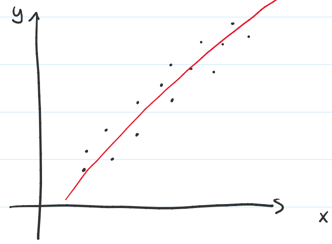
Section 8.1 discrete least squares

Monday, January 18, 2021 11:38 AM

Warm-up: Given a collection of pts $\{(x_j, y_j)\}_{j=1}^{n}$ in $\mathbb{R}^2$. How would you build a curve to go through all the points?



The pts $\{(x_j, y_j)\}_{j=1}^{n}$ is the data.

Soln: Interpolation question

Global polynomial of degree $n-1$

$$P_{n-1}(x) = \sum_{j=1}^{n} y_j \, l_j(x)$$

$$l_j(x_i) = \begin{cases} 1 & x_i = x_j \\ 0 & x_i \neq x_j \end{cases}$$

$$l_j(x) = \prod_{\substack{i=1 \\ i \neq j}}^{n} \frac{(x - x_i)}{(x_j - x_i)}$$

Other options:
- Newton Divided diff. NDD
- Monimial via Vandermonde.

+ whatever in between.

Pros: Goes through exact pts.

Cons: Possible conditioning problems., Can be wild; is it approx. anything.

What happens if the data is messy?

if the data messy you may not want to go through "exactly" all the pts. It is better for us to understand the trend.

How do we approximate "trends"?

Simple example: approximate with a line $a_0 + a_1 x$.
Our goal is to find a_0 & a_1. so that we have the "best" fit.

What do we mean by best fit? - Depends on norm.

1. minimax
minimize
$$\underbrace{E_\infty(a_0, a_1)}_{\text{our unknown constants.}} = \max_i \{ |y_i - (a_0 + a_1 x_i)| \}$$

2. absolute
minimize
$$E_1(a_0, a_1) = \sum_i |y_i - (a_0 + a_1 x_i)|$$

2. absolute minimize
$$E_1(a_0,a_1) = \sum_i |y_i - (a_0 + a_1 x_i)|$$

We focus on this

3. Least squares minimize
$$\boxed{E_2(a_0,a_1) = \left(\sum_i |y_i - (a_0 + a_1 x_i)|^2\right)^{1/2}}$$

4. General p-norm minimize
$$E_p(a_0,a_1) = \left(\sum_i |y_i - (a_0 + a_1 x_i)|^p\right)^{1/p}$$

How do we find our constants a_0 & a_1?

let's look at
$$g_2(a_0,a_1) = (E_2(a_0,a_1))^2$$
$$= \sum_i (y_i - (a_0 + a_1 x_i))^2$$

This is a quadratic upward facing. So the min happens when $\frac{dg_2}{da_0} = 0$ & $\frac{dg_2}{da_1} = 0$

expanded.

$$\frac{dg_2}{da_0} = -2 \sum_i (y_i - (a_0 + a_1 x_i)) = 2\left(a_0 \sum_i 1 + a_1 \sum_i x_i - \sum_i y_i\right) = 0$$

$$\frac{dg_2}{da_1} = -2 \sum_i x_i (y_i - (a_0 + a_1 x_i)) = 2\left(a_0 \sum_i x_i + a_1 \sum_i x_i^2 - \sum_i y_i x_i\right) = 0$$

This is a 2x2 linear system

$$\begin{bmatrix} \sum_i 1 & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i y_i x_i \end{bmatrix}$$

This is a non-singular matrix.

We can do higher order least squares fitting.

Approach: our approximation is $p_m(x) = \sum_{j=0}^{m} a_j x^j$

Then $(E_2(a_0, \cdots, a_m))^2 = \sum_i (y_i - (\sum_{j=0}^{m} a_j x_i^j))^2 = g_2$

Make linear system by equation $\frac{\partial g_2}{\partial a_j} = 0$ for $j = 0, \cdots, m$

Need $m < n$ (order of approx. less than # of data)

Alternatively we can take a linear algebra approach.

Go back to linear approx. for simplicity. $p_1(x) = a_0 + a_1 x$
let's write down the equations as if our line can go through the data.

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} a_0 \\ a_1 \end{bmatrix}}_{\bar{a}} = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{\bar{b}}$$

This is an overdetermined linear system.

It is unlikely that we will be able to get through all the data.

Finding our coefficients corresponds to finding the

$$\min \| M\bar{a} - \bar{b} \|_2 \quad \text{where} \quad \|\bar{a}\|_2^2 = a_0^2 + a_1^2 + \cdots + a_n^2$$

How do we minimize this 2-norm?

The min is the solution to the <u>normal equation</u>

$$M^* M \bar{a} = M^* b$$

(for $M \in \mathbb{R}^{n \times m}$ $M^* = M^T$
$M \in \mathbb{C}^{n \times m}$ $M^* = \overline{M}^T$)

if $M \in \mathbb{R}^{n \times m} \Rightarrow M^* M \in \mathbb{R}^{m \times m}$

We are back at a square linear system. :)

How does this relate to what we did earlier?

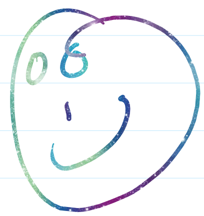
Write down the normal equation

$$\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \end{bmatrix}$$

$$\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\begin{bmatrix} \sum_i 1 & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i x_i \end{bmatrix}$$

This is the same as before.



Now we will show that the soln to the normal eqn. is the min. of $\|M\bar{a} - \bar{b}\|_2$

<u>Thm</u>: The vector $\bar{x}$ that minimizes $\|A\bar{x} - \bar{b}\|_2$ where $A \in \mathbb{R}^{m\times n}$ $m > n$ is given by the solution of

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

<u>Proof</u>:

<u>Idea (Goal)</u>: We want to show that if $A^T A\bar{x} = A^T b$

Then $\|A\bar{x} - \bar{b}\|_2 \leq \|A\bar{y} - \bar{b}\|_2 \quad \forall\, y \in \mathbb{R}^n$

Define the residuals $\bar{r}_x = A\bar{x} - \bar{b}$

$\bar{r}_y = A\bar{y} - \bar{b}$

look at $A^T r_x = A^T A\bar{x} - A^T b = \bar{0}$

look at $\bar{r}_y = A\bar{y} - \bar{b} = A\bar{y} - A\bar{x} + A\bar{x} - \bar{b} = A(\bar{y} - \bar{x}) + \bar{r}_x$

$$\|r_y\|_2^2 = (A(\bar{y}-\bar{x}) + \bar{r}_x)^T (A(\bar{y}-\bar{x}) + \bar{r}_x)$$

$$= \underbrace{(A(\bar{y}-\bar{x}))^T A(\bar{y}-\bar{x})}_{= \|A(\bar{y}-\bar{x})\|_2^2 \geq 0} + (A(\bar{y}-\bar{x}))^T \bar{r}_x + \bar{r}_x^T A(\bar{y}-\bar{x}) + \underbrace{\bar{r}_x^T \bar{r}_x}_{\|r_x\|^2}$$

$$= \|A(\bar{y}-\bar{x})\|_2^2 + (\bar{y}-\bar{x})^T \underbrace{A^T \bar{r}_x}_{=0} + \underbrace{\bar{r}_x^T}_{= (A^T r_x)^T = 0} A(\bar{y}-\bar{x}) + \|r_x\|^2$$

$$= \|A(\bar{y}-\bar{x})\|_2^2 + \|\bar{r}_x\|_2^2 \geq \|\bar{r}_x\|_2^2$$

TADA! //