最適化理論

カーブフィッティング、回帰、正則化

カーブフィッティング

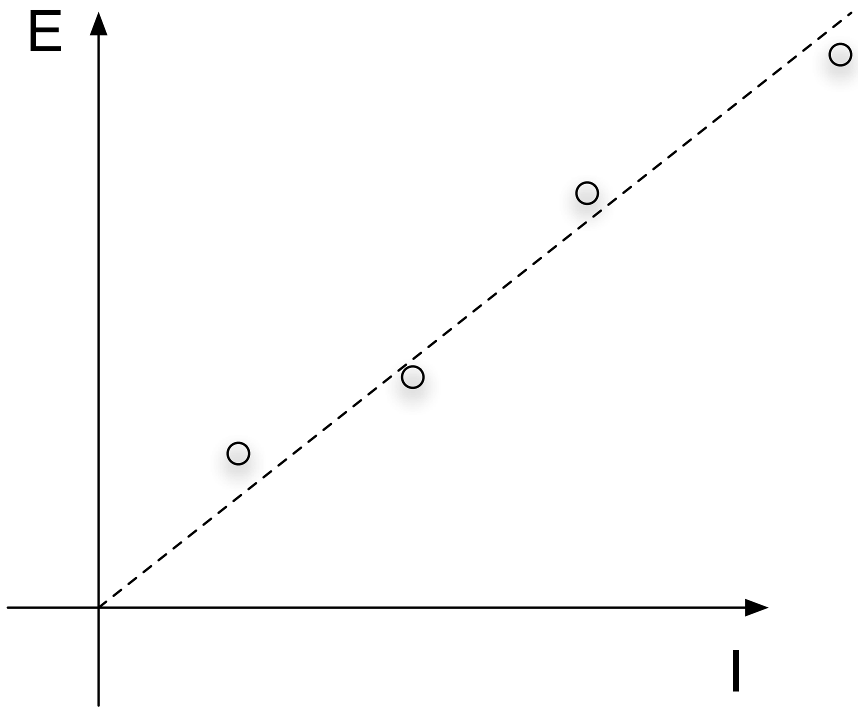
- 法則／モデル式にデータを当てはめる。
- データから未知のパラメータを推測する。
- データは測定誤差を含む

例：オームの法則

$$E = IR$$

複数の測定値の組 $(i_1, e_1), (i_2, e_2), \ldots, (i_m, e_m)$ から R を推定したい。

直線の当てはめ



- 推定直線の傾き$= R$の推定値
- "合理的" に直線を定めるにはどうする？

問題設定

$$b = x_1 + x_2 a + e$$

- 物理量a, bの間には比例関係がある。
- x_2が傾きを表し、x_1が切片を表すパラメータ(混乱しないように注意)。
- eは誤差を表す
- $(a_1, b_1), (a_2, b_2), \ldots, (a_m, b_m)$という$m$個の測定データがある。

問題を書き直す

$$
\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} - \begin{pmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \\ 1 & a_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}
$$

$\Rightarrow$

$$
e = b - Ax
$$

$e, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times 2}, x \in \mathbb{R}^2$

最小二乗法

戦略：二乗誤差を最小化するようにx_1, x_2を定める

$$\text{minimize } ||e||_2^2$$

$$\Leftrightarrow \text{minimize } ||b - Ax||_2^2$$

最小二乗法の解法 (1)

無制約な二次関数の最小化問題となる。

$$f_0(x) = ||b - Ax||_2^2 = x^T A^T Ax - 2b^T Ax + b^T b$$

f_0 **は凸関数であり、制約条件はついていないので、最適解の必要十分条件は** $\nabla f_0(x) = 0$

$$\nabla f_0(x) = 2A^T Ax - 2(b^T A)^T = 0$$

となる。これを解くと

最小二乗法の解法(2)

