}$$

Linear System for $\vec{a}$

$$\begin{pmatrix} - & \vec{x}^{(1)\top} & - \\ - & \vec{x}^{(2)\top} & - \\ & \vdots & \\ - & \vec{x}^{(n)\top} & - \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{pmatrix}$$

General Case

$$f(\vec{x}) = a_1 f_1(\vec{x}) + a_2 f_2(\vec{x}) + \cdots + a_m f_m(\vec{x})$$

$$\begin{pmatrix} f_1\left(\vec{x}^{(1)}\right) & f_2\left(\vec{x}^{(1)}\right) & \cdots & f_m\left(\vec{x}^{(1)}\right) \\ f_1\left(\vec{x}^{(2)}\right) & f_2\left(\vec{x}^{(2)}\right) & \cdots & f_m\left(\vec{x}^{(2)}\right) \\ \vdots & \vdots & \cdots & \vdots \\ f_1\left(\vec{x}^{(m)}\right) & f_2\left(\vec{x}^{(m)}\right) & \cdots & f_m\left(\vec{x}^{(m)}\right) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{pmatrix}$$

f can be nonlinear!

Two Important Cases

$$f(\vec{x}) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

"Vandermonde system"

$$f(x) = a cos(x + \phi)$$

Mini-Fourier

Something Fishy

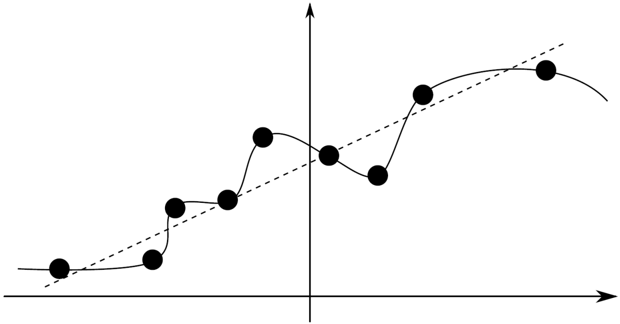
Why should you have to do exactly n experiments?

What if $y^{(k)}$ is measured with error?

Over fitting

Overfitting

Finding patterns in statistical noise



Interpretation of Linear Systems

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} - & \vec{r}_1^\top & - \\ - & \vec{r}_2^\top & - \\ \vdots & \cdots & \vdots \\ - & \vec{r}_n^\top & - \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \vec{r}_1 \cdot \vec{x} \\ \vec{r}_2 \cdot \vec{x} \\ \vdots \\ \vec{r}_n \cdot \vec{x} \end{pmatrix}$$

"Guess $\vec{x}$ by observing its dot products with $\vec{r}_i$'s."

What happens when $m > n$?

Rows are likely to be incompatible.

Next best thing:

$$A\vec{x} \approx \vec{b}$$

An over-determined least-squares problem.

Least Squares

$$A\vec{x} \approx \vec{b} \iff \min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2$$
$$\iff A^\top A\vec{x} = A^\top \vec{b}$$

Normal Equations

$$A^T A\vec{x} = A^T \vec{b}$$

$A^T A$ is the Gram matrix.

Regularization

Tikhonov regularization
("Ridge Regression;" Gaussian prior):

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 + \alpha\|\vec{x}\|_2^2$$

Lasso (Laplace prior):

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 + \beta\|\vec{x}\|_1$$

Elastic Net:

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 + \alpha\|\vec{x}\|_2^2 + \beta\|\vec{x}\|_1$$

Regularization

Tikhonov regularization
("Ridge Regression;" Gaussian prior):

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 + \alpha\|\vec{x}\|_2^2$$

Least-Squares "In the Wild"

Make lots of expressions approximately zero.

$$\sum_i [f_i(\vec{x})]^2$$

Example: Image Alignment

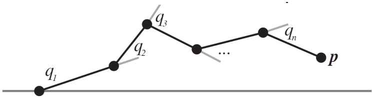
$$\vec{y}_k \approx A\vec{x}_k + \vec{b}$$

$$A \in \mathbb{R}^{2\times 2} \quad \vec{b} \in \mathbb{R}^2$$

(a) (b) Input images with keypoints (c) Aligned images

Example: Robotics

Planar Serial Chain Manipulator



Problem: How to change redundant joint angles $\vec{q}$ to move toward goal position?

