

## 1 最小二乗法 Least squares method

統計的手法において学習セットとして数値が与えられたときに、それらの数値を説明できるような数式を決定することを、回帰分析といいます。回帰分析の代表的な手法に、最小二乗法 (Least squares method) があります。

はじめての機械学習 小高 知宏 オーム社 21 頁

## 2 最小二乗法による計算式の決定

### lsm.c

```
/*
   lsm.c
   最小二乗法による計算式の決定
   統計的な機械学習の例題プログラムです
   使い方
   $./lsm <(入力ファイル)
   y = a0 + a1*x の a0 および a1 を出力します。
*/

#include <stdio.h>

#define TEXTLENGTH 4096

int main()
{
    char    text [TEXTLENGTH];
    double xi,yi;
    double sxi = 0, syi = 0, sxiyi = 0, sxi2 = 0;
    double a0, a1;
    int     n = 0;

    while(fgets(text,TEXTLENGTH,stdin) != NULL){
        if(sscanf (text, "%lf %lf", &xi, &yi) == 2){
            sxi += xi;
            syi += yi;
            sxiyi += xi * yi;
            sxi2 += xi * xi;
            ++n;
        }else{
            fprintf(stderr, "不正なデータです:%s", text);
        }
    }

    if(n > 1){
        a0 = (sxi2 * syi - sxiyi * sxi) / (n * sxi2 - sxi * sxi);
        a1 = (n * sxiyi - sxi * syi) / (n * sxi2 - sxi * sxi);
    }
}
```

```

        printf("%lf\n%lf\n", a0, a1);
    }else{
        fprintf(stderr,"データが足りません\n");
    }

    return 0;
}

```

### 3 「lsm」の操作方法

「lsm」の操作方法

```

./lsm
1 2.1
3 3.7
2.5 3.4
3.9 3.1

```

「CTL」+ 「D」で入力終了

```

2.010294
0.409502

```

$y = 2.010294 + 0.409502x$  の係数と切片が出力される。

上記の「./」は、Linuxでのプログラムを実行する時の例です。  
 ウィンドウズでは、lsm 「Enter」と入力してください。  
 「CTL」+ 「Z」で入力を終了します。

## 4 Rでグラフを書くときのスクリプトとデータ

### 4.1 スクリプト

```

#"lsm.txt"を読み込む
lsm<-read.csv("lsm.txt",header = F, sep=" ")
plot(lsm)
abline(2.01,0.40)

```

## 4.2 データファイル

1 2.1

3 3.7

2.5 3.4

3.9 3.1