$$x = (A^T A)^{-1} A^T b$$

が得られる。これが最小二乗法による推定パラメータとなる。

- 最小二乗法を実行するためには、勾配法もニュートン法も内点法も必要はない。逆行列の計算さえできればOK。

補足事項

$$x = (A^T A)^{-1} A^T b$$

における $(A^T A)^{-1} A^T$ のことをムーア・ペンローズの疑似逆行列と呼ぶこともある。

なぜ二乗誤差 (ℓ_2 ノルムの二乗) に注目するのか。

- 雑音 e の各要素がガウス分布に従う場合には、最小二乗法が最尤推定となるため。
- 計算が容易にできるため。

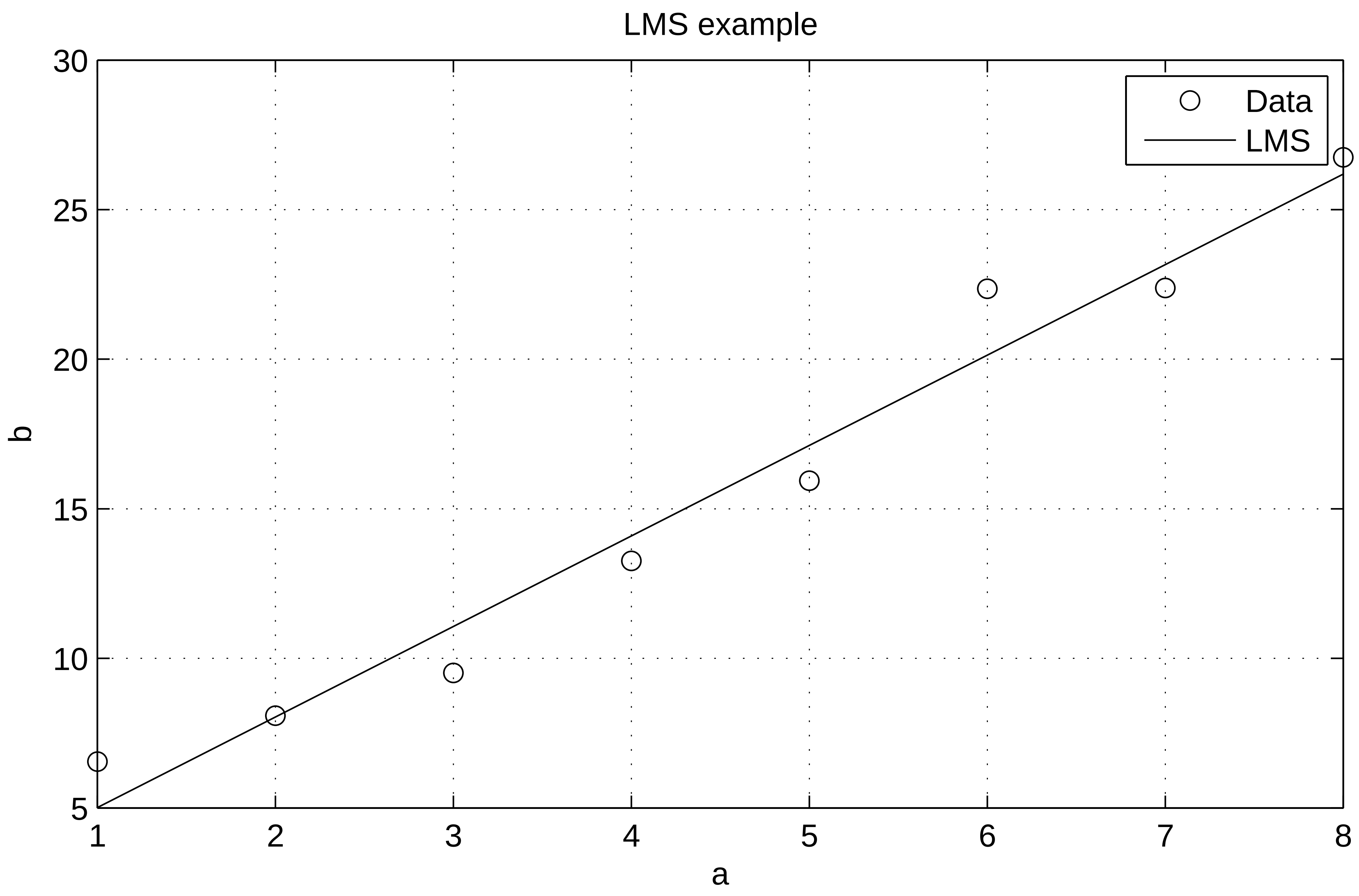
実行例

$$a = (1, 2, 3, 4, 5, 6, 7, 8)^T$$

$$b = (6.5442, 8.0859, 9.5084, 13.2577, 15.9384, 22.3505, 22.3844, 26.7481)^T$$

$$x = (A^T A)^{-1} A^T b = \begin{pmatrix} 1.9883 \\ 3.0253 \end{pmatrix}$$

生成モデルは $x_1 = 2.0, x_2 = 3.0$ **を利用した。**



LMS example
Data
LMS
30
25
20
15
10
5
1
2
3
4
5
6
7
8
a
b

パラメータに事前情報がある場合の最小二乗法

もし、$x \geq 0$だと"事前"に分かっている場合は、その情報も制約条件として、推定に織り込むことが望ましく思える。この場合、解くべき問題

$$\text{minimize } ||Ax - b||_2^2$$
$$\text{subject to } x \geq 0$$

これは、二次計画問題($\subset$ 凸計画問題)となり、無制約問題のような簡単な公式はない。

事前制約の例

- **非負性**: $x \geq 0$
- **下限と上限**: $l \leq x \leq u$
- **パラメータベクトルは確率**：$x \geq 0, \mathbf{1}^t x = 1$
- **ノルム球制約**：$||x - x_0|| \leq d$
- **多面体制約**：$Cx - d \leq 0$

上のような事前情報の制約のもとに誤差を最小化する問題は凸計画問題となる。

事前制約を持つ回帰問題 (凸計画形)

$A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$

$$\begin{aligned}
&\text{minimize } ||Ax - b|| \\
&\text{subject to} \\
&f_i(x) \leq 0, \quad i = 1, 2, \ldots, q \\
&a_i^T x = b, \quad i = 1, 2, \ldots, p
\end{aligned}$$

