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.

Regression

(回归)

Regression: for data analysis

Example: biological experiment

Plant growth: fertilizer, sunlight, water

Goal: predict the output of $f(\vec{x})$ for a new $\vec{x}$ without carrying out the full experiment

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 $\vec{a}$

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

注意a是我们带确定的参数向量。可以将a理解为某个属性在最终结果中所占据的权重。

General Case

或许可以参考一下泰勒展开，将非线性函数在某一点展开为基本(非)线性函数的组合

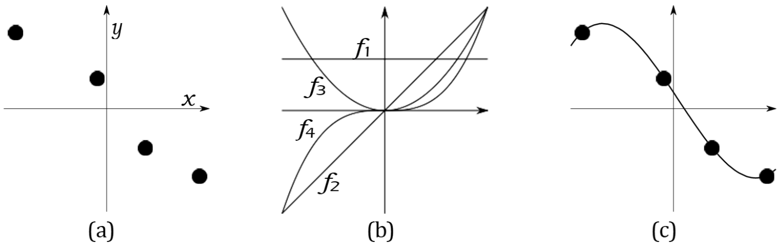
f can be nonlinear!

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

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

Key: write f as a linear combination of basis functions

General Case



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

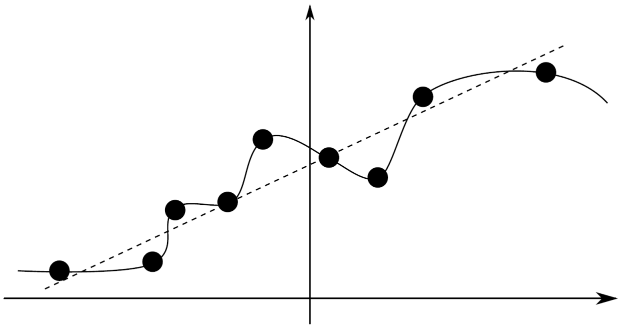
Why should you have to do exactly n experiments?

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

Drawbacks of fitting values exactly

Overfitting noisy data

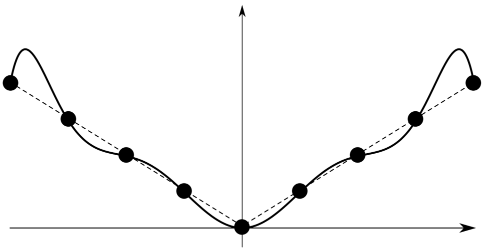
Finding patterns in statistical noise



Drawbacks of fitting values exactly

Wrong basis

Basis may not be tuned to the function sampled



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

$$\begin{aligned} A\vec{x} \approx \vec{b} &\iff \min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2 \\ &\iff \min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 \\ &\iff A^\top A\vec{x} = A^\top \vec{b} \end{aligned}$$

Minimizing residual square $\|A\vec{x} - \vec{b}\|_2^2$

$$\begin{aligned}
\|A\vec{x} - \vec{b}\|_2^2 &= (A\vec{x} - \vec{b}) \cdot (A\vec{x} - \vec{b}) \\
&= (A\vec{x} - \vec{b})^\top (A\vec{x} - \vec{b}) \\
&= \left(\vec{x}^\top A^\top - \vec{b}^\top\right)(A\vec{x} - \vec{b}) \\
&= \vec{x}^\top A^\top A\vec{x} - \vec{x}^\top A^\top \vec{b} - \vec{b}^\top A\vec{x} + \vec{b}^\top \vec{b} \\
&= \|A\vec{x}\|_2^2 - 2\left(A^\top \vec{b}\right) \cdot \vec{x} + \|\vec{b}\|_2^2
\end{aligned}$$

Minimum ($\nabla_{\vec{x}}$ must be zero)

$$\begin{aligned}
\vec{0} &= 2A^T A\vec{x} - 2A^T\vec{b} \\
\implies A^\top A\vec{x} &= A^\top \vec{b}
\end{aligned}$$

Normal Equations

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

$A^T A$ is the Gram matrix.

Normal Equations

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

$A^T A$ is the Gram matrix.

In the overdetermined case $(m > n)$, solving the least-squares problem $A\vec{x} \approx \vec{b}$ is equivalent to solving the square system $A^T A\vec{x} = A^T\vec{b}$.

Normal Equations

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

$A^T A$ is the Gram matrix.

In the overdetermined case $(m > n)$, solving the least-squares problem $A\vec{x} \approx \vec{b}$ is equivalent to solving the square system $A^T A\vec{x} = A^T\vec{b}$.

How about underdetermined case $(m < n)$?

Underdetermined case

More difficult: ambiguity, too much solutions

Add additional assumptions to get a unique solution (e.g. small norm, more zeros)

Application dependent

Methods commonly used in computer graphics, computer vision, statical analysis and machine learning

Regularization

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

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

$$0 < \alpha \ll 1$$

Regularization

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

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

$$0 < \alpha \ll 1$$

$$\implies \vec{0} = 2A^T A\vec{x} - 2A^T\vec{b} + 2\alpha\vec{x}$$
$$\implies (A^\top A + \alpha I_{n\times n})\vec{x} = A^\top\vec{b}$$

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:

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

Example

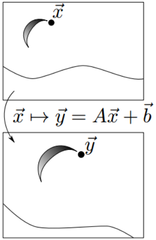
$$\begin{pmatrix} 1 & 1 \\ 1 & 1.00001 \end{pmatrix} \vec{x} = \begin{pmatrix} 1 \\ 0.99 \end{pmatrix}$$

$$\vec{x} = (1001, -1000)$$

Example: Image Alignment

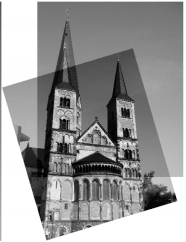
$$\vec{y}_k \approx A\vec{x}_k + \vec{b}$$

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



(a)

(b) Input images with keypoints



(c) Aligned images

Example: Image Alignment

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

$\Rightarrow$Residual: $\vec{r_k} = \vec{y}_k - A\vec{x}_k - \vec{b}$

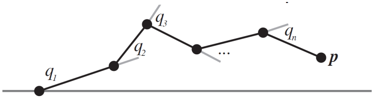
$\Rightarrow$Target: $\min_{A,b} \sum_k \|\vec{r_k}\|_2^2$

$$\Rightarrow A\vec{x} + \vec{b} = \vec{y} \quad \Rightarrow \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} A \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \Rightarrow Xp = Y$$

$\Rightarrow X^\top X p = X^\top Y \quad \Rightarrow p$ (LU, Cholesky or others)

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

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

$$\Downarrow$$

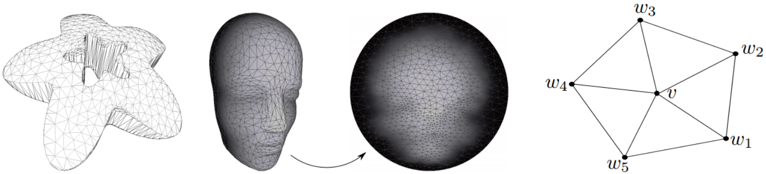
$$LL^\top = \begin{pmatrix} \times & \times & \times \\ \vec{\ell}_k^\top L_{11}^\top & \vec{\ell}_k^\top \vec{\ell}_k + \ell_{kk}^2 & \times \\ \times & \times & \times \end{pmatrix}$$

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

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

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

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