最小二乘法

目次

1 はじめに

本レポートでは, 最小二乗法について述べる. 最小二乗法は, 与えられたデータに合う数理モデルやそのパラメータを決定するために, 二乗誤差を目的関数とした最小化問題を解いて推定値を得る手法である. 以下で最小二乗法で用いる種々のアルゴリズムを示す. なお, 実験で用いたプログラムは添付する別ファイルにまとめる.

1.1 最小二乗法

次の関係を満たす方程式が与えられているとする.

$$z = f(\theta, x) \tag{1}$$

ここで $z \in \mathbb{R}^m, \theta \in \mathbb{R}^n, x \in \mathbb{R}^p$ とし, $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^m$ とする. これはパラメータを θ とし, x を入力, z を出力とするシステム f であるといえる. 一般には x, z の両方に観測誤差が乗るが, ここでは出力にのみ観測誤差のある場合を考える. このとき, 誤差を $w \in \mathbb{R}^m$ とおき, 加法的な誤差を考えると, 観測値は次のように与えられる.

$$y = z + w \tag{2}$$

N 組の入力および観測値のデータ $(x_i, y_i)_{i=1}^N$ が与えられたとき, 最小二乗法を解くとは, θ についての最小値問題

$$\mathop{min}_{\theta \in \mathbb{R}^n} \sum_{i=1}^{N} ||y_i - f(\theta, x_i)||^2 \tag{3}$$

を解くことである. ここで $|| \cdot ||$ は Euclid ノルムである.

1.2 回帰問題

θ が f に対して線形である場合を考える. このとき, 適当な行列値関数 $\varphi : \mathbb{R}^p \to \mathbb{R}^{m \times n}$ が存在して,

$$f(\theta, x) = \varphi(x)\theta \tag{4}$$

となる場合を考えていることになる. ここで, $\sum_{i=1}^N \varphi(x_i)^\top \varphi(x_i)$ が正則ならば, 最適化問題 (3) の最適解は次で求められる.

$$\hat{\theta}_N = \left(\sum_{i=i}^{N} \varphi(x_i)^\top \varphi(x_i) \right)^{-1} \sum_{j=1}^{N} \varphi(x_j)^\top y_j \tag{5}$$

この $\hat{\theta}_N$ を最小二乗誤差推定値と呼ぶ. このとき, 観測誤差 w_i の共分散行列の推定値は

$$\hat{V}_N = \frac{1}{N-n} \sum_{i=1}^{N} (y_i - \varphi(x_i)\hat{\theta}_N)(y_i - \varphi(x_i)\hat{\theta}_N)^\top \tag{6}$$

で与えられる．また, もし観測誤差の共分散行列 V が既知であるならば, 推定誤差の共分散行列は

$$\mathbb{E}[(\theta-\hat{\theta}_N)(\theta-\hat{\theta}_N)^\top] = \Phi_N \sum_{i=1}^{N} \varphi_i^\top V \varphi_i \Phi_N \tag{7}$$

となる．ただし,

$$\Phi_N := \left(\sum_{i=1}^{N} \varphi_i^\top \varphi_i\right)^{-1}$$

とした．観測誤差の推定値が分からない場合は, (7) の V を $\hat{V}_N$ に置き換えればよい.
このようにして得られたパラメータ $\hat{\theta}_N$ が「よい推定」であるかを考える際, 観測誤差の確率分布が不明である場合に用いるのが決定変数とよばれる量である．決定変数は以下で与えられる.

$$C := \frac{\sum_{i=1}^{N} ||\varphi_i\hat{\theta}_N - \bar{y}||^2}{\sum_{i=1}^{N} ||y_i - \bar{y}||^2}$$

最小二乗推定値 $\hat{\theta}_N$ の決定変数は [0,1] の範囲に値を取り, その値が 1 に近いほどよい推定であるとみなされる.

1.3 重み付き最小二乗法

データの信用度が同等でないとき, それを加味して最小二乗法を解くために用いるのが重み付き最小二乗法である．重み付き最小二乗法では, 重み行列 $Q_i \geq 0,\ \ i=1,2,\ldots,N$ を用いて

$$\min_{\theta\in\mathbb{R}^n} \sum_{i=1}^{N} (y_i-\varphi_i\theta)^\top Q_i (y_i-\varphi_i\theta) \tag{8}$$

を解くことを考える．このとき, $\sum_{i=1}^{N} \varphi_i^\top Q_i \varphi_i$ が正則ならば, 最適な推定値は

$$\hat{\theta}_N = \left(\sum_{i=1}^{N} \varphi_i^\top Q_i \varphi_i\right)^{-1} \sum_{j=1}^{N} \varphi_j^\top Q_j y_j \tag{9}$$

となる.

