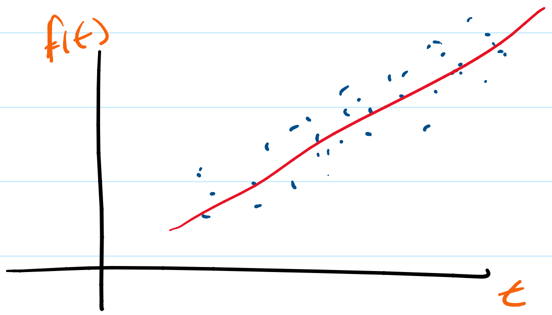
Discrete Least Squares approximations

Friday, October 20, 2023 1:45 PM

What if I have data that has been collected and I need to fit a function to it?



Do we want to do interpolation?

No don't, at best it will be wild.

How do we mathematically approximate trends?

For simplicity let's assume we are building a linear fit. $p(x) = a_0 + a_1 x$

Goal: find a_0 & a_1 to make p the best fit for the data.

What do we mean by best fit?

We look at norms.

Given data $\{(x_i, y_i)\}_{i=0}^{n}$

1- Minimax

$$\text{minimize } E_\infty(a_0,a_1) = \min \max_i \{|y_i - (a_0 + a_1 x_i)|\}$$

2- l_1-minimization - minimize the l_1-norm

$$\min E_1(a_0,a_1) = \sum_{i=0}^{n} |y_i - (a_0 + a_1 x_i)|$$

3- Least square or l_2-minization

$$\text{minimize } E_2(a_0,a_1) = \min\left(\sum_{i=0}^{n} |y_i - (a_0 + a_1 x_i)|^2\right)^{1/2}$$

We Consider the least squares problem.
How do we find a_0 & a_1?

let's consider the function

$$g_2(a_0,a_1) = (E_2(a_0,a_1))^2$$

$$= \sum_{i=0}^{n} (y_i - (a_0 + a_1 x_i))^2$$

This is an upward facing parabola.
So we look for where the derivatives $=0$.

$$0 = \frac{dg}{da_0} = -2 \sum_{i=0}^{n} (y_i - a_0 - a_1 x_i)$$

$$0 = \frac{dg}{da_0} = -2 \sum_{i=0} (y_i - a_0 - a_1 x_i)$$

$$= -2\left(\sum_{i=0}^{n} y_i - a_0 \sum_{i=0}^{n} 1 - a_1 \sum_{i=0}^{n} x_i\right)$$

$$0 = \frac{dg}{da_1} = -2 \sum_{i=0}^{n} x_i (y_i - a_0 - a_1 x_i)$$

$$= -2 \sum_{i=0}^{n} x_i y_i - a_0 \sum_{i=0}^{n} x_i - a_1 \sum_{i=0}^{n} x_i^2$$

This results in a 2x2 linear system

$$\underbrace{\begin{bmatrix} \sum_{i=0}^{n} 1 & \sum_{i=0}^{n} x_i \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 \end{bmatrix}}_{A} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} y_i x_i \end{bmatrix}$$

A is non-singular so you can invert this to find a_0 & a_1

We can do higher order least squares fitting.

Approach: write $p(x) = \sum_{i=0}^{m} a_i x^i$ most often. $m \neq n$

Then $g_2(a_0, \cdots a_m) = \sum_{i=0}^{n} \left(y_i - \left(\sum_{j=0}^{m} a_j x_i^j \right) \right)^2$

$$= \sum_{i=0}^{n} \left(y_i - \left(a_0 + a_1 x_i + \cdots + a_m x_i^m \right) \right)^2$$

Stats:

→ $n+1$ data, $m+1$ coefficients. for degree m polynomial

We need $m < n$ inorder to uniquely find the coef.

option 1 → Create our linear system by setting $\frac{\partial g}{\partial a_j} = 0$ for $j = 0, \cdots, m$

option 2 Pure brute force to go through the pts.

let's assume we have $m = 1$.

