

最小二乗法：各データの説明

1. 各課題のデータについて

※ 下記の課題番号は通し番号のため、ずれることがあります.

課題 8 (重回帰問題)

mmse_kadai1.txt または mmse_kadai1.xls を使う

データは, 10000 行 3 列で, 最初の 2 列が $x=(x_1, x_2)$, 3 列目が y .

$\theta=(1.5, 2)$ が真値.

課題 9 (多項式回帰)

mmse_kadai2.txt または mmse_kadai2.xls を使う

データは, 10000 行 2 列で, 1 列目が x , 2 列目が y .

$\theta=(-0.5, 2, 0.2, -0.1)$ が真値.

課題 10 (分散 ∞ の観測誤差の場合)

mmse_kadai3.txt または mmse_kadai3.xls を使う

データは, 10000 行 2 列で, 最初の 1 列目が x , 2 列目が y .

$\theta=(1.5, 2)$ が真値.

課題 11 (多項式回帰, 範囲限定)

mmse_kadai4.txt または mmse_kadai4.xls を使う

データは, 1000 行 2 列で, 最初の 1 列目が x , 2 列目が y .

$\theta=(-0.5, 2, 0.2, -0.1)$ が真値.

課題 12 (重み付き最小二乗法 1)

mmse_kadai5.txt または mmse_kadai5.xls を使う

データは, 1000 行 3 列で, 最初の 1 列目が x , 2, 3 列目が $y=(y_1, y_2)$.

真値: $\theta=(3, -2)$

課題 13 (重み付き最小二乗法 2)

mmse_kadai6.txt または mmse_kadai6.xls を使う

データは、1000 行 2 列で、最初の 1 列目が x , 2, 3 列目が $y=(y_1, y_2)$.
真値: $\theta=(3,-2)$

課題 14 (推定値の合成)

mmse_kadai7.txt または mmse_kadai7.xls を使う

データは、10000 行 2 列で、最初の 1 列目が x , 2 列目が y .
真値: $\theta=(0,3,-2)$

課題 15 (異なる雑音の場合の推定値の合成)

mmse_kadai8.txt または mmse_kadai8.xls を使う

データは、10000 行 2 列で、最初の 1 列目が x , 2 列目が y .
真値: $\theta=(0.1, 3.1, -2.1)$

課題 16~20 (システム同定~Kalman スムーザ)

データは自分で作成し、作成したデータを元にパラメータ推定を行う. Bernolli 過程であることは重要ではないので、平均 0 で分散有限の確率分布から生成するのもよい.

課題 20 (交互最小二乗法)

mmse_kadai13.txt または mmse_kadai13.xls を使う

データは、10000 行 2 列で、最初の 1 列目が x , 2 列目が y .
真値: $\alpha=(1, -2)$, $\beta=(0.5, -1, 2)$

課題 21 (K-クラスタリング)

mmse_kadai14.txt または mmse_kadai14.xls を使う

データは、1000 行 2 列で、最初の 1 列目が x , 2 列目が y .

データの生成については、次の 3 つの Gauss 分布からランダムに 1000 の 2 次元ベクトルを生成.

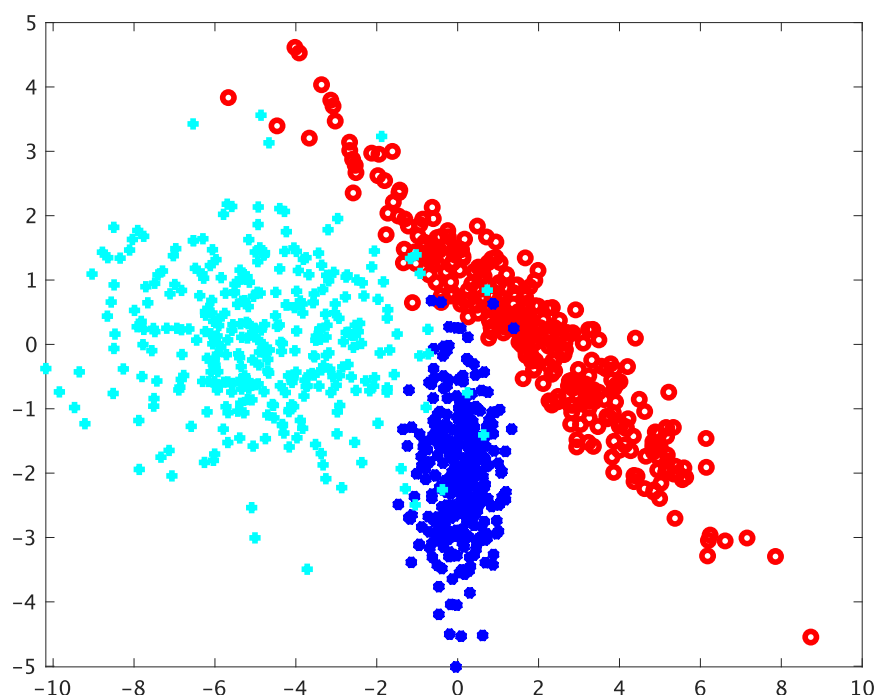
$N(\mu_1, V_1)$: $\mu_1 = (2; 0)$, $V_1 = [2, -1; -1, 1]$

$N(\mu_2, V_2)$: $\mu_2 = (0; -2)$, $V_2 = [0.5, 0; 0, 1]$

$N(\mu_3, V_3)$: $\mu_3 = (-5, 0)$, $V_3 = [2, 0; 0, 1]$

データは、 1000×2 のベクトルで、各要素がどの分布から生成されたかの正解データは mmse_kadai14_hosoku に保存されている（1 列目が $N(\mu_1, V1)$ ，2 列目が $N(\mu_2, V2)$ ，3 列目は $N(\mu_3, V3)$ で、要素に 0 が入っているのは保存する際に同じ長さに揃える必要があったため）。

色分けして可視化すると、下図になる。



2. 発展課題

発展課題 1

hatten_kadai1.txt または hatten_kadai1.xls を使う

データは、1000 行 2 列で、最初の 1 列目が x ，2 列目が y 。

発展課題 2

hatten_kadai2.txt または hatten_kadai2.xls を使う

データは、20001 行 2 列で、最初の 1 列目が ϕ ，2 列目が F 。

発展課題 3

hatten_kadai3.txt を使う（データ数が多すぎて Excel では扱えない）

データは，500001 行 2 列で，最初の 1 列目が t （時間），2 列目が y .