

实验3 曲线拟合的最小二乘法

计52 路橙 2015010137

实验要求

实验题目

从随机的数据中找出其规律性，给出其近似表达式的问题，在生产实践和科学实验中大量存在，通常利用数据的最小二乘法求得拟合曲线。

1. 用最小二乘法求拟合曲线 $y = p_n(x), (n = 1, 2, 3)$ ；即分别实现线性函数拟合、二次函数拟合、三次函数拟合。
2. 打印出拟合函数 $p_n(x), (n = 1, 2, 3)$ 表达式，并打印出 $p_n(x), (n = 1, 2, 3)$ 与实际值 y 的误差；
3. 绘制出散点图和 $y = p_n(x), (n = 1, 2, 3)$ 曲线拟合图（绘图部分可以采用 matlab 等来绘制图像）。

初始数据

x	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
y	-4.467	-0.452	0.551	0.048	-0.447	0.549	4.552

算法描述

本次实验较简单，只需要计算对应内积即可。具体而言：

对于一组给定的基函数 $\Phi_1(x), \Phi_2(x), \Phi_3(x), \dots, \Phi_n(x)$ 和数据点 $(x_i, f(x_i)), i = 0, 1, \dots, m$ ，最小二乘法求 $f(x)$ 的拟合曲线 $y = \sum_{i=1}^n a_i \Phi_i(x)$ 中系数 $X = [a_1, a_2, \dots, a_n]^T$ 的计算方法为：

$$A = \begin{bmatrix} (\Phi_1(x), \Phi_1(x)), & (\Phi_1(x), \Phi_2(x)), & \dots, & (\Phi_1(x), \Phi_n(x)) \\ (\Phi_2(x), \Phi_1(x)), & (\Phi_2(x), \Phi_2(x)), & \dots, & (\Phi_2(x), \Phi_n(x)) \\ \vdots & & & \\ (\Phi_n(x), \Phi_1(x)), & (\Phi_n(x), \Phi_2(x)), & \dots, & (\Phi_n(x), \Phi_n(x)) \end{bmatrix}$$

$$D = [(\Phi_1(x), f(x)), (\Phi_2(x), f(x)), \dots, (\Phi_n(x), f(x))]^T$$

其中内积 $(f(x), g(x)) = \sum_{i=0}^m f(x_i)g(x_i)$ ，则 X 为线性方程 $AX = D$ 的解。

本次实验中，我选择函数基 $\Phi_i(x) = x^i, i = 1, 2, \dots, n$ 。

程序清单说明

实验在 1.cpp 下, ALU.h 是自己实现的线性方程求解的运算器。

使用的是标准c++库函数，直接编译即可运行。

其中定义了一系列函数指针，便于计算内积时的通用表示。

运行结果与分析

表达式：

- 线性拟合： $y = 2.00436x - 0.954464$
- 二次拟合： $y = -0.0037619x^2 + 2.00812x - 0.951643$
- 三次拟合： $y = 2.00356x^3 - 3.0091x^2 + 0.00456349x + 0.551024$

由表达式可以看出，线性拟合和二次拟合的公式几乎一样，因为二次拟合的二次项系数很小，且一次项系数、常数项系数与线性拟合的基本相等。

误差：

- 线性拟合：

x	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
误差 $y(x_i) - f(x_i)$	1.50818	-1.50464	-1.50546	-0.000285714	1.49689	1.50307	-1.49775

- 二次拟合：

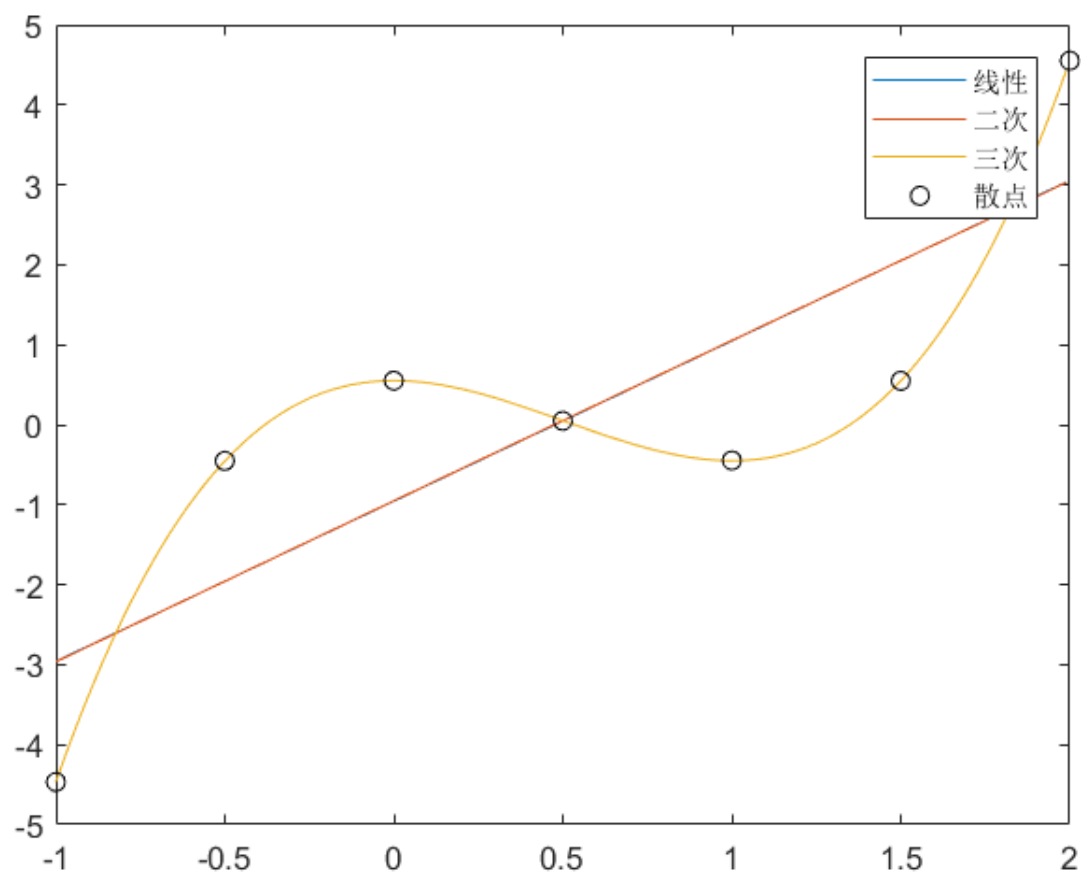
x	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
误差 $y(x_i) - f(x_i)$	1.50348	-1.50464	-1.50264	0.00347619	1.49971	1.50307	-1.50245

- 三次拟合：

x	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
误差 $y(x_i) - f(x_i)$	0.000809524	-0.00197619	0.0000238095	0.00347619	-0.00295238	0.000404762	0.000214286

由此可以看出，线性拟合与二次拟合的误差都较大，三次拟合的结果较好。

绘图：



由图像也可以清晰地看出，线性拟合与二次拟合的结果几乎重叠，而三次拟合很好地覆盖了各个散点。

体会与展望

本次实验我完成了最小二乘法的实现，基于函数指针的编程方法使得程序的实现与原理几乎完全对应。而本次数据中线性拟合与二次拟合的结果几乎相同，这是一件很神奇的事情，值得后续深入的讨论。