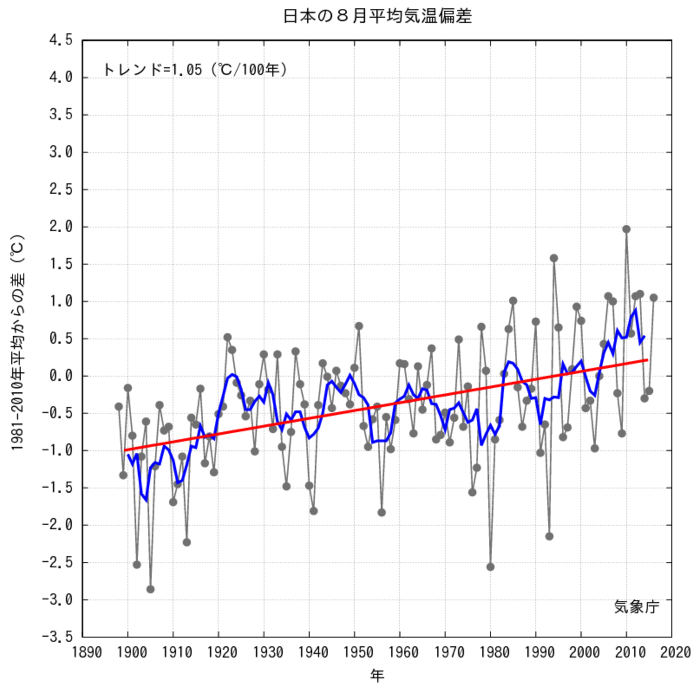
最小二乗法

目次

- 最小二乗法

最小二乗法
はじめに

- 様々なデータを解析する際, 何らかの形の関数をフィッティングしたいことがあるだろう.
- ここでは最小二乗法を使ってみよう.



http://www.data.jma.go.jp/cpdinfo/temp/an_jpn.html
downloaded at 2016/10/01.

最小二乗法
概要 (1)

- 下のようなデータがあるとする.

$$(x_1, y_1), (x_2, y_1), \dots, (x_N, y_N)$$

  – そして, これらの各データ, y_i , の分散（誤差）が下のように書けるとする.

$$\sigma_1, \sigma_2, \dots, \sigma_N$$

- このとき, このデータに対して, 下の関数をフィッティングすることを考えよう.

$$y = f(x)$$

最小二乗法
概要 (2)

- 「フィッティングする」とは, 言うなれば「よく合っているようにする」ことである.
- 「よく合っている」の基準の一つは, 二乗誤差が小さいことである.
  - つまり, 下の値が最小値を取れば「よく合っている」だろう.

$$E = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2$$

- 二乗誤差が最小値を取るのは, その微分がゼロになるときである. この条件を基に, $y = f(x)$ を求めればよい.
  - このときは, 本当は「最小値」ではなく「極小値」だけど.

最小二乗法
一次関数 (1)

- ここから先は, 具体的な例を使って考える.
- フィッティングしたい関数が一次関数の時を考える.

$$y = ax + b$$

- 二乗誤差は下のようになる.

$$E = E(a, b) = \sum_{i=1}^{N} \left(\frac{y_i - ax_i - b}{\sigma_i} \right)^2$$

- そして, 一次関数の係数を, 下の条件を基に決めればよい.

$$\frac{\partial E(a, b)}{\partial a} = 0 \qquad \frac{\partial E(a, b)}{\partial b} = 0$$

最小二乗法
一次関数 (2)

- 先の式より, 下のような, a, b の連立方程式となる.

$$\frac{\partial E(a,b)}{\partial a} = -\sum_{i=1}^{N} \frac{2x_i}{\sigma_i}\left(\frac{y_i - ax_i - b}{\sigma_i}\right)$$

$$= 2a\sum_{i=1}^{N}\frac{{x_i}^2}{{\sigma_i}^2} + 2b\sum_{i=1}^{N}\frac{x_i}{{\sigma_i}^2} - 2\sum_{i=1}^{N}\frac{x_i y_i}{{\sigma_i}^2} = 0$$

$$\frac{\partial E(a,b)}{\partial b} = -\sum_{i=1}^{N} \frac{2}{\sigma_i}\left(\frac{y_i - ax_i - b}{\sigma_i}\right)$$

$$= 2a\sum_{i=1}^{N}\frac{x_i}{{\sigma_i}^2} + 2b\sum_{i=1}^{N}\frac{1}{{\sigma_i}^2} - 2\sum_{i=1}^{N}\frac{y_i}{{\sigma_i}^2} = 0$$

最小二乗法
一次関数 (3)

- 先の式より, a, b は下のように求められる.

