データ分析入門 第8回
回帰分析

京都大学情報学研究科／高度情報教育基盤コア

関戸 啓人（せきど ひろと）

回帰曲線と回帰分析 1

★ 2つの確率変数 X と Y を考える

★ $X = x$ という条件下での Y の平均 $E[Y|X = x]$ を x の関数と思ったとき，それを回帰曲線という

★ 回帰分析とは，大雑把に言えば，回帰曲線を推定することにより，2つの確率変数 X と Y の関係を調べること，である

- ★ $E[Y|X = x]$ を推定するときは，X は説明変数で，Y は目的変数と呼ばれる
- ★ つまり，Y がどのような値を取るかは X によって定まる，と考えている

★ 説明変数は複数あっても良い

★ 説明変数が $X_1, X_2, \ldots, X_n$ で，$E[Y|X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]$ を考えても良い

★ 説明変数が1個の場合を単回帰分析，複数の場合を重回帰分析という

回帰曲線と回帰分析 2

★ 例えば，小学生を対象に，Xを朝食を食べる割合，Yでテストの点数とすれば，回帰分析で，これらの関係がわかるであろう

★ 多くの場合は，

$$E[Y|X = x] = ax + b \quad (a, b \in \mathbb{R})$$

という関係を仮定する，もしくは，

$$E[Y|X = x] = ax + b + \varepsilon \quad (a, b \in \mathbb{R})$$

として，誤差項（Xだけでは説明できない部分）εをできるだけ小さくするようにa, bを決めることが多い

★ 重回帰分析の場合は，

$$E[Y|X_k = x_k] = \sum_{k=1}^{n} a_k x_k + b + \varepsilon \quad (a_k, b \in \mathbb{R})$$

とする場合が多い．

★ 勿論，もっと複雑な式を考えることもある

回帰曲線と回帰分析 3

★ ただし，推定した結果，αが明らかに正であるからと言って，朝食を食べることが直接テストの点数を上げるとは限らない．例えば，食生活をきちんと躾けていれば，朝食を食べる割合が高くなり，テストの点数も良い傾向にあり，朝食を食べたからといってテストの点数が上がるわけではないかもしれない．

★ αが0か，正か，負かのみが重要な場合もあるが，そのようなときは，検定などを用いれば良い．他にも，α, bを推定した後，ε（残差）がどのようになっているか調べ，$E[Y|X=x]$の式の形を修正するなどを考えることも必要かもしれない．今回は，どうやって，関数$E[Y|X=x]$を推定するかについて述べる．多くの場合は，最小二乗法を用いる．

最小二乗法の概要

★ 未知な関数を得られたデータから推定したい

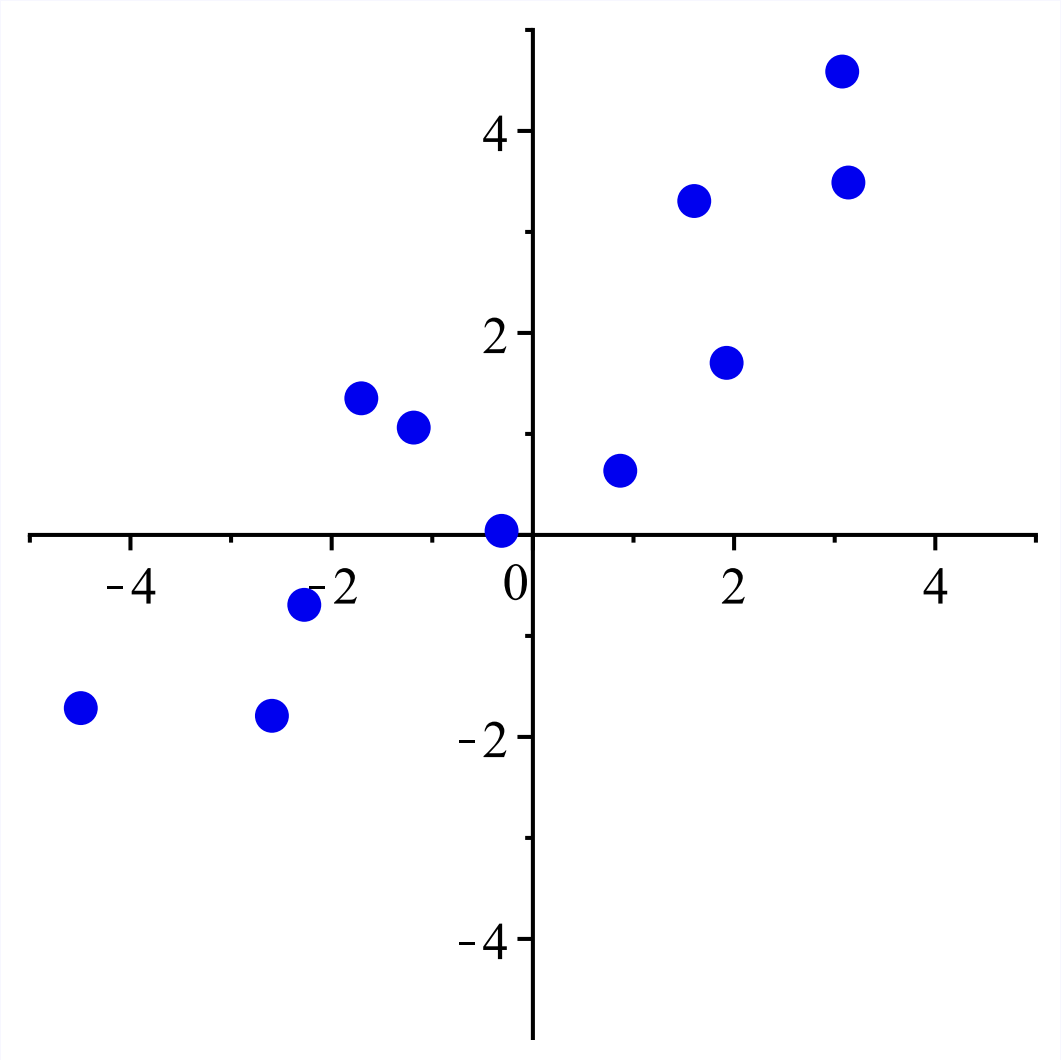
★ 未知関数 $f(x)$ の形はわかっていて，未知パラメータを含む形で書かれる