1.4 逐次最小二乗法

データが時間経過によって順次得られるときは, 逐次最小二乗法を用いる．逐次最小二乗法は, N 個のデータから推定された $\hat{\theta}_N$ と新たなデータ (x_{N+1}, y_{N+1}) を用いて再帰的に $\hat{\theta}_{N+1}$ を求める方法である．式 (5) から

$$\begin{aligned}
\hat{\theta}_{N+1} &= \Phi_{N+1} \sum_{j=1}^{N+1} \varphi(x_j)^\top y_j \\
&= (\Phi_N^{-1} + \varphi_{N+1}^\top \varphi_{N+1})^{-1} \left(\varphi_{N+1}^\top y_{N+1} + \sum_{j=1}^{N} \varphi_j^\top y_j\right) \\
&= \left\{\Phi_N - \Phi_N \varphi_{N+1}^\top \left(I_m + \varphi_{N+1}\Phi_N\varphi_{N+1}^\top\right)^{-1} \varphi_{N+1}\Phi_N\right\} \left(\varphi_N^\top y_{N+1} + \sum_{j=1}^{N} \varphi_j^\top y_j\right) \\
&= \hat{\theta}_N + K_{N+1}\left(y_{N+1} - \varphi_{N+1}\hat{\theta}_N\right)
\end{aligned}$$

ここで,

$$K_{N+1} := \Phi_N \varphi_{N+1}^\top \left(I_m + \varphi_{N+1} \Phi_N \varphi_{N+1}^\top\right)^{-1}$$

である.
以上から,

$$\tilde{\theta}_{N+1} = \tilde{\theta}_N + \tilde{K}_{N+1} \left(y_{N+1} - \varphi_{N+1} \tilde{\theta}_N\right) \tag{10}$$

$$\tilde{K}_{N+1} = \tilde{\Phi}_N \varphi_{N+1}^\top \left(I_m + \varphi_{N+1} \tilde{\Phi}_N \varphi_{N+1}^\top\right)^{-1} \tag{11}$$

$$\tilde{\Phi}_{N+1} = \tilde{\Phi}_N - \tilde{K}_{N+1} \varphi_{N+1} \tilde{\Phi}_N \tag{12}$$

とすることで, オンラインでパラメータを更新することが可能になった.
ただし, 初期値には次の値を用いる. $\epsilon > 0$ を十分小さい正の実数として

$$\tilde{\Phi}_0 = \frac{1}{\epsilon} I_n, \quad \tilde{\theta}_0 = 0$$

このように初期値を決めると, データ数 N が十分大きいとき $\tilde{\Phi}_N \simeq \Phi_N$ となる.

1.5 Kalman フィルタ

次の差分方程式を考える.

$$\theta_k = a_k \theta_{k-1} + v_k \tag{13}$$

$$y_k = c_k \theta_k + w_k, \quad k = 1, 2, \ldots \tag{14}$$

ここで $a_k, c_k \in \mathbb{R}$ であり, $v_k, w_k in \mathbb{R}$ は互いに独立な平均 0, 分散がそれぞれ σ_v^2, σ_w^2 の Gauss 分布に従って発生し, さらに異なる時刻で独立であるとする. また, $c_k \neq 0, \ \ k = 1, 2, \ldots$ とし, θ_0 は平均 $\bar{\theta}$, 分散 P を持つ分布に従って発生するとする. いま, v_K や w_k, θ_0 は観測できないが, それらの分布は既知であり, a_k, c_k も既知であるとする. 観測値 y_k はセンサから各時刻で得られるが, θ_k は直接知ることができないとする. このとき, 次の規則で推定値 $\hat{\theta}_k$ を更新する推定則を Kalman フィルタと呼ぶ.

$$\hat{\theta}_k = a_k \hat{\theta}_{k-1} + F_k (y_k - c_k a_k \hat{\theta}_{k-1}) \tag{15}$$

$$X_k = a_k^2 V_{k-1} + \sigma_v^2 \tag{16}$$

$$V_k = X_k - \frac{\sigma_w^2 X_k}{c_k^2 X_k + \sigma_w^2} \tag{17}$$

$$F_k = \frac{c_k X_k}{c_k^2 X_k + \sigma_w^2} \tag{18}$$

Kalman フィルタはシステム制御や時系列解析, 信号処理等で広く用いられる手法である.

1.6 交互最小二乗法

交互最小二乗法は, 双線形性のある関数に対して用いる最小二乗法である. 関数 $f(\theta, x)$ が $\theta = (\alpha, \beta)^\top$ に対して双線形であるとき, 関数を改めて $g(\alpha, \beta, x)$ とおき, 次のようにパラメータを求める.

Step 1: 初期値 $\hat{\alpha}_0, \hat{\beta}_0$, 終了条件 $\epsilon > 0$ を決める. $j := 0$ とする.

Step 2-1: $\beta = \hat{\beta}_{j-1}$ に固定し, α を最小化する.

$$\hat{\alpha}_j = \underset{\theta \in \mathbb{R}^{n_1}}{\arg\min} \sum_{i=1}^{N} ||y_i - g(\alpha, \hat{\beta}_{j-1}, x_i)||^2$$

Step 2-2: $\alpha = \hat{\alpha}_j$ に固定し, β を最小化する.

$$\hat{\beta}_j = \underset{\theta \in \mathbb{R}^{n_2}}{\arg\min} \sum_{i=1}^{N} ||y_i - g(\hat{\alpha}_j, \beta, x_i)||^2$$

Step 3: $||\alpha_{j-1} - \alpha_j||^2 + ||\beta_{j-1} - \beta_j||^2 < \epsilon$ なら終了. そうでなければ, $j \leftarrow j+1$ として, Step2 へ戻る.

交互最小二乗法は局所的最適解への収束は保証されていないので, 初期値をうまく選ぶ必要がある.

