

13. 一次の最小二乗法

2018/1/22

1 最小二乗法

実験やシミュレーションによって、データが得られた際に、そのデータが従う式を推定することが多くあります。最小二乗法は、そのような方式の代表的なものです。

二次元のデータ点の列 $\{x_k, y_k\} (k = 0, \dots, n-1)$ があるとし、その関数形を $y = f(x)$ と予想したとします。関数には、一般にパラメタが含まれていることに注意します。そのときに、データと予想した関数との二乗誤差

$$S = \sum_{k=0}^{n-1} [y_k - f(x_k)]^2 \quad (1.1)$$

を最小化することで、データに合う曲線 $y = f(x)$ を求めるのが最小二乗法です。

その最も簡単な形式が、予想した関数として一次関数 $y = ax + b$ を仮定するものです。このときの二乗誤差は

$$S = \sum_{k=0}^{n-1} [y_k - ax_k - b]^2 \quad (1.2)$$

となります。 S を最小化するような係数 a と b は以下のように表される。

$$a = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\sigma^2} \quad (1.3)$$

$$b = \frac{\langle x^2 \rangle \langle y \rangle - \langle xy \rangle \langle x \rangle}{\sigma^2} \quad (1.4)$$

ここで $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$ です。また

$$\langle f \rangle = \frac{1}{n} \sum_{k=0}^{n-1} f_i \quad (1.5)$$

は、サンプル平均です。

2 準備

それでは、最小二乗法に対するプログラムを作成する準備をしましょう。プロジェクト `LeastSquare` を作成します。さらに、以下の URL から Java ソースファイルを取得し、プロジェクト `LeastSquare` のソースファイルの下に置きます。また、ライブラリ `MyLib` も設定しましょう。

```
http://http://aoba.cc.saga-u.ac.jp/lecture/ModelingAndSimulation/  
javasrc/LeastSquare/src.zip
```

3 データの準備

プロジェクト `LeastSquare` のソースファイルには、二つのパッケージがあります。最初のパッケージ `data` は、最小二乗法で合わせることができるデータを生成します。クラス `DataGenerator` を実行すると、三つのデータファイルを生成します。直線で合わせることができるデータ `out-linear.txt`、指数関数で合わせることができるデータ `out-exp.txt`、そしてべき関数で合わせることができるデータ `out-pow.txt` の三つです。

課題 1 `gnuplot` をもちいて、三つのデータを図示しなさい。それぞれ、直線になるように、適切に片対数、両対数で表示しなさい。

4 データの関数形の推計

もう一つのパッケージ `leastSquare` は、データに対して最小二乗法で関数形を推計するクラスが入っています。まず、データをファイルから読む必要があります。クラス `FileRead` は、そのようなメソッドもつライブラリです。

メソッド `openReader` は、ファイル名を指定して、`BufferedReader` クラスのインスタンスを開きます。メソッド `readData` は、ファイル名を指定すると、内部でメソッド `openReader` を使って `BufferedReader` クラスのインスタンスを開き、そこから、`double` 型数値がスペース区切りで一行に二つあると仮定して、データを読み出します。各行のデータは、`xy`-座標のクラス `Point2D.Double` のインスタンスとしてリストになって戻り

ます。

クラス `LeastSquare` は、ファイル名を指定してデータを読み出し、直線を使った最小二乗法を実行します。このクラスは、前述の指数関数、べき関数にも対応します。その区別は

```
static public enum Type {  
    Normal, SemiLog, LogLog  
}
```

という `enum` 型で定義され、メソッド `fit` に引数として指定します。

ソースコード 4.1 `fitSub` メソッド

```
1 private Point2D.Double fitSub(List<Point2D.Double> d) {  
2     double ax = 0.; //x の平均  
3     double ax2 = 0.; //x^2 の平均  
4     double ay = 0.; //y の平均  
5     double axy = 0.; //x*y の平均  
6     int n = 0; //データ点の数  
7     for (Point2D.Double p : d) { //全てのデータに対して  
8         //x, x^2, y, xy の和を計算  
9  
10  
11  
12     }  
13     ax /= n;  
14     ax2 /= n;  
15     ay /= n;  
16     axy /= n;  
17     double sigma2 = ax2 - ax * ax;  
18     double a = (axy - ax * ay) / sigma2;  
19     double b = (ay * ax2 - axy * ax) / sigma2;  
20     return new Point2D.Double(a, b);  
21 }
```

最小二乗法の本体はプライベートメソッド `fitSub` です (ソースコード 4.1)。引数は `xy`-座標となったデータ点のリストです。指数関数やべき関数での推計が対象の場合には、すでに対数をとる処理がなされています。クラス `Point2D.Double` のインスタンスを `p` とすると、`p.x` が `x`-座標の値、`p.y` が `y`-座標の値になります。

課題 2 ソースコード 4.1 を完成させなさい。

クラス `LeastSquare` を実行すると、先に作った `out-linear.txt`、`out-exp.txt`、及

び `out-pow.txt` の三つを読み込み、それぞれの関数形を推計して画面に表示します。

課題 3 データとともに、それらの関数形を図示し、正しくフィットできたことを確認しなさい。