★ データ (x_j, y_j) は $f(x_j)$ での値が y_j であることを「示唆」する

★ データは厳密に「正しい」訳ではない．つまり厳密に $f(x_j) = y_j$ とは限らない（測定誤差などが含まれている）

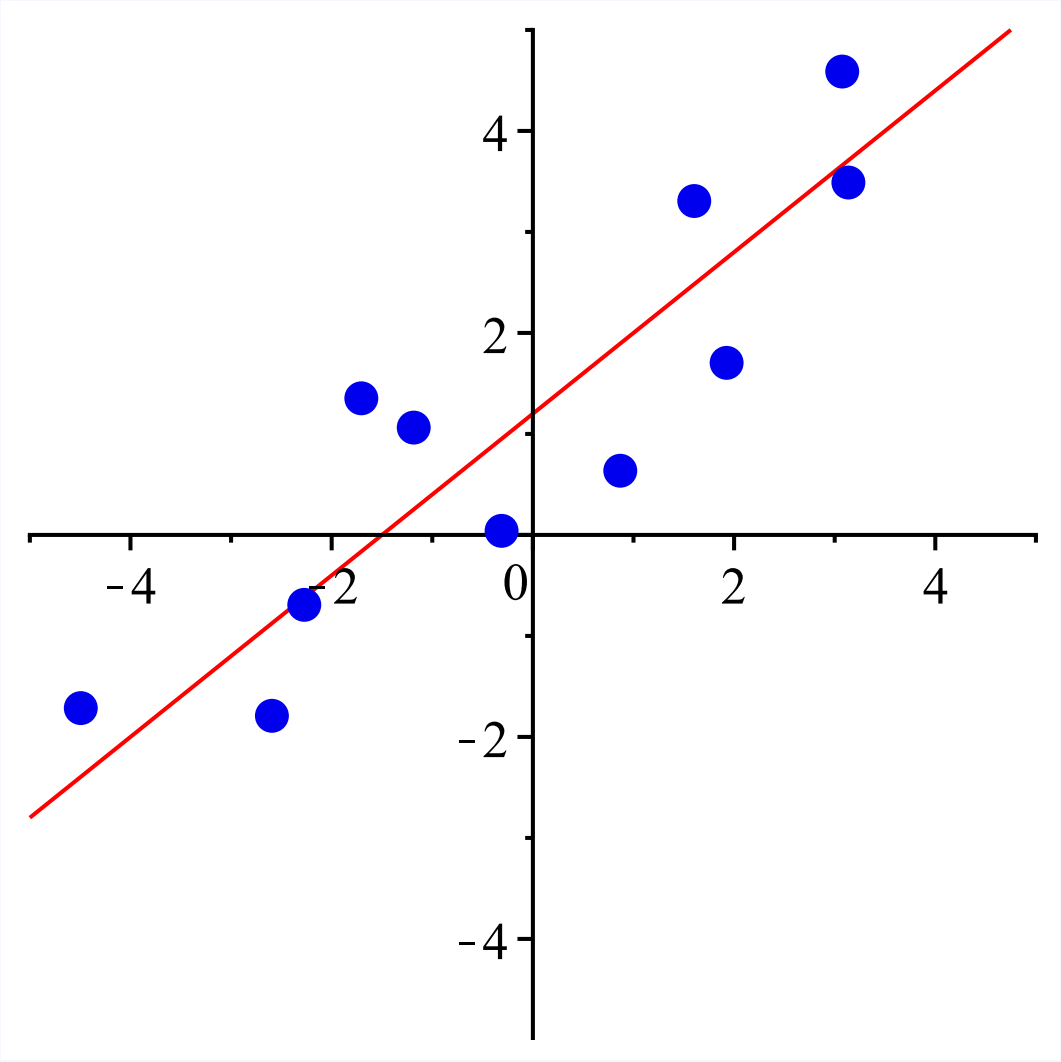
★ 2変数以上の場合は x はベクトルだと思えば良い

最小二乗法の例 (その1)