1.7 K-平均法

K-平均法は, 与えられたデータを K 個のクラスタに分類する手法である.
${x_i}_{i=1}^N \subset \mathbb{R}^p$ がデータとして与えられているとする. ここで, $\nu := 1, 2, \ldots, N$ とし, ν の空でない部分集合 $\nu_1, \ldots, \nu_K$ を, $\cup_{l=1}^K \nu_l = \nu$ かつ $\nu_l \cap \nu_k = \emptyset, l \neq k$ を満たすように取る. この ν_l がクラスタを表し, そのクラスタを特徴付ける量を次で定める.

$$\mu(\nu_l) := \underset{\mu \in \mathbb{R}^p}{arg\ min} \sum_{i_l \in \nu_l} ||x_{i_l} - \mu||^2,\ l = 1, \ldots, K$$

これは, 次のようにクラスタの中心を表す.

$$\mu(\nu_l) = \frac{1}{|\nu_l|} \sum_{i_l \in \nu_l} x_{i_l}$$

ここで, クラスタは次の最小値問題の解として与えられる.

$$\underset{\nu_1, \ldots, \nu_k}{min} \sum_{l=1}^{K} \sum_{i_l \in \nu_l} ||x_{i_l} - \mu(\nu_l)||^2 \tag{19}$$

ここで

$$r_{i,l} := \begin{cases} 1, & i \in \nu_l \\ 0, & i \notin \nu_l \end{cases}$$

を導入すれば, (19) は

$$\sum_{l=1}^{K} \sum_{i=1}^{n} r_{i,l} ||x_i - \mu(\nu_l)||^2$$

となる. 各データが所属するクラスタを更新する方法を述べる. まずクラスタの中心 $\mu(\nu_l)$ を固定して μ_l と表し, $r_{i,l}$ を以下の式で更新する.

$$r_{i,l} = \begin{cases} 1, & l = \underset{k=1,\ldots,K}{\arg\min} ||x_i - \mu_k||^2 \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

次に, 新しい所属データを用いて $\mu(\nu_l)$ を計算する. 計算には以下の式を用いる.

$$\mu_l = \frac{1}{\sum_{i=1}^{N} r_{i,l}} \sum_{i=1}^{N} r_{i,l} x_i \tag{21}$$

収束判定条件 $\epsilon > 0$ について, $\max_l ||\mu_l^{j-1} - \mu_l^{j}|| < \epsilon$ を満たすまで式 (20),(21) を交互に求める. K-平均法では大域的最適解を求められる保証はなく, 初期値の選び方が重要である.

2 課題 6

各入力 $x_i \in \mathbb{R}^2,\ i = 1, 2, \ldots, 10000$ に対し, 次の式で観測データ $y_i \in \mathbb{R}$ が与えられているとする.

$$y_i = x_i^\top \theta + w_i$$

ここで観測誤差は $\mathcal{N}(0,1)$ に従って発生している (ただし, 平均の情報のみ知っており, 分散も分布も知らないものとする).

2.1 6-1

θ の最小二乗誤差推定量を, 与えられた全てのデータを使って求める. また, そのときの推定誤差共分散行列を求める. 計算の結果, θ の最小二乗誤差推定量は (1.50655081, 1.99769567) となった. また, 推定誤差共分散行列は

$$\begin{pmatrix} 9.86649107e-05 & -4.08165714e-07 \\ -4.08165714e-07 & 1.00524760e-04 \end{pmatrix}$$

となった.

2.2 6-2

用いるデータを増やしていったとき, $\hat{\theta}_N$ の各要素が収束していくことを確かめる. 結果を図 1 に示す. なお, θ の真値は (1.5,2) である.

2.3 6-3

全てのデータを用いたときの決定変数を求める. 計算の結果, 決定変数は 0.86297344 となった.

2.4 結果の考察

図 1 を見ると, $N = 100$ 程度でかなり収束していることがわかる. また, 決定変数の値から, それなりによい推定ができていることがわかる.

3 課題 7

各入力 $x_i \in \mathbb{R},\ i = 1, 2, \ldots, 10000$ に対し, 次の式で観測データ $y_i \in \mathbb{R}$ が与えられているとする.