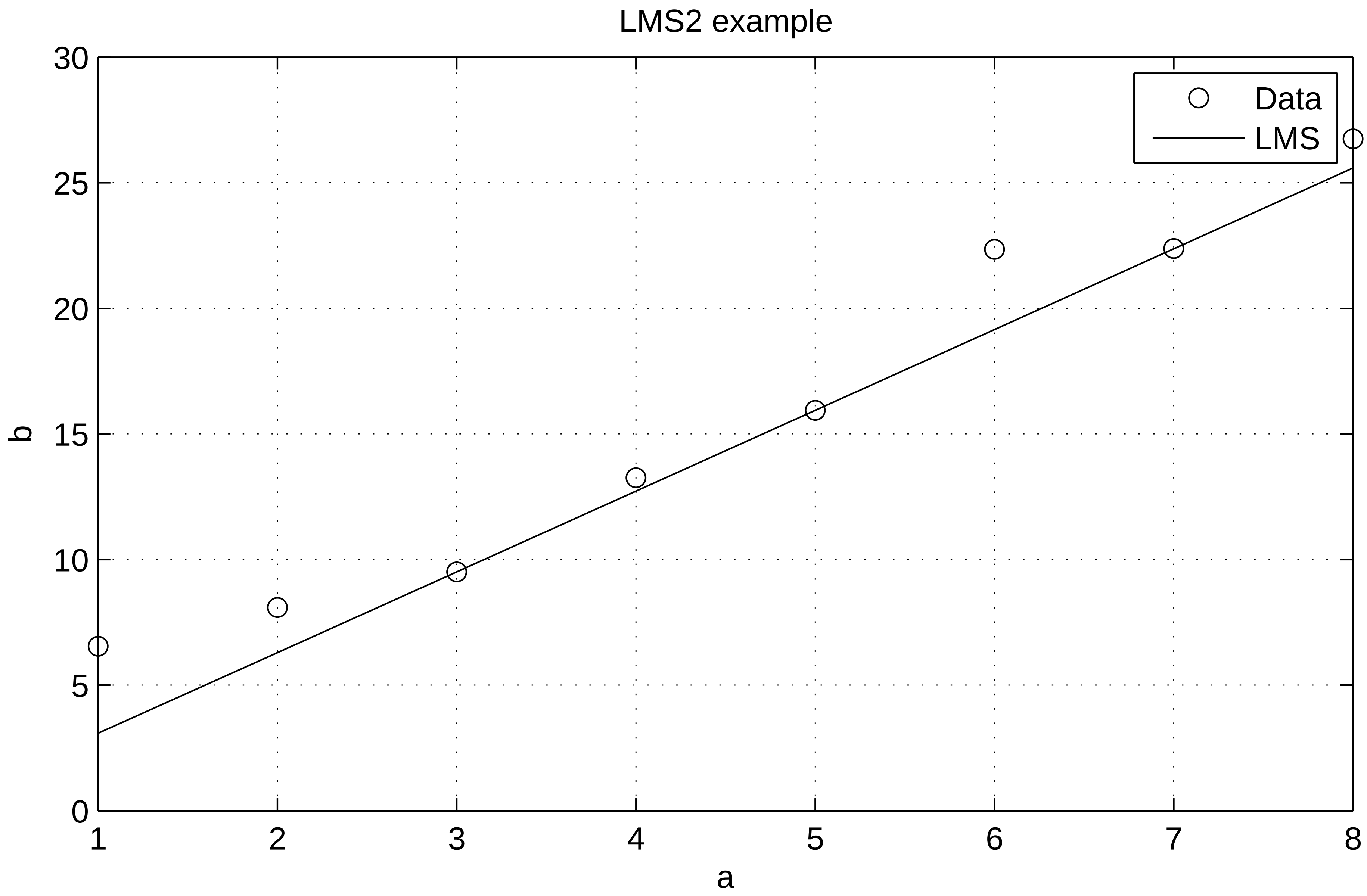
- f_i は凸関数。
- ノルム近似問題とも呼ばれる。

例題

誤差が必ず非負となるように制約する場合：

$$\text{minimize } ||Ax - b||_2^2$$
$$\text{subject to } Ax - b \geq 0$$

- **二次計画問題となる。**


LMS2 example
Data
LMS
30
25
20
15
10
5
0
b
1
2
3
4
5
6
7
8
a

関数近似問題

$x \in \mathbb{R}^n$ とし、未知の関数 $g(y) : \mathbb{R}^n \to \mathbb{R}$ を

$$h(y) = x_1 h_1(y) + x_2 h_2(y) + \cdots + x_k h_n(y)$$

で近似したい $(g(y) \simeq h(y))$。ただし、以下のデータのみが与えられている：

$$\begin{array}{cc} g(y_1) & y_1 \\ g(y_2) & y_2 \\ \vdots & \vdots \\ g(y_m) & y_m \end{array}$$

機械学習の観点から見ると

未知のシステム$g(y)$をその入力／出力関係の有限数データから近似する。

- **システム$g(y)$の内部構造／法則の推定**
- **システム$g(y)$の振る舞いの予測**

例：

$g(y) =$ 次の日の日経平均、$y =$ (日米為替レート、ダウ平均、消費者物価指数)

さまざまな基底関数

- **多項式近似**

$$h(y) = x_1 + x_2 y + x_3 y^2 + \cdots + x_n y^{n-1}$$

- **ガウス基底関数**

$$h(y) = \sum_{i=1}^{n} x_i \exp\left(-\frac{(y-\mu_i)^2}{2\sigma^2}\right)$$

- **三角関数**

$$h(y) = \sum_{i=1}^{n} x_i \sin(i\omega y)$$

問題を書き直す

$$\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix} = \begin{pmatrix} g(y_1) \\ g(y_2) \\ \vdots \\ g(y_m) \end{pmatrix} - \begin{pmatrix} h_1(y_1) & \cdots & h_n(y_1) \\ h_1(y_2) & \cdots & h_n(y_2) \\ \vdots & & \\ h_1(y_m) & \cdots & h_n(y_m) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$\Rightarrow$