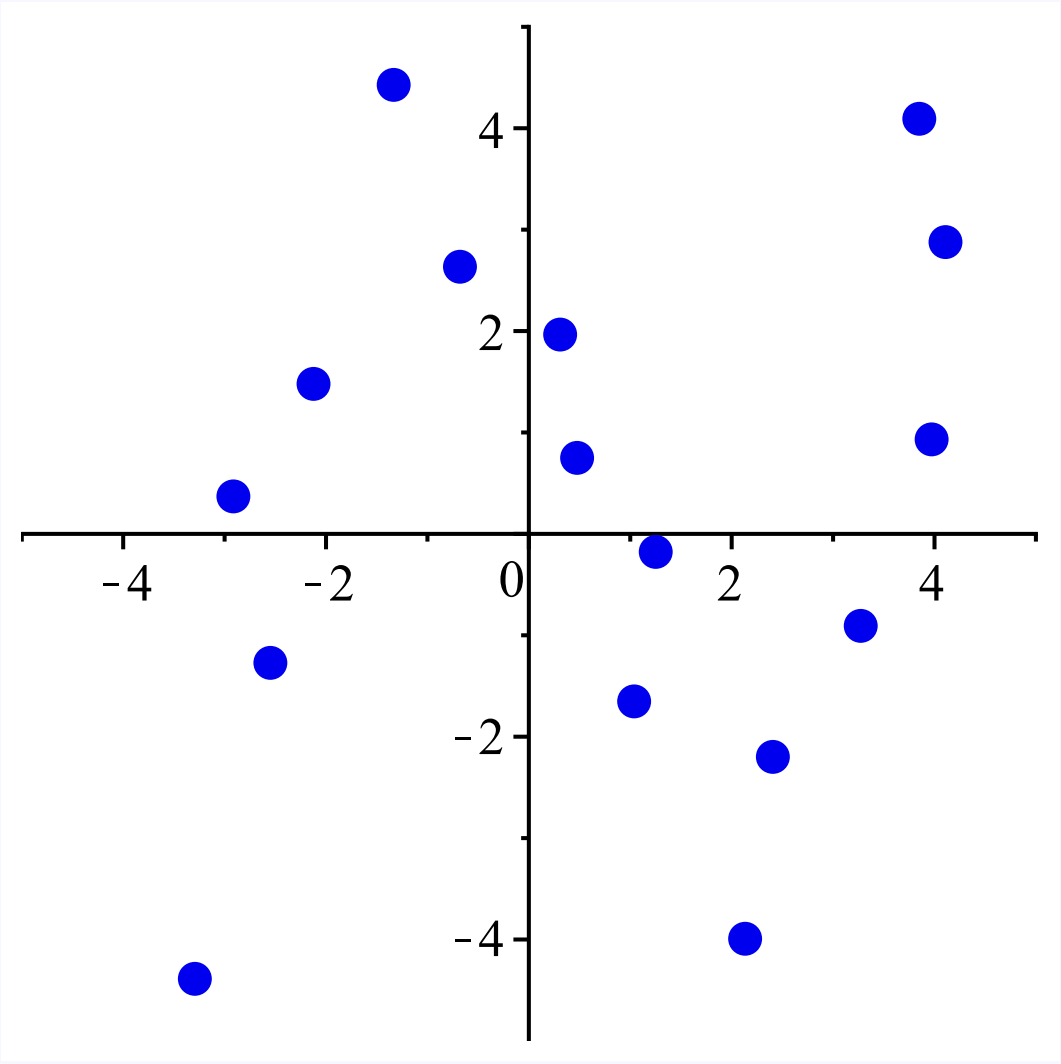
$$f(x) = \theta_1 x + \theta_0$$

最小二乗法の例 (その1)