$$y_i = \varphi(x_i)\theta + w_i,\ \varphi(x) := [1\ x\ x^2\ x^3]$$

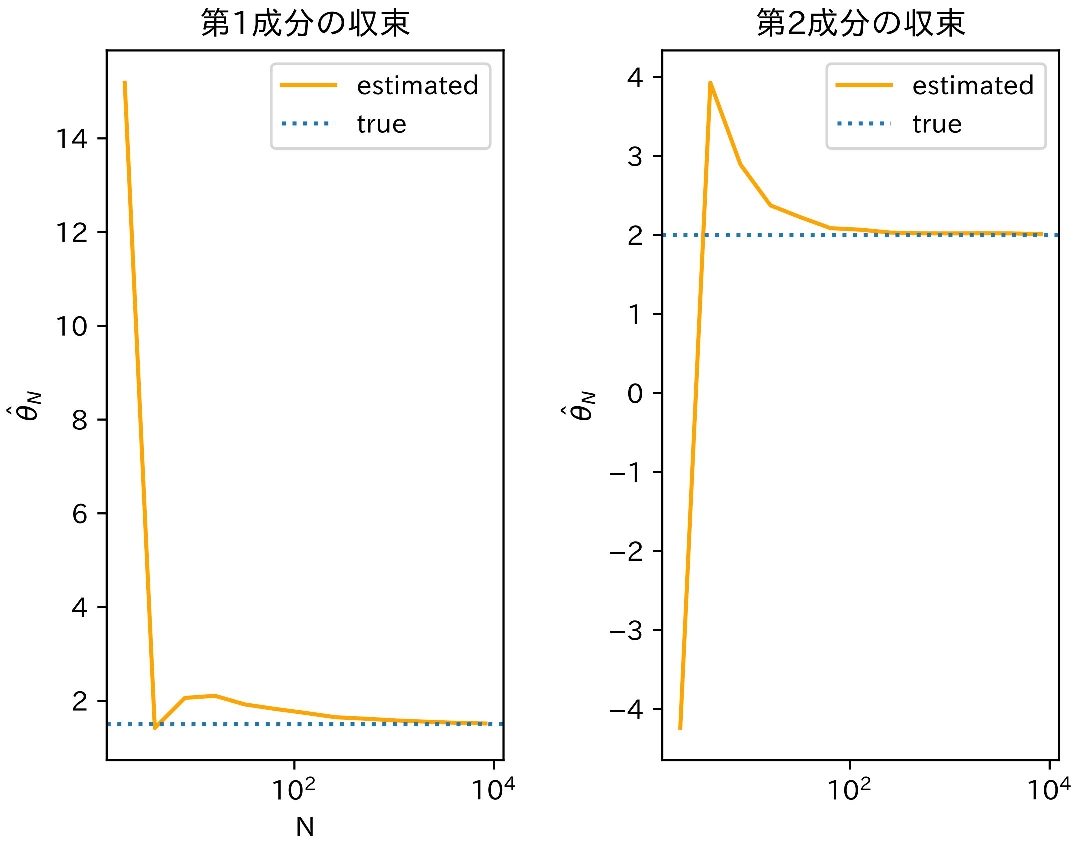

図 1: 各成分の収束

ここで観測誤差は $\mathcal{N}(0,9)$ に従って発生している (ただし, 平均の情報のみ知っており, 分散も分布も知らないものとする).

3.1 7-1

θ の最小二乗誤差推定量を, 与えられた全てのデータを使って求める. また, そのときの推定誤差共分散行列を求める. 計算の結果, θ の最小二乗誤差推定量は (-0.50902942,1.97586067,0.19774405,-0.09866691) となった. また, 推定誤差の共分散行列は

$$\begin{pmatrix} 2.02344200e-03 & -1.18649745e-05 & -1.34831217e-04 & 3.97116548e-07 \\ -1.18649745e-05 & 6.75301451e-04 & -1.89854517e-07 & -3.79471936e-05 \\ -1.34831217e-04 & -1.89854517e-07 & 1.60391021e-05 & 6.35182005e-08 \\ 3.97116548e-07 & -3.79471936e-05 & 6.35182005e-08 & 2.52871068e-06 \end{pmatrix}$$

となった.

3.2　7-2

用いるデータを増やしていったとき, $\hat{\theta}_N$ の各要素が収束していくことを確かめる．結果を図 2 に示す．なお, θ の真値は (-0.5,2,0.2,-0.1) である．