$p(x) = a_0 + a_1 x$.

$$\left.\begin{array}{l} a_0 + a_1 x_0 = y_0 \\ \vdots \\ a_0 + a_1 x_n = y_n \end{array}\right\} \to n+1 \qquad \underbrace{\begin{array}{c} m+1 \\ \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \\ 1 & x_n \end{bmatrix} \end{array}}_{M} \underbrace{\begin{bmatrix} a_0 \\ a_1 \end{bmatrix}}_{\bar{a}} = \underbrace{\begin{bmatrix} y_0 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}}_{\bar{b}}$$

This is an over determined system.

It is unlikely that we will go through all the pts. so this still a minimization problem.

Finding the coefficients is the same as minimizing

$\| M \bar{a} - \bar{b} \|_2$ where $\| \bar{a} \|_2 = \left(\sum_{i=0}^{n} |a_i|^2 \right)^{1/2}$

How do we minimize the 2-norm?

$$M^* M \bar{a} = M^* b \xrightarrow{\text{for Real matrices}} M^T M \bar{a} = M^T b$$

Def: This is the normal eqn.

What is the size of $M^T M$?

$$M \in \mathbb{R}^{(n+1)\times(m+1)} \Rightarrow M^T \in \mathbb{R}^{(m+1)\times(n+1)}$$

$$\underset{(m+1)\times(n+1)}{M^T} \; \underset{(n+1)\times(m+1)}{M} \in \mathbb{R}^{(m+1)\times(m+1)}$$

As long as $M^T M$ is invertible, we are good.

What is $\text{cond}(M^T M)$?

$$\text{cond}(M^T M) \approx (\text{cond}(M))^2$$

so be careful

How does the matrix least squares problem relate to the continuous problem?

(fingers crossed they are the same)

let's start w/ the normal eqn.

$$\begin{array}{c c} & n+1 \\ 2 & \begin{bmatrix} 1 & \cdots & 1 \\ x & \cdots & x \end{bmatrix} \end{array} \begin{array}{c} 2 \\ \begin{bmatrix} 1 & x_0 \\ \vdots & \vdots \end{bmatrix} \end{array} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ x_0 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_0 \\ \vdots \end{bmatrix}$$

$$2\begin{bmatrix} 1 & \cdots & 1 \\ x_0 & \cdots & x_n \end{bmatrix}_{n+1} \begin{bmatrix} 1 & x_0 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ x_0 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_0 \\ \vdots \\ y_1 \end{bmatrix}$$

$$\begin{bmatrix} \sum_{i=0}^{n} 1 & \sum_{i=0}^{n} x_i \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \end{bmatrix}$$

This is the same as what we had at the beginning

i.e. setting $\frac{dg}{da_j} = 0$ for $j = 0, \cdots, m$

What about the theory?

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

Idea of proof:

Goal: show that if $A^T A x = A^T b$ then

$\|Ax - b\|_2 \leq \|Ay - b\| \quad \forall y \in \mathbb{R}^m$

Define $r_x = Ax - b$ & $r_y = Ay - b$

let's look at $A^T r_x$

$$A^T r_x = A^T A x - A^T b = 0$$

let's look at r_y & rewrite it interms of r_x

$$r_y = Ay - b = Ay \underbrace{- Ax + Ax}_{\text{add zero}} - b$$

$$= Ay - Ax + r_x = A(y-x) + r_x$$

$$\|r_y\|^2 = (A(y-x) + r_x)^T (A(y-x) + r_x)$$

$$=$$

$$= \|A(y-x)\|_2^2 + \underbrace{(A(y-x))^T r_x}_{①} + \underbrace{r_x^T A(y-x)}_{②} + \|r_x\|_2^2$$

① $(A(y-x))^T r_x = (y-x)^T A^T r_x = 0$

② $r_x^T A(y-x) = (A^T r_x)^T (y-x) = 0$

$$\|r_y\|^2 = \|A(y-x)\|_2^2 + \|r_x\|_2^2 \geq \|r_x\|_2^2 \quad //$$