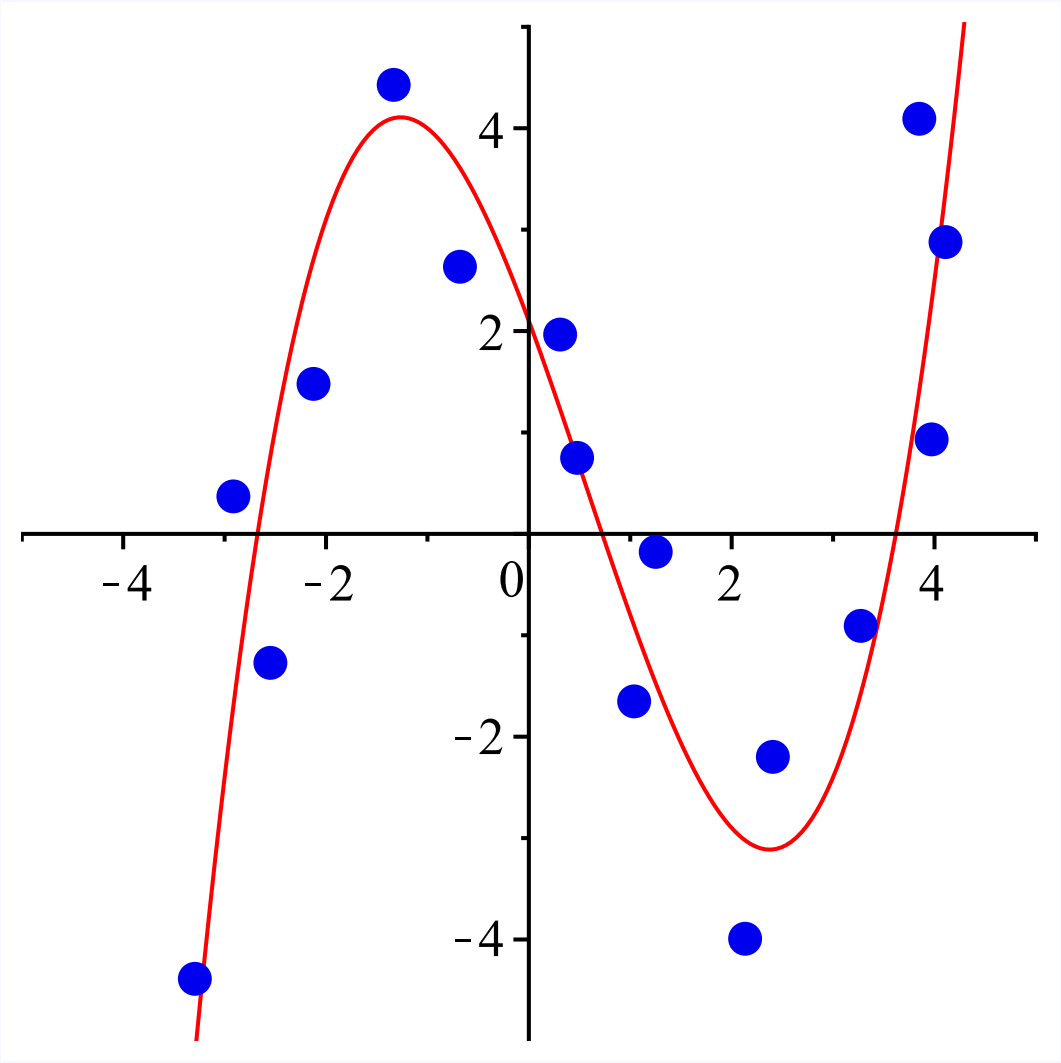
$$f(x) = 0.8x + 1.2$$

最小二乗法の例 (その2)



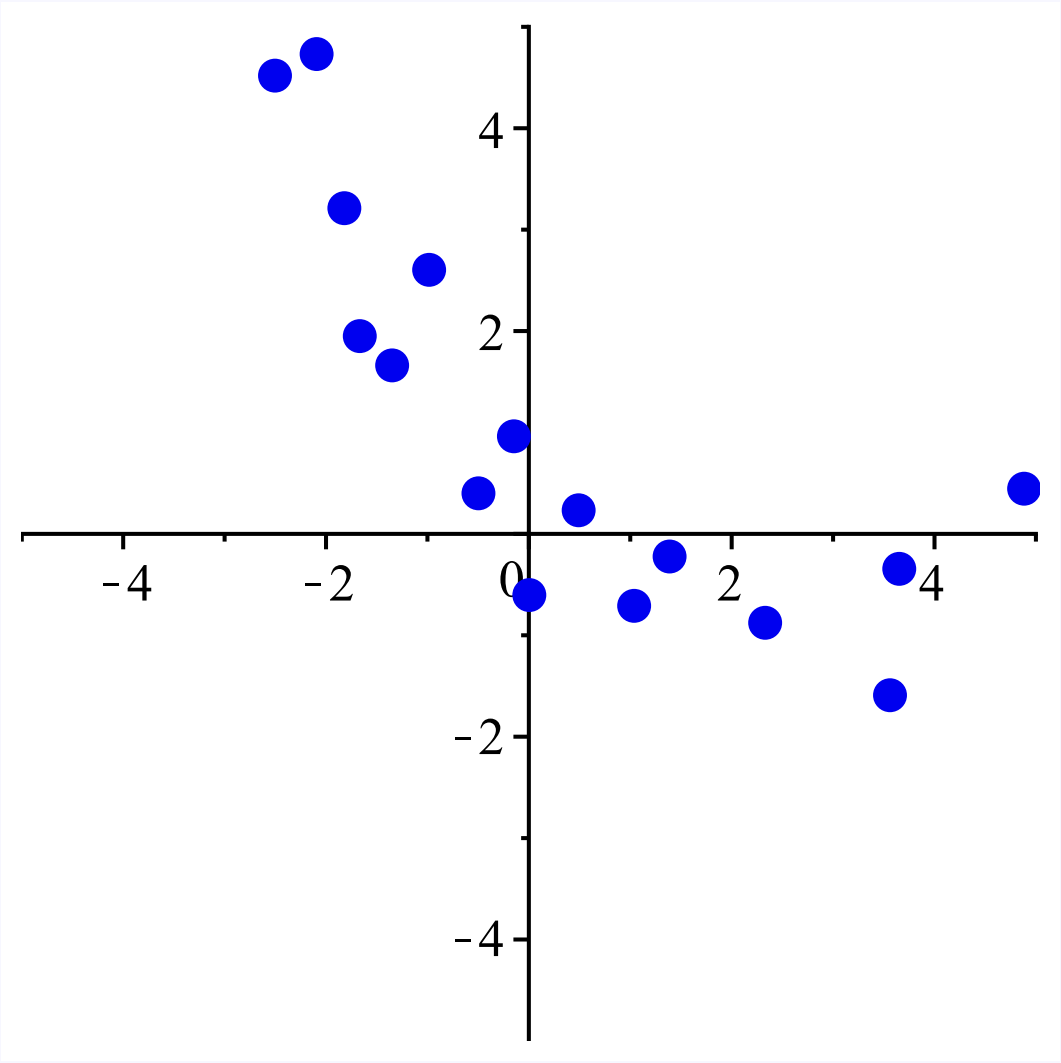
$$f(x) = \theta_3 x^3 + \theta_2 x^2 + \theta_1 x + \theta_0$$