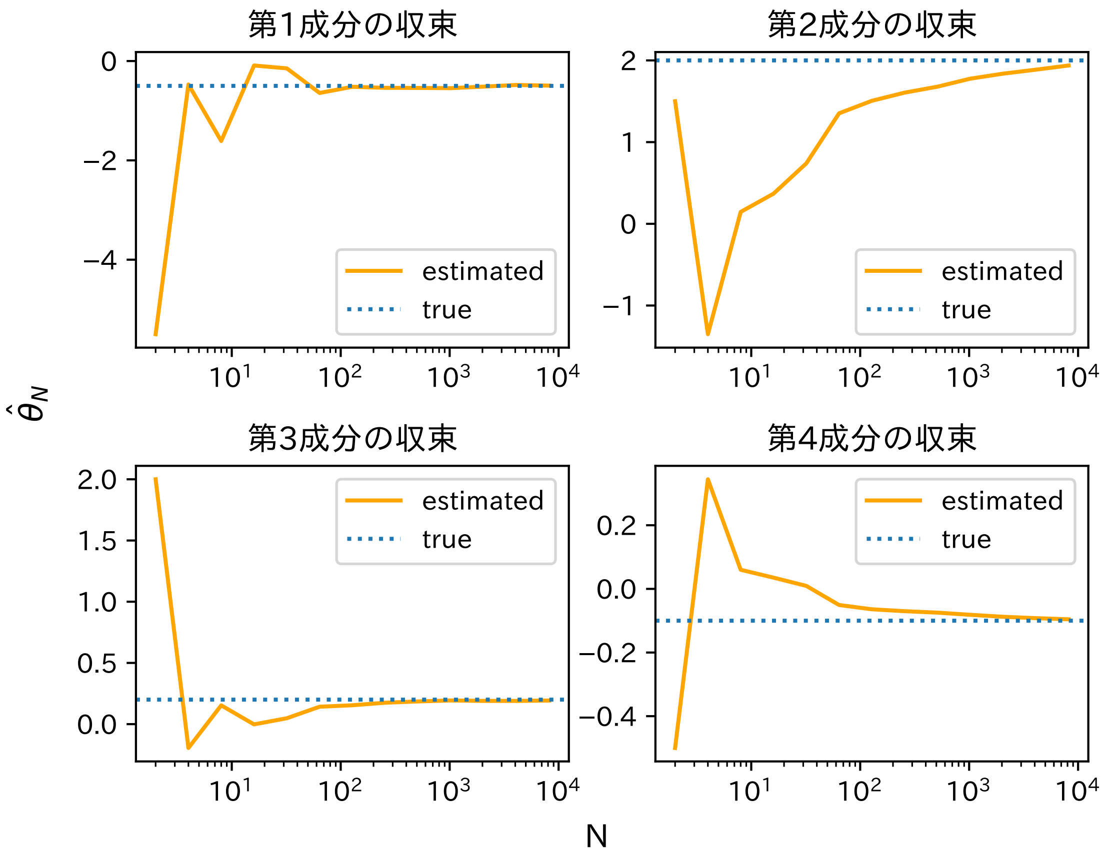

図 2: 各成分の収束

3.3　7-3

全てのデータを用いたときの決定変数を求める．計算の結果, 決定変数は 0.461855 となった．

3.4　結果の考察

図 2 を見ると, 前問と違って第二成分の収束が遅いことがわかる．また, 決定変数を見ると, あまり役に立たない推定であると言える．

4　課題 10

$x_i \in \mathbb{R}, i = 1, \ldots, 1000$ に対し, 観測値が 2 次元となる場合を考える.

$$y_i = \varphi(x_i)\theta + w_i$$

ここで w_i は $\mathcal{N}(0, V)$ の独立同一分布に従うとし,

$$\varphi = \begin{bmatrix} 1 & x \\ 1 & x^2 \end{bmatrix}, \quad V = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}$$

とする．このとき, 重みなしと重み付きのそれぞれで推定値を求め, それぞれの推定誤差共分散行列を求める. ただし, $Q_i = V$ とする．また, $\hat{\theta}_N$ の各成分の収束の仕方をプロットする.
計算の結果, 重みなしでは

$$\hat{\theta}_N = \begin{pmatrix} 2.99456713 \\ -2.06897079 \end{pmatrix},$$
$$\mathbb{E}[(\theta - \hat{\theta}_N)(\theta - \hat{\theta}_N)^\top] = \begin{pmatrix} 0.03591211 & -0.01277071 \\ -0.01277071 & 0.01108541 \end{pmatrix}$$

重み付きでは

$$\hat{\theta}_N = \begin{pmatrix} 2.93908407 \\ -1.98646467 \end{pmatrix},$$
$$\mathbb{E}[(\theta - \hat{\theta}_N)(\theta - \hat{\theta}_N)^\top] = \begin{pmatrix} 0.24522728 & -0.09870736 \\ -0.09870736 & 0.05095924 \end{pmatrix}$$

となった．また, 収束の様子を図 3 に示す．なお, θ の真値は (3,-2) である.

4.1　結果の考察

図 3 を見ると, 重みなしでは $N = 100$ 程度までは誤差がまばらに出ているが, 重み付きでは $N = 10$ 程度からよく収束しており, 重み付き最小二乗法がこのデータに適した推定方法であることがわかる.

5　課題 12

$x_i \in \mathbb{R}, i = 1, \ldots, 10000$ に対し, 次の式で観測データ $y_i \in \mathbb{R}$ が生成されているとする.

$$y_i = \varphi(x_i)\theta + w_i$$

ここで

$$\varphi(x) := \begin{bmatrix} 1 & exp\left(-\frac{(x-1)^2}{2}\right) & exp\left(-(x+1)^2\right) \end{bmatrix}$$

であり, w_i は平均 0, 分散有限の同一独立分布から生成されているものとする．最初の 6000 組と残りの 4000 組のデータそれぞれで θ の推定値を求め, それらを合成する．その後, それが全データを使った推定値と一致