- Joint angles: $\vec{q} = (q_1, q_2, \cdots, q_n)^T$
- End-effector position: $\vec{p} = \begin{pmatrix} x \\ y \end{pmatrix}$
- Kinematic model: $\vec{p} = \vec{f}(\vec{q}) \xrightarrow{\text{Linearize}} \Delta\vec{p} = J\Delta\vec{q}$
- An under-determined linear least-squares problem.
- Minimum-norm solution for $\Delta\vec{q}$ given $\Delta\vec{p}$.

A Ridiculously Important Matrix

$$A^T A$$

$A^T A$ is the Gram matrix.

Properties of $A^T A$

Symmetric

B is symmetric if $B^T = B$.

Properties of $A^T A$

Symmetric

B is symmetric if $B^T = B$.

Positive (Semi-)Definite

B is positive semidefinite if for all $\vec{x} \in \mathbb{R}^n$, $\vec{x}^T B \vec{x} \geq 0$. B is positive definite if $\vec{x}^T B \vec{x} > 0$ whenever $\vec{x} \neq \vec{0}$.

Pivoting for SPD C

Goal:

Solve $C\vec{x} = \vec{d}$ for symmetric positive definite C.

$$C = \begin{pmatrix} c_{11} & \vec{v}^\top \\ \vec{v} & \tilde{C} \end{pmatrix}$$

Forward Substitution

$$E = \begin{pmatrix} 1/\sqrt{c_{11}} & \vec{0}^\top \\ \vec{r} & I_{(n-1)\times(n-1)} \end{pmatrix}$$

Symmetry Experiment

Try post-multiplication:

$$ECE^T$$

What enabled this?

- Positive definite $\Rightarrow$ existance of $\sqrt{c_{11}}$
- Symmetry $\Rightarrow$ apply E to both sides

Cholesky Factorization

$$C = LL^T$$

Observation about Cholesky

$$L = \begin{pmatrix} L_{11} & 0 & 0 \\ \vec{\ell}_k^\top & \ell_{kk} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix}$$

$$\Downarrow$$

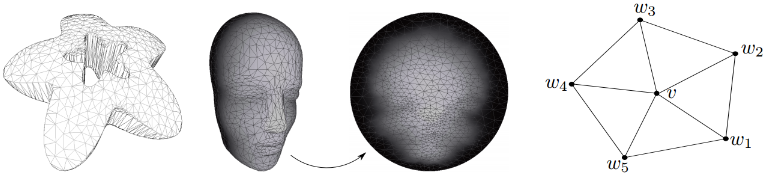
$$LL^\top = \begin{pmatrix} \times & \times & \times \\ \vec{\ell}_k^\top L_{11}^\top & \vec{\ell}_k^\top \vec{\ell}_k + \ell_{kk}^2 & \times \\ \times & \times & \times \end{pmatrix}$$

Observation about Cholesky

$$\ell_{kk} = \sqrt{c_{kk} - \left\|\vec{\ell}_k\right\|_2^2}$$

$$L_{11}\vec{\ell}_k = \vec{c}_k$$

Harmonic Parameterization



(a) Triangle mesh (b) Parameterization (c) Harmonic condition

E.g., mesh Laplacian matrices.

Storing Sparse Matrices

Want $O(n)$ storage if we have $O(n)$ nonzeros!

Examples:

- List of triplets (r,c,val)
- For each row r, matrix[r] holds a dictionary c $\rightarrow$ A[r][c]

Fill

$$\begin{pmatrix} \otimes & \times & \times & \times & \times \\ 0 & \times & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & 0 \\ 0 & 0 & 0 & \times & 0 \\ 0 & 0 & 0 & 0 & \times \end{pmatrix} \Rightarrow \begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{pmatrix}$$

Storing Sparse Matrices

- Common strategy: Permute rows/columns
- Mostly heuristic constructions
  Minimizing fill in Cholesky is NP-complete!
- Alternative strategy:
  Avoid Gaussian elimination altogether
  Iterative solution methods - only need matrix-vector multiplication! More in a few weeks.

Banded Matrices

$$\begin{pmatrix} \times & \times & & & \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & & & \times & \times \end{pmatrix}$$

Cyclic Matrices

$$\begin{pmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{pmatrix}$$