最小二乗法の例 (その2)



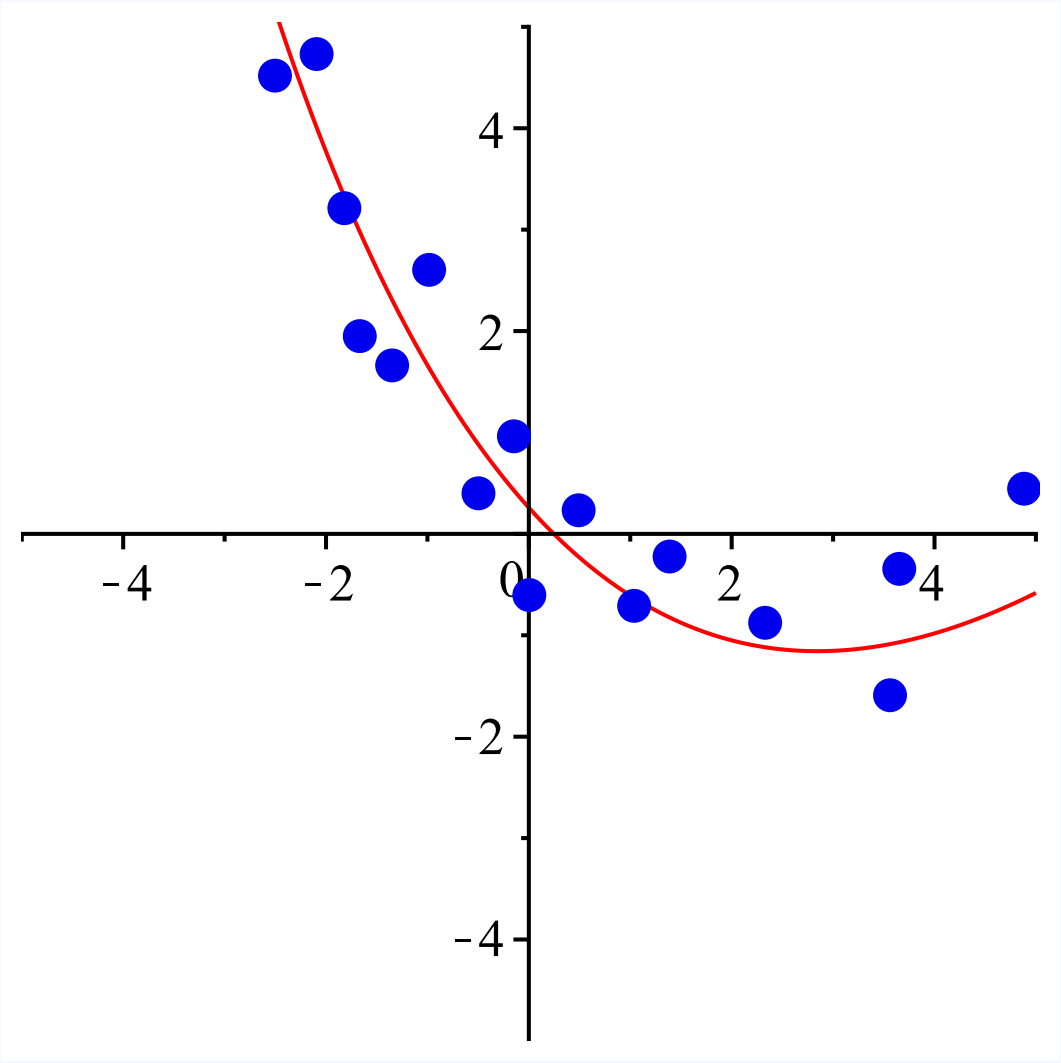
$$f(x) = 0.3x^3 - 0.5x^2 - 2.7x + 2.1$$

最小二乗法の例 (その3)