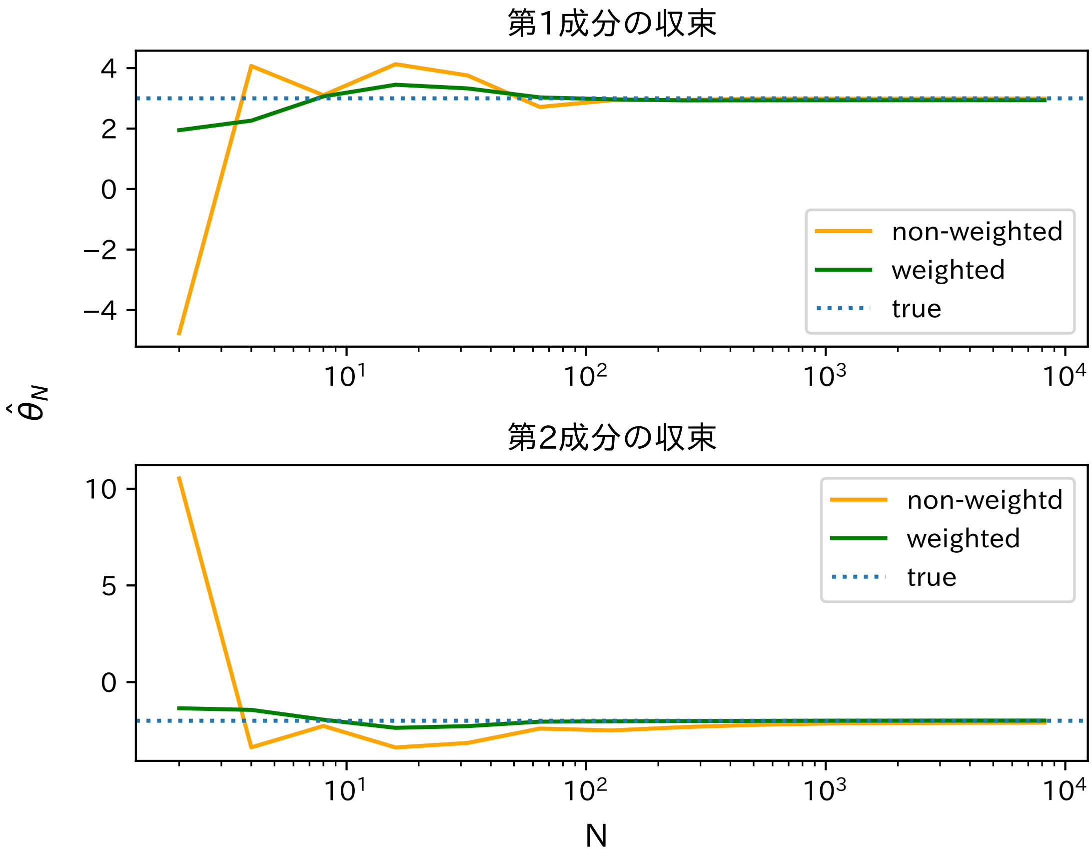

図 3: 各成分の収束

することを確かめる.

計算の結果, 最初の 6000 組では

$$\hat{\theta}_{6000} = \begin{pmatrix} -0.00387938 \\ 3.0107149 \\ -1.98943435 \end{pmatrix}$$

残りの 4000 組では

$$\hat{\theta}_{4000} = \begin{pmatrix} -0.02537589 \\ 3.03489937 \\ -1.97772731 \end{pmatrix}$$

合成すると

$$\hat{\theta}_{compounded} = \begin{pmatrix} -0.0124955 \\ 3.02042997 \\ -1.98486916 \end{pmatrix}$$

となり, 全データを用いた値

$$\hat{\theta}_{10000} = \begin{pmatrix} -0.0124955 \\ 3.02042997 \\ -1.98486916 \end{pmatrix}$$

と一致した.

6　課題 13

$x_i \in \mathbb{R}, i = 1, \ldots, 10000$ に対し, 次の式で観測データ $y_i \in \mathbb{R}$ が生成されているとする.

$$y_i = \varphi(x_i)\theta + w_i$$

ここで w_i は最初の 6000 組と残りの 4000 組のデータで, 平均 0 の異なる独立な分布に従うものとする.

$$\varphi(x) := \begin{bmatrix} 1 & exp\left(-\frac{(x-1)^2}{2}\right) & exp\left(-(x+1)^2\right) \end{bmatrix}$$

である. 10000 組のデータのうち, 最初の 6000 組と残りの 4000 組のデータそれぞれで θ の推定値と偶然誤差 w_i の分散の推定値を求め, 推定された分散を用いて推定値を合成する. その後, 全データを使って重みなしでもとめた $\hat{\theta}_{10000}$ と値を比較する.
計算の結果,

$$\hat{\theta}_{6000} = \begin{pmatrix} 0.00707684 \\ 3.28054335 \\ -2.1908997 \end{pmatrix}$$

$$\hat{\theta}_{4000} = \begin{pmatrix} 0.09848311 \\ 3.10040485 \\ -2.09100212 \end{pmatrix}$$

$$\hat{\theta}_{compounded} = \begin{pmatrix} 0.09846859 \\ 3.10043371 \\ -2.09101821 \end{pmatrix}$$

$$\hat{\theta}_{10000} = \begin{pmatrix} 0.04373209 \\ 3.20866562 \\ -2.15117856 \end{pmatrix}$$

となった. また, 真値との誤差 (ノルム) は重みなしで 0.13264064245003918, 重み付きで 0.009121725469491195 となり, 重み付きでは精度が改善されていることがわかる.

7　課題 14

次のバネマスダンパ系を表す直線上の運動方程式を考える.

$$M\frac{d^2}{dt^2}y(t) + D\frac{d}{dt}y(t) + Ky(t) = F(t)$$

ここで M, K, D は正定数で, $F(t)$ は外力である. $y(t)$ は観測でき, $F(t)$ はこちらで設計できるものとする. このとき, 得られるデータから (M, K, D) を決定したい. 微小な δt では

$$\begin{aligned} \frac{d}{dt}y(t) &\simeq \frac{y((k+1)\delta t) - y(k\delta t)}{\delta t} \\ \frac{d^2}{dt^2}y(t) &\simeq \frac{y((k+2)\delta t) - 2y((k+1)\delta t) + y(k\delta t)}{\delta t^2} \end{aligned} \tag{22}$$

と近似できることを利用し, $y_k := y(k\delta t),\ F(k) := F(k\delta t)$ とすると,

$$y_k = (2 - \frac{D}{M}\delta t)y_{k-1} + (1 + \frac{D}{M}\delta t - \frac{K}{M}\delta t^2)y_{k-2} + \frac{\delta t^2}{M}F_{k-2} + w_k$$

を得る. ここで w_k は近似の際に出る誤差である. したがって, 問題は

$$y_k = x_k^\top \theta + w_k,\ x_k := \begin{bmatrix} y_{k-1} \\ y_{k-2} \\ F_{k-2} \end{bmatrix}$$

の θ を求める問題に帰着できる. パラメータの真値を (M,D,K)=(2,1,3) とし, $\delta t = 0.01$, w_k は [-1,1] の一様分布に従う各時刻で独立な確率変数であるとして, 以下の問いに答える.

