MATH3290 Mathematical Modeling

Tutorial 2

27th September 2017

Outline

Introduction

Given a data set (x_i, y_i), where $i = 1, \cdots, m$, and a type of model function $y = f(x)$, depending on some parameters, we want to find the model function that **best fits** the data. In general, there are 3 commonly used criteria.

1. **Chebyshev criterion:** find the parameters in the model function $f(x)$ such that the largest absolute derivation is minimized. That is, we **minimize** the following quantity

$$\max|y_i - f(x_i)|, \quad i = 1, \cdots, m.$$

Introduction (Cont.)

2. Minimize the L^1 norm: find the parameters in the model function $f(x)$ such that the following quantity is minimized. That is, we **minimize** the quantity

$$\sum_{i=1}^{m} |y_i - f(x_i)|.$$

Solving the minimization in 1. and 2. are usually difficult due to the non-differentiability of the absolute-value function

$$|\cdot| : \mathbb{R} \to \mathbb{R}^+.$$

The good news is that these problems are so-called convex minimization that there is a mature methodology to deal with.

Introduction (Cont.)

3. Least-squares method: find the parameters in $f(x)$ such that the following quantity (L^2 norm) is minimized:

$$\sum_{i=1}^{m} |y_i - f(x_i)|^2.$$

It is a very popular way because the solution of this problem can be easily obtained by calculus methods.

Assumption

Assume that the model function is $y = f(x; p_1, \cdots, p_k)$, where $p_1, \cdots, p_k$ are the model parameters. We consider the least-square problem: find the parameters $p_1, \cdots, p_k$ such that

$$S(p_1, \cdots, p_k) := \sum_{i=1}^{m} |y_i - f(x_i; p_1, \cdots, p_k)|^2$$

is minimized. In this tutorial, we introduce another equivalent way to solve the problem above.

Example I

Given the data set (x_i, y_i) with $i = 1, \cdots, m$ and assume that $f(x; p_1, p_2) = p_1 x + p_2$. We define $S(p_1, p_2)$ as follows:

$$S(p_1, p_2) = \sum_{i=1}^{m} (y_i - (p_1 x_i + p_2))^2.$$

Using the calculus method stated in lecture notes, to solve the parameters p_1 and p_2, we need to solve the following linear system:

$$\begin{aligned} \left(\sum_{i=1}^{m} x_i^2\right) p_1 + \left(\sum_{i=1}^{m} x_i\right) p_2 &= \sum_{i=1}^{m} x_i y_i, \\ \left(\sum_{i=1}^{m} x_i\right) p_1 + m p_2 &= \sum_{i=1}^{m} y_i. \end{aligned}$$

Example I (Cont.)

Write it into matrix form and we obtain

$$\begin{pmatrix} \sum_{i=1}^m x_i^2 & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & m \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m x_i y_i \\ \sum_{i=1}^m y_i \end{pmatrix}.$$

If we denote

$$A = \begin{pmatrix} x_1 & \cdots & x_m \\ 1 & \cdots & 1 \end{pmatrix}^T, \quad p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \quad \text{and } b = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

then, the system above is equivalent to

$$A^T A p = A^T b.$$

It is the so-called **normal equation**. Hence, p can be obtained by solving the normal equation.

Example II

Given the data set (x_i, y_i) with $i = 1, \cdots, m$ and assume that $f(x; p_1, p_2, p_3) = p_1 x^2 + p_2 x + p_3$. We define $S(p_1, p_2, p_3)$ as follows:

$$S(p_1, p_2, p_3) = \sum_{i=1}^{m} (y_i - (p_1 x_i^2 + p_2 x_i + p_3))^2.$$

To solve the parameters p_1, p_2 and p_3, we need to solve the following linear system:

$$\begin{pmatrix} \sum_{i=1}^m x_i^4 & \sum_{i=1}^m x_i^3 & \sum_{i=1}^m x_i^2 \\ \sum_{i=1}^m x_i^3 & \sum_{i=1}^m x_i^2 & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i^2 & \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m x_i^2 y_i \\ \sum_{i=1}^m x_i y_i \\ \sum_{i=1}^m y_i \end{pmatrix}.$$

Example II (Cont.)

If we denote

$$A = \begin{pmatrix} x_1^2 & \cdots & x_m^2 \\ x_1 & \cdots & x_m \\ 1 & \cdots & 1 \end{pmatrix}^T, \quad p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}, \quad \text{and } b = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

then the above linear system is equivalent to

$$A^T A p = A^T b.$$

We obtain the normal equation again.

Assumptions

In general, assume the model function is linear with respect to the parameters p_j, that is

$$f(x; p_1, \cdots, p_k) = \sum_{j=1}^{k} p_j g_j(x),$$

where $g_j(x)$ is a function of x (e.g. $g_j(x) = x^j$). Then, in order to solve those parameters, we do the following steps:

- Define matrix A, vectors P and b as follows:

$$A = \begin{pmatrix} g_1(x_1) & g_2(x_1) & \cdots & g_k(x_1) \\ g_1(x_2) & g_2(x_2) & \cdots & g_k(x_2) \\ \cdots & \cdots & \cdots & \cdots \\ g_1(x_m) & g_2(x_m) & \cdots & g_k(x_m) \end{pmatrix}_{m\times k},$$

$$p = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}_{k\times 1}, \quad \text{and} \quad b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}_{m\times 1}.$$

- Write down the relation $Ap = b$ and the normal equation

$$A^T Ap = A^T b.$$

- Solve the normal equation, obtains $p_1, \cdots, p_k$.

Remark

- It is easy to remember the relation $Ax = b$ in the sense that we want all the points (x_i, y_i) in the data set satisfy the model function f. For instance, if $f(x; p_1, p_2) = p_1 x + p_2$ and we have data points (x_1, y_1), (x_2, y_2) and (x_3, y_3). Then we write,

$$\begin{aligned} p_1 x_1 + p_2 &= y_1, \\ p_1 x_2 + p_2 &= y_2, \\ p_1 x_3 + p_2 &= y_3. \end{aligned}$$

- The normal equation is always solvable and as the matrix $A^T A$ is symmetric, it reduces the computational cost.
- For the non-linear model function (e.g. $y = Ce^{ax}$), we need to apply some linearized techniques to the model before we use the normal-equation approach.

Recall the practice problem: a data set $\{(x_i, y_i)\}_{i=1}^{N}$ is given. Find a parabola

$$f(x; p_0, p_1, p_2) := p_0 + p_1 x + p_2 x^2$$

such that the mean square error is minimized. Now we can you use MATLAB to solve the problem by writing down the normal equation.

```
clear;
load LSdata.mat;
pnum = 3;
N = size(x,1);
A = zeros(N,pnum);
for i=(pnum-1):-1:0
 A(:,pnum-i) = x.^i;
end
p = (A'*A)\(A'*y);
```

Practice Problem

We obtain $p_0 = 2.2348$, $p_1 = -0.7486$ and $p_2 = 3.1478$.

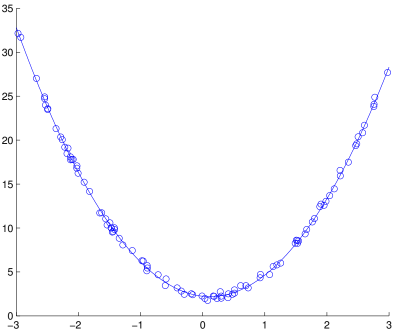

Figure: The parabola and the data.