$$f(x) = \frac{\theta_2 x^2 + \theta_1 x + \theta_0}{x + \theta_3}$$

最小二乗法の例 (その3)



$$f(x) = \frac{2.1x^2 - 13.1x + 3.1}{x + 12.0}$$

最小二乗法の例

★ その1: 直線で近似する場合
  ★ $f(x) = \theta_0 + \theta_1 x$
  ★ （単純な）単回帰分析
★ その2: 未知関数がパラメータについて線形（線形最小二乗法）
  ★ $f(x) = \theta_0 f_0(x) + \theta_1 f_1(x) + \cdots + \theta_{m-1} f_{m-1}(x)$
  ★ （単純な）重回帰分析，以下では主にこれを説明する
★ その3: 未知関数がパラメータについて非線形（非線形最小二乗法）
  ★ $f(x) = f(x; \theta_0, \theta_1, \ldots, \theta_{m-1})$
  ★ 複雑な式の形を指定した場合，解く場合は最適化の理論を用いる

回帰モデルの例 (1) — 単回帰モデル

★ 体重を意味する確率変数を W

★ 身長を意味する確率変数を H

★ モデル：$W = \theta_1 H + \theta_0 + \varepsilon$

★ データは，例えば

	体重(kg)	切片	身長(cm)
A氏	56.8	1	163.3
B氏	52.1	1	160.2
C氏	52.6	1	158.0
D氏	23.4	1	129.0
E氏	32.1	1	139.7
F氏	40.6	1	141.4

回帰モデルの例 (2-1) — 重回帰モデル

★ 体重を意味する確率変数を W

★ 身長を意味する確率変数を H

★ モデル：$W = \theta_2 H^2 + \theta_1 H + \theta_0 + \varepsilon$

★ データは，例えば

	体重(kg)	切片	身長(cm)	身長^2(cm^2)
A氏	56.8	1	163.3	26666.89
B氏	52.1	1	160.2	25664.04
C氏	52.6	1	158.0	24964.00
D氏	23.4	1	129.0	16641.00
E氏	32.1	1	139.7	19516.09
F氏	40.6	1	141.4	19993.96

回帰モデルの例 (2-2) — 重回帰モデル

★ 体重を W，身長 H，体脂肪率を F，性別を S

★ 性別は女性を 1，男性を 0 で表す

★ モデル：$W = \theta_3 S + \theta_2 F + \theta_1 H + \theta_0 + \varepsilon$

★ データは，例えば

	体重(kg)	切片	身長(cm)	体脂肪率(%)	性別
A氏	56.8	1	163.3	14.3	0
B氏	52.1	1	160.2	15.3	0
C氏	52.6	1	158.0	21.2	1
D氏	23.4	1	129.0	13.3	1
E氏	32.1	1	139.7	16.8	0
F氏	40.6	1	141.4	19.6	1

線形最小二乗法の定義，および，性質 1

★ 観測と応答の関係

$$Y = \sum_{k=0}^{m-1} \theta_k f_k(x) + \varepsilon = f(x, \theta) + \varepsilon$$