$$e = b - Ax$$

$e, b \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n$

最小二乗法

制約条件がない場合には、二乗誤差の最小化問題

$$\text{minimize } ||Ax - b||_2^2$$

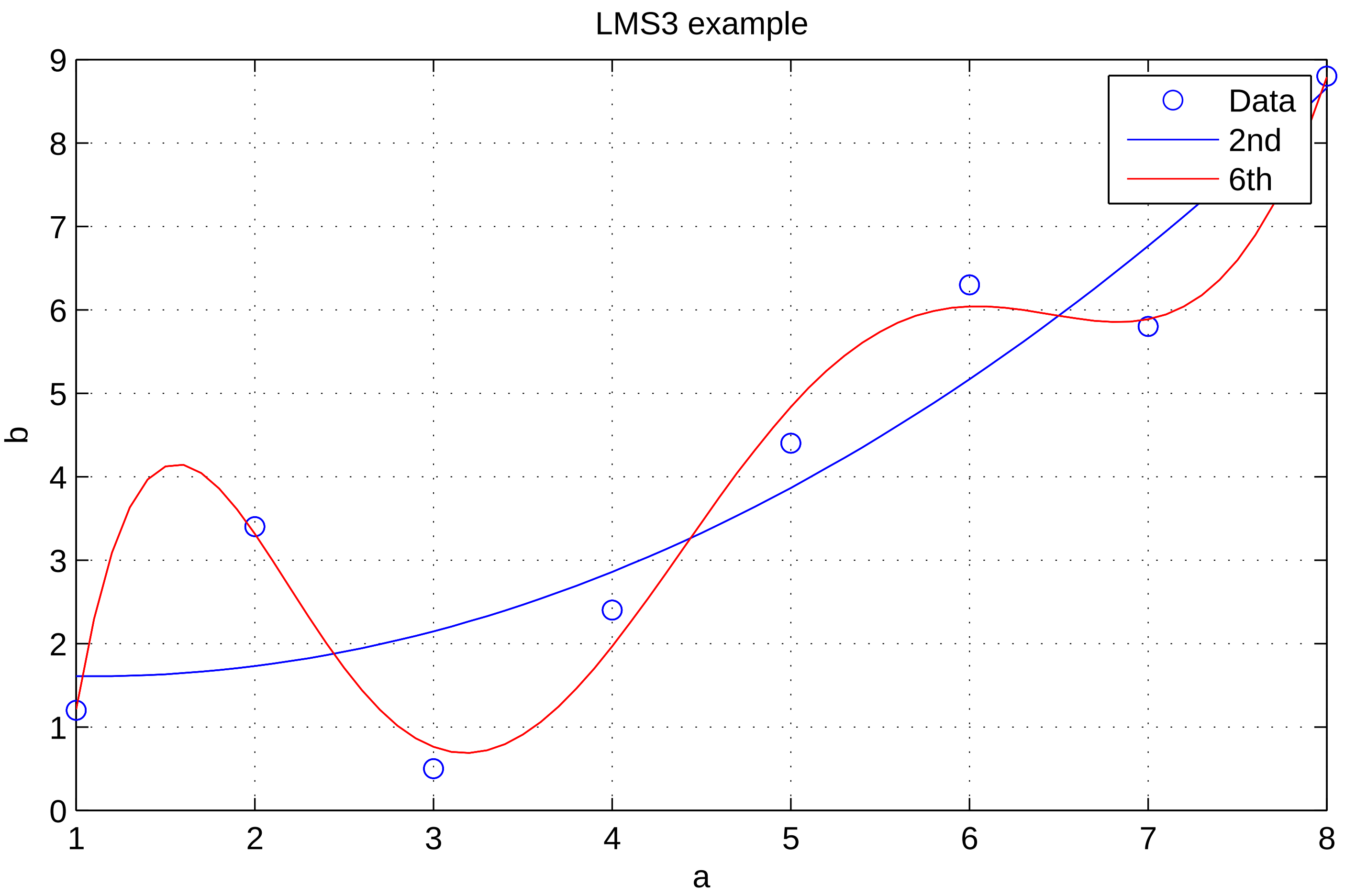
の解はやはり

$$x = (A^T A)^{-1} A^T b$$

となる。

多項式基底の例

二次と６次の近似



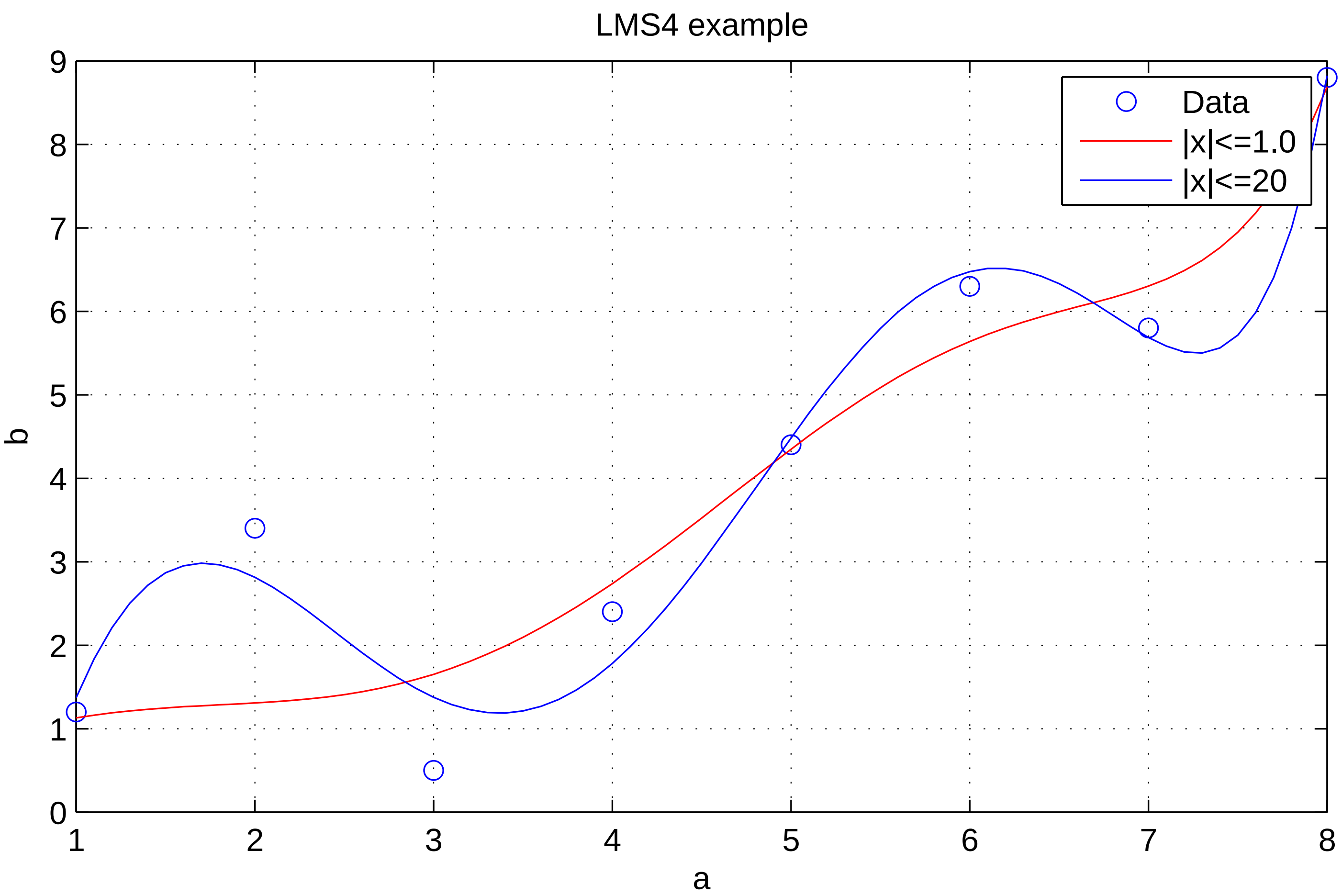
事前制約の例(ふたたび)

- 非負性: $x \geq 0$
- 下限と上限: $l \leq x \leq u$
- パラメータベクトルは確率：$x \geq 0, \mathbf{1}^t x = 1$
- ノルム球制約：$||x - x_0|| \leq d$
- 多面体制約：$Cx - d \leq 0$

上のような事前情報の制約のもとに誤差を最小化する問題は凸計画問題となる。

パラメータのノルムに関する制約

6次多項式：$||x||_2 \leq 1$ v.s. $||x||_2 \leq 20$



正則化(regularization)

