University of Idaho

Lecture 43

Least squares

General linear system: $A\vec{x} = \vec{b}$

where

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

$m \times n$

m equations

n: # of unknowns

Aside

$x_1 + 2x_2 = 5$

let $x_2 = s$: free parameter

$x_1 = 5 - 2s$

∞ solutions

Solvability conditions

1) if $n > m$ (more unknowns than equations)

- system is underdetermined
- ∞ many solutions
- can't solve

2) $n = m$

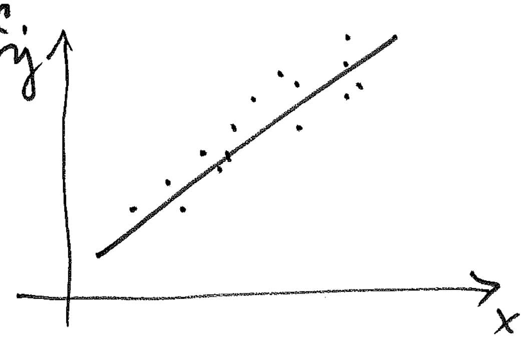
$\Rightarrow$ there is the unique solution if A is invertible
can obtain solution by Gaussian elimination or iterative methods

3) if $m > n$ (more equations than unknowns)
system is overdetermined
in general, no solution exists

$2x_1 = 5$
$x_1 = 3$

Problem

Have some data points: $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$
The data are noisy, but we believe that data obey some functional form: linear, polynomial, exp, rational function etc.



Pick a function and choose parameters to minimize the "distance" between function and data points/values.

Least squares with linear functions

Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$

Assume: $y = a_0 + a_1 x$

If data were exactly linear:

$$\begin{cases} y_1 = a_0 + a_1 x_1 \\ y_2 = a_0 + a_1 x_2 \\ \quad - \; - \; - \\ y_m = a_0 + a_1 x_m \end{cases} \Rightarrow \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_m \end{pmatrix}}_{A:\ m \times 2} \underbrace{\begin{pmatrix} a_0 \\ a_1 \end{pmatrix}}_{\vec{z}:\ 2\times 1} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}}_{\vec{b}:\ m \times 1}$$

# of equations $= m >$ # of unknowns $= 2$

$\Rightarrow$ system is overdetermined

$\Rightarrow$ no solution, in general

alternative: find $\vec{z} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$ such that

$$E = \| \underbrace{A\vec{z} - \vec{b}}_{\text{residual}} \| \text{ is minimized.}$$

Def Error measures:

1) ∞-norm $\quad E = \max_i |y_i - (a_0 + a_1 x_i)|$

2) ℓ_1-norm $\quad E = \sum_{i=1}^{m} |y_i - (a_0 + a_1 x_i)|$

3) ℓ_2-norm (squared) $\quad E = \sum_{i=1}^{m} |y_i - (a_0 + a_1 x_i)|^2$

Aside

$\sqrt{f(x)} \to \min$

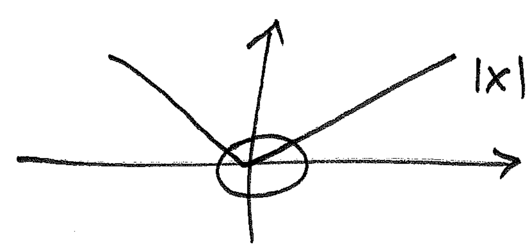
$\Updownarrow$

$f(x) \to \min$

Note

1) ∞-norm is very sensitive to outliers

2) ℓ_1-norm: E is continuous, but not differentiable at zero



* minimize l_2-norm (w/ to a_0, a_1) : $E = E(a_0, a_1)$

$$E = \sum_{i=1}^{m} \left[y_i - (a_0 + a_1 x_i)\right]^2$$

$$\frac{\partial E}{\partial a_0} = \sum_{i=1}^{m} (-2)\left[y_i - (a_0 + a_1 x_i)\right] = 0$$

$$\frac{\partial E}{\partial a_1} = -2 \sum_{i=1}^{m} x_i \left[y_i - (a_0 + a_1 x_i)\right] = 0$$

$$\Rightarrow \begin{cases} \sum_{i=1}^{m} (a_0 + a_1 x_i) = \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} (a_0 x_i + a_1 x_i^2) = \sum_{i=1}^{m} x_i y_i \end{cases}$$

or
$$a_0 m + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i$$

$$a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} x_i y_i$$

We can write this linear system (for a_0 & a_1) as

$$\begin{pmatrix} m & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \end{pmatrix}$$