は線形回帰モデルと呼ばれる

★ $f_k(x)$ は既知の関数

★ θ_k は未知のパラメータ，$\theta = (\theta_0, \theta_1, \ldots, \theta_{m-1})^T$

★ ε は確率変数で平均 0（$E[\varepsilon] = 0$）

★ 実際に n 個のデータ $(x_1, y_1), \ldots, (x_n, y_n)$ を用いて

$$y_j = f(x_j, \theta) + \varepsilon_j, \quad j = 1, 2, \ldots, n$$

とする

★ y_j, ε_j は確率変数

★ ε_j は j 回目の観測における誤差

線形最小二乗法の定義，および，性質 2

★ $y_j = f(x_j, \theta) + \varepsilon_j, \quad j = 1, 2, \ldots, n$

★ 今回は，誤差 ε_j に対して以下の仮定を置く

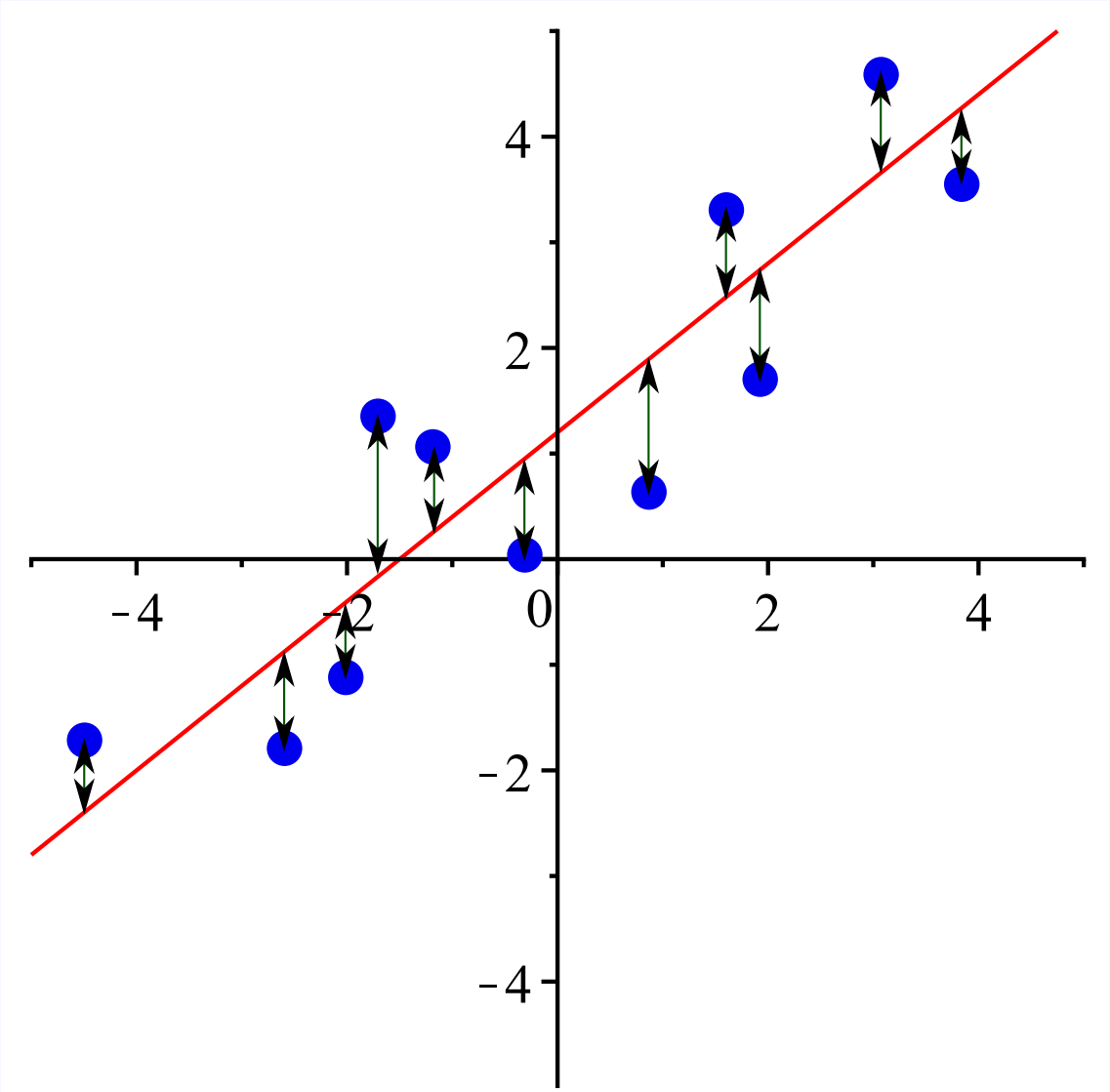
- ★ 平均は0．つまり，$E[\varepsilon_j] = 0$
- ★ 誤差の分散は等しく，正．つまり，$V[\varepsilon_j] = \sigma^2 > 0$
- ★ 誤差は互いに無相関．つまり，$E[\varepsilon_i \varepsilon_j] = 0, \quad i \neq j$

★ 残差二乗和

$$S(\beta) = \sum_{k=1}^{n} (y_k - f(x_k, \beta))^2$$

を最小化する未知パラメータベクトル β を最小二乗推定量 $\hat{\theta}$ と言う

絵で見る最小二乗法



緑の線の長さの二乗和を最小化するように，未知パラメータθを推定

線形最小二乗法の定義，および，性質 3

★ 最小二乗推定量 $\hat{\theta}$ は，最良線形不偏推定量である

- ★ $E[\hat{\theta}] = \theta$（不偏）
- ★ $\hat{\theta}$ は，y_j について線形の式で書ける（線形）
- ★ その中で，分散がある意味で最小（最良）
  - ★ 任意の不偏性と線形性を満たす β に対して，$Cov[\beta] - Cov[\hat{\theta}]$ が非負定値

★ 誤差 ε が正規分布に従うとき，最小二乗推定量 $\hat{\theta}$ は，最尤推定量である

- ★ つまり，$x_1, \ldots, x_n$ を固定して，測定結果として $y_1, \ldots, y_n$ が得られる確率を θ の関数として考えたとき，その確率の値が最大となるのが $\theta = \hat{\theta}$ のとき

最小二乗法推定量 (その1)