1. $y_{-2} = 0, y_{-1} = 0$ とし, $F_k = 1, k = -2, -1, \ldots$ として一定の外力を加えたとき, パラメータ θ を逐次最小二乗法で求める. ただし, 初期値 0, k=10000 までとする.
2. $y_{-2} = 0, y_{-1} = 0$ とし, $F_k = sin(\pi k/5), k = -2, -1, \ldots$ として一定の外力を加えたとき, パラメータ θ を逐次最小二乗法で求める. ただし, 初期値 0, k=10000 までとする.
3. $y_{-2} = 0, y_{-1} = 0$ とし, パラメータの推定がうまくいくように F_k を定めて実装し, 逐次最小二乗法で求める. ただし, 初期値 0, k=10000 までとする.

7.1 14-1

計算の結果, 推定値は (1.99498442 -0.99512508 -0.00352517) となり,

$$\begin{aligned} M &= -0.028367387689851437 \\ D &= -0.014227885266867711 \\ K &= -0.039901050060332136 \end{aligned}$$

となった.

7.2 14-2

計算の結果, 推定値は (1.99298678 -0.99314178 0.01133397) となり,

$$\begin{aligned} M &= 0.008823031659173442 \\ D &= 0.006187786588893572 \\ K &= 0.013675966149161915 \end{aligned}$$

となった.

7.3　14-3

誤差 w は [-1,1] の範囲で発生するので, F を非常に大きくとればその影響を小さくできる．今回は $F = 10^{10}$ とした.
計算の結果, 推定値は (1.99500000e+00 -9.95150000e-01 4.99999926e-05) となり,

$$
\begin{aligned}
M &= 2.0000002948653495 \\
D &= 1.0000000440502688 \\
K &= 3.0000000068468657
\end{aligned}
$$

となった.

7.4　結果の考察

θ の真値が (1.995 -0.99515 0.00005) であることを踏まえると, 14-2,14-2 では第三成分の推定が上手くいっておらず, 結果 M, D, K は真値からかけ離れた値になってしまっている．一方, 14-3 では非常に高精度で推定することができていることがわかる.

8　課題 16

次の離散時間ダイナミクスおよび観測方程式が与えられているとする.

$$
\begin{aligned}
\theta_k &= 0.9\theta_{k-1} + v_k \\
y_k &= 2\theta_k + w_k
\end{aligned}
$$

ここで $\theta_k, y_k \in \mathbb{R}$ であり, v_k, w_k は互いに独立な $\mathcal{N}(0,1)$ に従う確率変数である．また, θ_0 は平均 3, 分散 2 をもつ確率分布に従うとする．このとき, $y_{l}{}_{l=1}^{k}$ までを使った $\theta_k, k = 1, \ldots, 100$ の推定値 $\hat{\theta}_k$ を求め, θ_k とともに時間変化をプロットする．結果を図 4 に示す.

9　課題 18

$x_i \in [-1, 1],\ i = i, \ldots, 10000$ に対して, 次の方程式で定められる入出力データが与えられているとする.

$$
y_i = \alpha^\top \varphi(x_i)\beta + w_i
$$

ここで $\alpha \in \mathbb{R}^2, \beta \in \mathbb{R}^3$, w_i は区間 [-1,1] の一様分布であるとし,

$$
\varphi(x) = \begin{bmatrix} 1 & x & x^2 \\ x & x^2 & x^3 \end{bmatrix}
$$

である場合を考える．10000 組のデータから, 交互最小二乗法を用いて α, β を推定する．ここで, 初期値は各成分を区間 [-3,3] の一様分布からランダムに取った値を 10 個用意する．それぞれの初期値で交互最小二乗法を用い, 結果と真値との二乗誤差の和が最も小さかったものを採用する.
計算の結果, 初期値 $\alpha_0 = (-1.87581231\ \ 2.78280854), \beta_0 = (0.58598988\ \ -2.60494637\ \ -0.43212561)$ で $\alpha = (1.03692569\ \ -2.07856709), \beta = (0.48800822\ \ -0.96193953\ \ 1.924057)$ となり, それなりによい推定ができた.