Solve this 2×2 system to find a_0 and a_1

<u>Recall</u>

$A\vec{z} = \vec{b}$ where $A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_m \end{pmatrix}$, $\vec{z} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$, $\vec{b} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$

$$A^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_m \end{pmatrix}$$

transpose matrix

$A : m \times 2, \qquad A^T : 2 \times m$

$$A^T A = \underset{2 \times m}{\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_m \end{pmatrix}} \underset{m \times 2}{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}} = \underset{2 \times 2}{\begin{pmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{pmatrix}}$$

$$\therefore \quad A^T A = \begin{pmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{pmatrix}$$

$$A^T b = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_m \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{pmatrix}$$

Normal equations

$A\vec{z} = \vec{b}$: $m \times n$ system with $m \geq n$

$A^T A \vec{z} = \vec{A}^T \vec{b}$: Normal equations : $n \times n$ linear system

$$\underbrace{\overset{n\times m}{A^T}\,\overset{m\times n}{A}}_{n\times n}\vec{z} = \underbrace{\overset{n\times m}{A^T}\,\overset{m\times 1}{\vec{b}}}_{n\times 1}$$

Solution to normal equations is the solution to $A\vec{z} = \vec{b}$ in the least squares sense, meaning that the ℓ_2-norm $E = \| A\vec{z} - \vec{b} \|_2$ is minimized.

Ex find $y = a_0 + a_1 x$

x_i	1.0	1.1	1.3	1.5	1.9	2.1
y_i	1.84	1.96	2.21	2.45	2.94	3.18

$$A = \begin{pmatrix} 1 & 1.0 \\ 1 & 1.1 \\ 1 & 1.3 \\ 1 & 1.5 \\ 1 & 1.9 \\ 1 & 2.1 \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} 1.84 \\ 1.96 \\ 2.21 \\ 2.45 \\ 2.94 \\ 3.18 \end{pmatrix} \qquad \vec{z} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 6 & 8.9 \\ 8.9 & 14.17 \end{pmatrix} \qquad A^T \vec{b} = \begin{pmatrix} 14.58 \\ 22.806 \end{pmatrix}$$

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

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

Inverse of a 2×2 matrix

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow B^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\vec{z} = (A^T A)^{-1} \vec{A}^T \vec{b} = \frac{1}{5.81} \begin{pmatrix} 14.17 & -8.9 \\ -8.9 & 6 \end{pmatrix} \begin{pmatrix} 14.58 \\ 22.806 \end{pmatrix} = \begin{pmatrix} 0.6209 \\ 1.2196 \end{pmatrix} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$$

$$\therefore \boxed{y = 0.6209 + 1.2196x}$$

$$E = \sum_{i=1}^{6} [y_i - (a_0 + a_1 x_i)]^2 = 2.719 \times 10^{-5}$$

Generalization to n^{th} degree polynomial

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \quad : \; n^{th} \text{ degree polynomial}$$

where $m \geq n+1$

$$\vec{z} = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} \qquad A = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

$n+1$ components

Normal equation: $\underbrace{A^T A \vec{z} = A^T \vec{b}}_{(n+1) \times (n+1) \text{ system}}$

→ solution in the least squares sense.

Least squares w/ exponentials

$$\ln e^{\alpha x} = \alpha x$$

1) Assume $y = \beta e^{\alpha x}$

$$\underline{\ln y} = \ln \beta + \ln(e^{\alpha x}) = \ln \beta + \alpha \underline{x} \quad : \; \ln y \text{ is a linear function of } x$$