- 観察: x のノルムが小さいほうがオーバーフィッティング(過学習)の影響が小さくなる傾向にある
- モデルの複雑さを押さえる働きをする
- 同じことを無制約問題として実現したい $\Rightarrow$ 目的関数にパラメータのノルム項を加える(正則化)

正則化最小二乗法

$$\text{minimize}\ ||Ax - b||_2^2 + \lambda||x||_p^p$$

$p = 2$ の場合

$$\text{minimize}\ ||Ax - b||_2^2 + \lambda||x||_2^2$$

正則化最小二乗法の解

(Tikhonov regularization)

$$\text{minimize } ||Ax - b||_2^2 + \lambda||x||_2^2$$

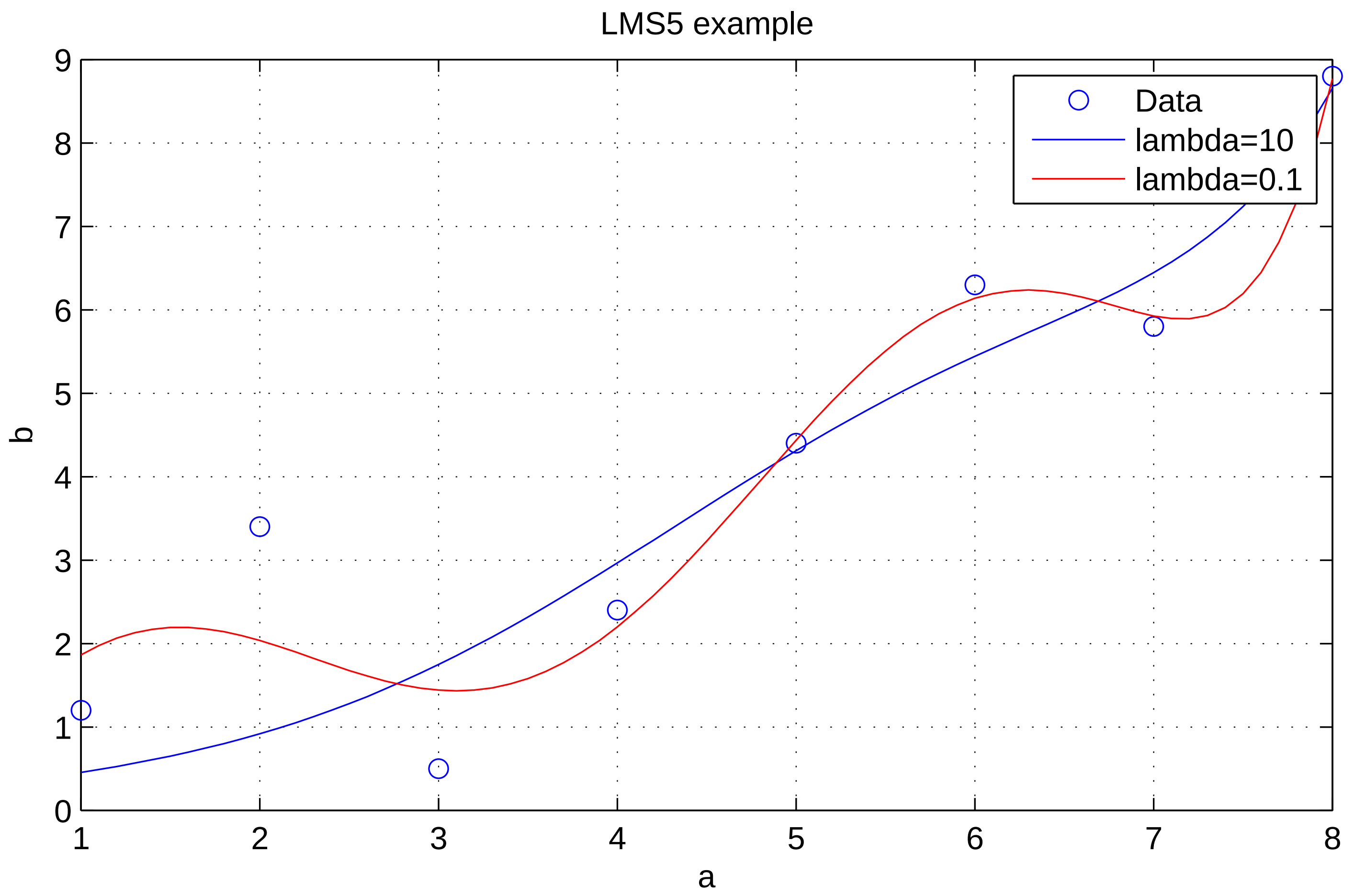
の解は

$$x = (A^T A + \lambda I)^{-1} A^T b$$

となる。

正則化の例

正則化係数：$\lambda = 0.1$ v.s. $\lambda = 10$

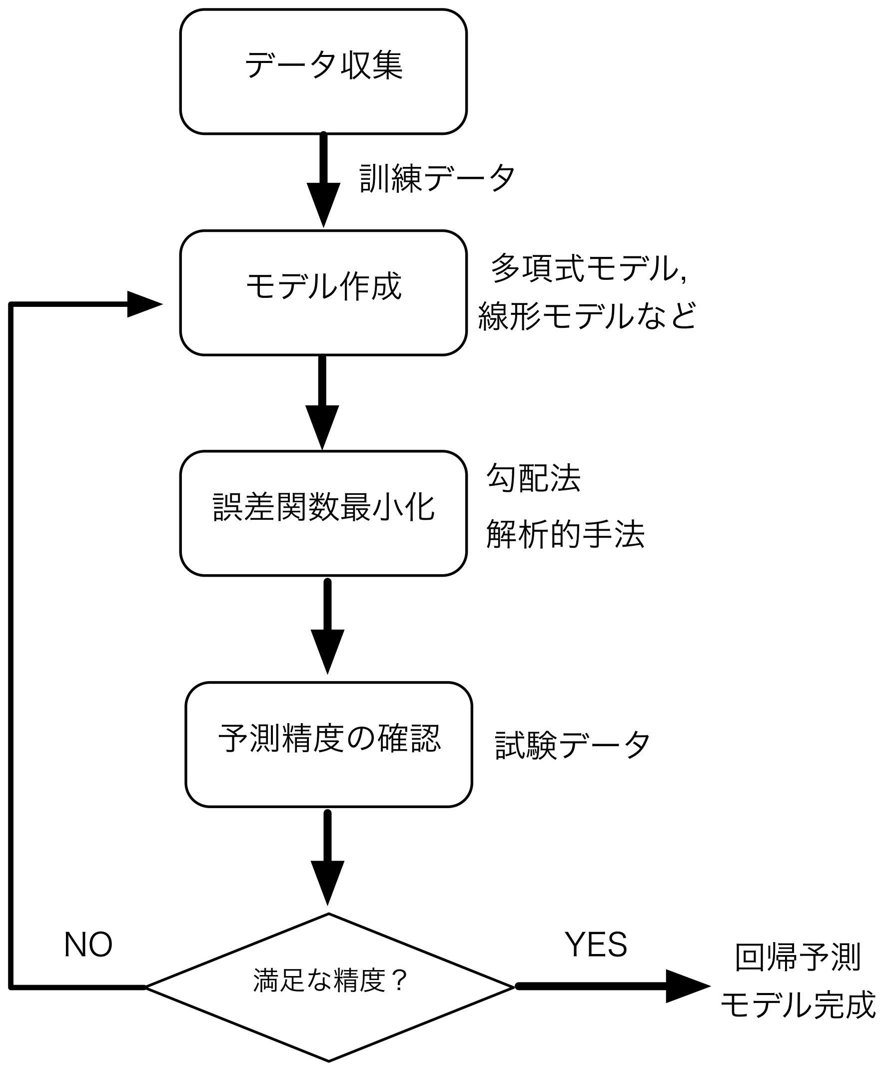


ℓ_1-正則化 (lasso)

$$\text{minimize } ||Ax - b||_2^2 + \lambda ||x||_1$$

- 疎な解が得られる
- 0 となる係数に対応する基底関数は近似に影響しない

回帰のためのレシピ


データ収集
訓練データ
モデル作成
多項式モデル,
線形モデルなど
誤差関数最小化
勾配法
解析的手法
予測精度の確認
試験データ
NO
満足な精度？
YES
回帰予測
モデル完成