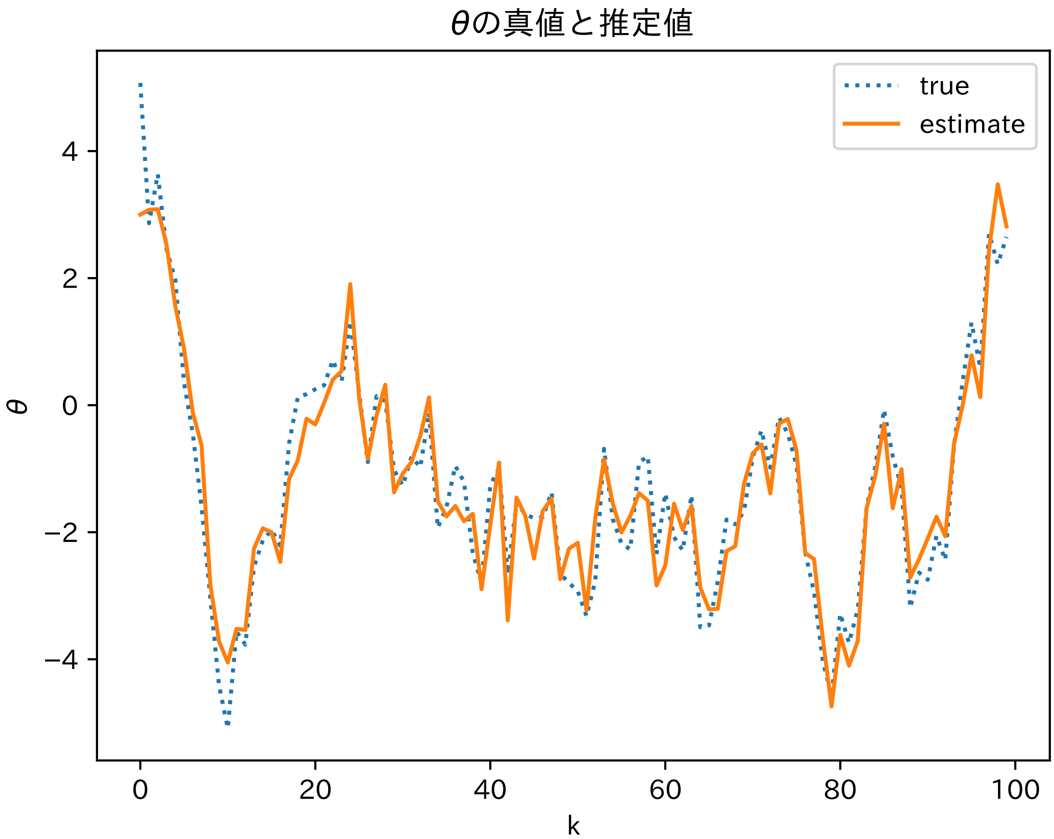

図 4: θ の真値と推定値

10 課題 19

与えられたデータ $x_i \in \mathbb{R}^2, i = 1, \ldots, 1000$ を 3 つのクラスタに分類する．ここで, 初期値は重心が (0,0) の三角形を 1 倍から 3 倍まで 0.1 倍刻みで回転させながら拡大していく形で 21 個取る．それぞれの初期値でクラスタリングを行い, 各クラスタに属する点とそれが属するクラスタとの距離の総和が最も小さいものを採用する．

計算の結果, 初期値は (-2.59807621, -1.5)(2.59807621, -1.5)(9.18485099e-16, 3.00000000e+00) が選ばれ, 散布図は図 5 のようになった．

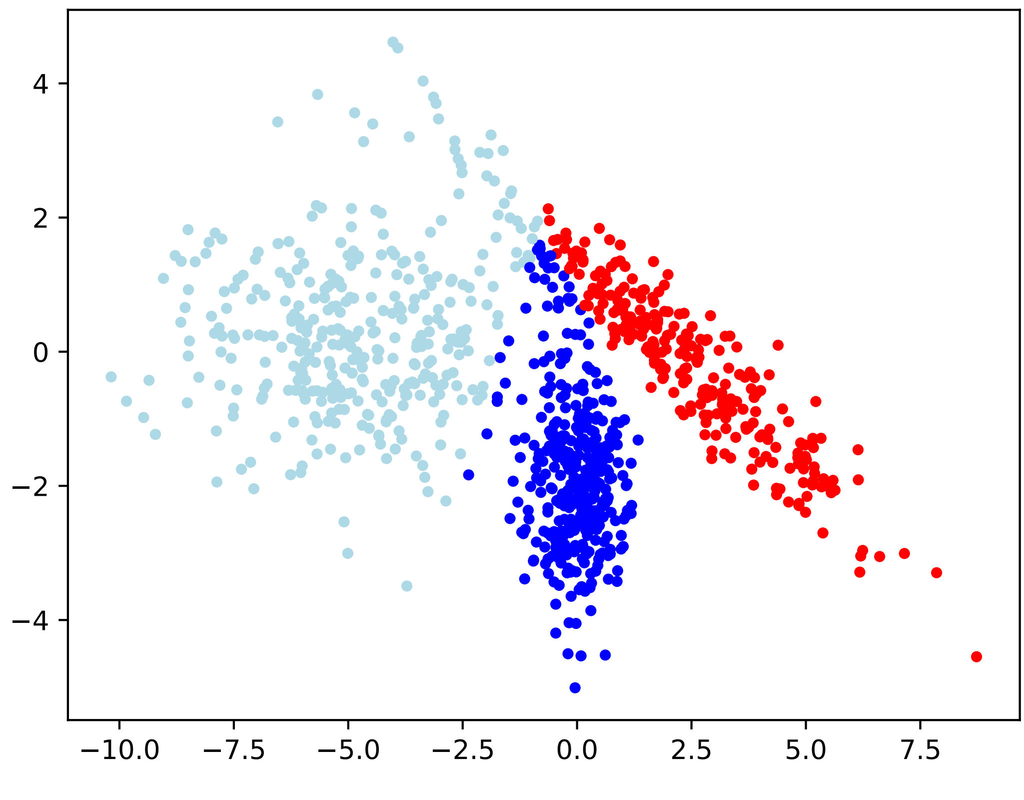

図 5: クラスタリングの結果

11　参考文献

参考文献

[1] 実験演習ワーキンググループ (2022),「数理工学実験テキスト_2022 年版」