物理学情報処理演習
12．データ処理

2015年7月17日

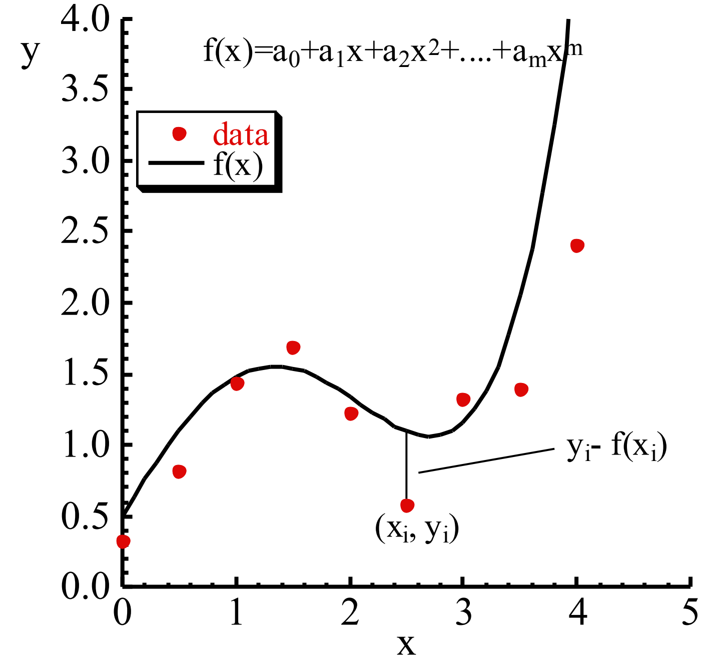
本日の推奨作業directory
lesson12

12.1　最小二乗法

参考文献
- やさしいC++ 第4版　高橋 麻奈 (著)
  ソフトバンククリエイティブ
- プログラミング言語C++第4版
  ビャーネ・ストラウストラップ, Bjarne Stroustrup, 柴田 望洋
- Numerical Recipes: The Art of Scientific Computing, Third Edition in C++

身内賢太朗
レポート提出：fsci-phys-jouhou@edu.kobe-u.ac.jp

12.1 最小二乗法



データ点が与えられているとき、データ点を最も再現する理論式$f(x)$を求める方法。

理論式: $f(x)=a_0+a_1x+a_2x^2+\ldots+a_mx^m$

とデータ点(x_i, y_i)との距離の2乗は、

$$\left(y_i - f(x_i)\right)^2$$

である。したがって各点でのそれらの総和Iは、

$$I(a_0, a_1, \ldots, a_m) = \sum_i \left(y_i - f(x_i)\right)^2 = \sum_i \left\{y_i - \left(a_0 + a_1 x_i + a_2 x_i^{\ 2} + \cdots + a_i x^m\right)^2\right\}$$

この$I(a_0,a_1,\ldots,a_m)$ を最小にする係数の組$(a_0, a_1, \ldots, a_m)$を求めることは、Iを$(a_0, a_1, \ldots, a_m)$の関数として見て、この関数の最小値を求める問題である。関数の最小値は変数の微分がゼロになる点であるから、

$$\frac{\partial I(a_0, a_1, \ldots, a_m)}{\partial a_i} = 0$$

を満足すればよい。したがって、係数の組を変数としたm個の連立方程式の解を求める問題に帰着する。

最小二乗法

m個のデータ点(x_i, y_i)を一次式の理論式$f(x)=a_0+a_1x$でフィットすることを考えよう。

理論式$f(x)=a_0+a_1x$とデータ点(x_i, y_i)との距離の2乗の総和は

$$I(a_0, a_1) = \sum_i \left(y_i - f(x_i)\right)^2 = \sum_i \left\{y_i - \left(a_0 + a_1 x_i\right)\right\}^2$$

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

この$I(a_0,a_1)$は係数(a_0, a_1)の２次式である。このIの最小値は２次式の極値であるから各係数による微分がゼロである。よって、

$$\frac{\partial I(a_0, a_1)}{\partial a_0} = 0 \qquad \frac{\partial I(a_0, a_1)}{\partial a_1} = 0$$