$$a = \frac{\left(\sum_{i=1}^{N} \frac{x_i y_i}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{1}{{\sigma_i}^2}\right) - \left(\sum_{i=1}^{N} \frac{x_i}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{y_i}{{\sigma_i}^2}\right)}{\left(\sum_{i=1}^{N} \frac{{x_i}^2}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{1}{{\sigma_i}^2}\right) - \left(\sum_{i=1}^{N} \frac{x_i}{{\sigma_i}^2}\right)^2}$$

$$b = \frac{\left(\sum_{i=1}^{N} \frac{{x_i}^2}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{y_i}{{\sigma_i}^2}\right) - \left(\sum_{i=1}^{N} \frac{x_i y_i}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{x_i}{{\sigma_i}^2}\right)}{\left(\sum_{i=1}^{N} \frac{{x_i}^2}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{1}{{\sigma_i}^2}\right) - \left(\sum_{i=1}^{N} \frac{x_i}{{\sigma_i}^2}\right)^2}$$

練習問題

1. 下の URL からデータをダウンロードして中身を確認しなさい. このデータは, 全球平均月平均気温の, 基準からのずれである. また, 温度の時間変化を図示しなさい.

   http://itpass.scitec.kobe-u.ac.jp/~yot/itpass/num_calc/data/worldtemp.data

2. 最小二乗法を使って, 上記のデータに一次関数をフィッティングしたときの係数を求めるプログラムを作りなさい. そして, 求めた一次関数とデータを図示し, フィッティングできていることを確認しなさい.
3. 2 のプログラムで求めた一次関数を用いて, 線形の経年変化を除いた温度の時間変化のデータを作りなさい.

最小二乗法
三角関数 (1)

- 最小二乗法で, 三角関数をフィッティングすることを考えよう.
$$y = f(x) = A_m \cos\left(m\left(x - x_{0,m}\right)\right)$$
- このとき, $A_m, x_{0,m}$ を最小二乗法で直接求めることは勧めない.
  - $A_m, x_{0,m}$ を求めるための方程式が非線形となり, 解くのが難しい.
    - 数値的には解けるけど.

最小二乗法
三角関数 (2)

- 代わりに, 関数を下のように変形して計算すると良い.

$$y = f(x) = A_m \cos\left(m\left(x - x_{0,m}\right)\right)$$
$$= A_m \cos(mx)\cos\left(mx_{0,m}\right)$$
$$+A_m \sin(mx)\sin\left(mx_{0,m}\right)$$
$$= B_m \cos(mx) + C_m \sin(mx)$$

- この関数は, B_m, C_m に対して線形であり, これらの値を最小二乗法で決めればよい.

最小二乗法
三角関数 (3)

- 二乗誤差は下のようになる.

$$E = E(B_m, C_m) = \sum_{i=1}^{N} \left(\frac{y_i - B_m \cos(m x_i) - C_m \sin(m x_i)}{\sigma_i} \right)^2$$

- そして, 一次関数の係数を, 下の条件を基に決めればよい.

$$\frac{\partial E(B_m, C_m)}{\partial B_m} = 0 \qquad \frac{\partial E(B_m, C_m)}{\partial C_m} = 0$$

最小二乗法
三角関数 (4)

- 先の式より, 下のような, B_m, C_m の連立方程式となる.

$$\frac{\partial E(B_m, C_m)}{\partial B_m} = -\sum_{i=1}^{N} \frac{2\cos(mx_i)}{\sigma_i} \left(\frac{y_i - B_m \cos(mx_i) - C_m \sin(mx_i)}{\sigma_i} \right)$$

$$= 2B_m \sum_{i=1}^{N} \frac{\cos^2(mx_i)}{{\sigma_i}^2} + 2C_m \sum_{i=1}^{N} \frac{\cos(mx_i)\sin(mx_i)}{{\sigma_i}^2} - 2\sum_{i=1}^{N} \frac{y_i \cos(mx_i)}{{\sigma_i}^2} = 0$$

$$\frac{\partial E(B_m, C_m)}{\partial C_m} = -\sum_{i=1}^{N} \frac{2\sin(mx_i)}{\sigma_i} \left(\frac{y_i - B_m \cos(mx_i) - C_m \sin(mx_i)}{\sigma_i} \right)$$

$$= 2B_m \sum_{i=1}^{N} \frac{\sin(mx_i)\cos(mx_i)}{{\sigma_i}^2} + 2C_m \sum_{i=1}^{N} \frac{\sin^2(mx_i)}{{\sigma_i}^2} - 2\sum_{i=1}^{N} \frac{y_i \sin(mx_i)}{{\sigma_i}^2} = 0$$

練習問題

1. 前の練習問題の 3 で作成したデータを用いて, 最小二乗法を使って, 年変化の振幅と位相を求めるプログラムを作りなさい.