を満たすa_0, a_1を求めればよい。

最小二乗法:

$$\frac{\partial I(a_0, a_1)}{\partial a_0} = ma_0 + \left(\sum_i x_i\right)a_1 - \sum_i y_i = 0$$

$$\frac{\partial I(a_0, a_1)}{\partial a_1} = a_0\sum_i x_i + a_1\sum_i x_i^2 - \sum_i x_i y_i = 0$$

より理論式$f(x)=a_0+a_1x$ のa_0, a_1は

$$a_0 = \frac{\sum_i y_i \sum_i x_i^2 - \sum_i x_i \sum_i x_i y_i}{m\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \qquad a_1 = \frac{m\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{m\sum_i x_i^2 - \left(\sum_i x_i\right)^2}$$

となる。

最小二乗法

より一般的なn次多項式 $f(x)=a_0+a_1x+a_2x^2+\ldots+a_mx^m$ の場合や、n次指数関数の多項式 $f(x)= a_0+a_1e+a_2e^2+\ldots+a_me^m$ の場合でも、差の２乗

$$\left(y_i - f(x_i)\right)^2$$

は係数の組$(a_0, a_1, ..., a_m)$の２次関数である。したがって、

$$\frac{\partial I(a_0, a_1, ..., a_m)}{\partial a_i} = 0$$

は１次方程式である。$i=0\ldots m$ なので、$m+1$の連立１次方程式を解くことになる。連立１次方程式は、前出のGauss-Jordan法で解けばよい。

最小二乗法の例

右の11個のデータ点を最小二乗法で一次式:

$f(x) = a_0 + a_1 x$

でfittingするプログラムを作成し、その係数a_0, a_1を求めよ。さらに、近似した直線との差の平均とその標準偏差を求めよ。

a_0, a_1は、

$$a_0 = \frac{\sum_i y_i \sum_i x_i^2 - \sum_i x_i \sum_i x_i y_i}{m\sum_i x_i^2 - \left(\sum_i x_i\right)^2} \qquad a_1 = \frac{m\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{m\sum_i x_i^2 - \left(\sum_i x_i\right)^2}$$

であり差（残差）の平均、標準偏差は次の式で与えられる；

$$\overline{residual} = \frac{1}{n}\sum_{i=1}^{n}\left(f(x_i) - y_i\right) \qquad \sigma_{residual} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f(x_i) - y_i\right)^2}$$

x	y
0.0000	-0.27832
1.0000	3.3703
2.0000	3.7018
3.0000	4.7366
4.0000	7.6784
5.0000	12.226
6.0000	13.992
7.0000	18.176
8.0000	22.618
9.0000	25.301
10.000	28.685

レポート提出：fsci-phys-jouhou@edu.kobe-u.ac.jp
提出期限　2015年7月24日13:00（提出は任意）

ソースコードファイル名：2015_jouhou_12_学籍番号の下4桁.cxx
出力画像ファイル名：2015_jouhou_12_学籍番号の下4桁.pdf

課題12：モンテカルロ法を用いて、以下の試行実験を行なえ。

① シンチレーション検出器を用いて、ガンマ線の検出を行う。ガンマ線とシンチレータとの反応で生じる可視光が確率的に光電子に変換され、電気信号として検出される。光電子の個数は5000個/MeVであると考える。
検出される光電子の数は統計的にゆらぎ、それぞれの光電子が期待値1となるポアソン統計に従って検出されると考えることで再現できるとしてよい（例：1個の電子に対して100試行×1%の確率での電子発生を行う。）。

② 1.33MeV、1.275MeV、1.173MeV、662keV、511keVのガンマ線をそれぞれ10000事象検出した結果をそれぞれfileに出力する。

③ 横軸は検出された光子数、縦軸にその頻度を（エネルギースペクトルという）5つの結果を重ね書きする。横軸は1刻みとして、0から7000までをプロットする。

（ヒント）loopはγ線の数、光電子の個数、各電子に対するポアソン統計の計算の3層になる。いきなりガンマ線を10000発回すと時間がかかるので、最初は少ない数で試すこと。