★方針

★残差二乗和

$$S(\beta) = \sum_{k=1}^{n} (y_k - f(x_k, \beta))^2$$

を最小化したいのだから，$\beta_0, \beta_1, \ldots, \beta_{m-1}$ で偏微分して 0 になる β を見つければ良い

最小二乗法推定量 (その1)

★ $f(x, \beta) = \beta_1 x + \beta_0$ の場合

★ $S(\beta) = \sum_{k=1}^{n} (y_k - \beta_1 x_k - \beta_0)^2$ であるから

★ $\frac{\partial}{\partial \beta_1} S(\beta) = 2 \sum_{k=1}^{n} (x_k^2 \beta_1 + x_k \beta_0 - x_k y_k) = 0$

★ $\frac{\partial}{\partial \beta_0} S(\beta) = 2 \sum_{k=1}^{n} (x_k \beta_1 + \beta_0 - y_k) = 0$

★ つまり，次の連立一次方程式を解けば良い

★ $\begin{pmatrix} \sum x_k^2 & \sum x_k \\ \sum x_k & n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_0 \end{pmatrix} = \begin{pmatrix} \sum x_k y_k \\ \sum y_k \end{pmatrix}$

$$\beta_1 = \frac{n \sum x_j y_j - \sum x_j \sum y_j}{n \sum x_j^2 - (\sum x_j)^2}$$

$$\beta_0 = \frac{\sum x_j^2 \sum y_j - \sum x_j y_j \sum x_j}{n \sum x_j^2 - (\sum x_j)^2}$$

最小二乗法推定量 (その2)

★ $f(x, \beta) = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_{m-1} f_{m-1}(x)$ の場合

★ $S(\beta) = \sum_{k=1}^{n} \left(y_k - \sum_{k=0}^{m-1} \beta_k f_k(x_k) \right)^2$ であるから

★ $\frac{\partial}{\partial \beta_i} S(\beta) = 2 \sum_{k=1}^{n} f_i(x_k) \left(\left(\sum_{j=0}^{m-1} f_j(x_k) \beta_j \right) - y_k \right) = 0$

★ $\sum_{j=0}^{m-1} \sum_{k=1}^{n} f_i(x_k) f_j(x_k) \beta_j = \sum_{k=1}^{n} f_i(x_k) y_k$

正規方程式

★ つまり，連立一次方程式 $B\beta = b$ を解けば良い

★ $B \in M_m(\mathbb{R}), \quad B_{ij} = \sum_{k=1}^{n} f_i(x_k) f_j(x_k)$

★ $b \in \mathbb{R}^m, \quad b_i = \sum_{k=1}^{n} f_i(x_k) y_k$

★ 行列 B がフルランクであれば，最小二乗推定量が一意に定まる

★ $B\beta = b$ は正規方程式と呼ばれる

★ 数値計算する際は，この方程式を直接解くよりも高精度な方法が存在する

正規方程式

★ 行列 $A \in M_{n,m}(\mathbb{R})$ を以下で定義（ヤコビアン，データ行列）

- ★ $A_{ij} = f_j(x_i) = \frac{\partial}{\partial \beta_j} f(x_i)$

★ $B = A^T A$

★ $b = A^T y$　（ただし $y = (y_1 \cdots y_n)^T$）

★ 正規方程式は以下のように書き直される

- ★ $A^T A \beta = A^T y$

★ 行列 A が列フルランクの場合

- ★ 最小二乗推定量は $\hat{\theta} = (A^T A)^{-1} A^T y$

QR分解を用いて解く

★ 行列Aは列フルランクでQR分解できたとする

- ★ $A = QR$
  - ★ $Q \in M_{n,m}(\mathbb{R})$ は列ベクトルが長さ1で互いに直交
  - ★ $R \in M_m(\mathbb{R})$ は正則な上三角行列

★ このとき，正規方程式は

- ★ $A^T A\beta = A^T y$
- ★ $(QR)^T QR\beta = (QR)^T y$
- ★ $R^T Q^T QR\beta = R^T Q^T y$
- ★ $R^T R\beta = R^T Q^T y$ $(Q^T Q = I)$
- ★ $R\beta = Q^T y$ (R^T は正則)

★ Rは上三角行列であるから，これは簡単に解ける

行列Aが列フルランクでない場合

★ 行列Aが列フルランクでない場合は，最小二乗推定量は一意に定まらない

★ 最小二乗推定量の中で，$\|\beta\|_2$を最小とするものを求めることが多い

★ $\|\beta\|_2 = \|\beta\| = \sqrt{\beta_0^2 + \beta_1^2 + \cdots \beta_{m-1}^2} = \sqrt{\beta^T \beta}$

★ 結論を言うと，A の Moore–Penrose の一般逆行列を A^+ と書くと $A^+ y = R^+ Q^T y$ が答え

★ ある程度ロバストに計算できる方法は特異値分解

★ 高速に計算するなら完全ピポット選択付きQR分解をして直交変換

一般逆行列

★ 正則でなくても，長方行列でも良い行列 $A \in M_{mn}(\mathbb{R})$ に対して，$AXA = A$ を満たす行列 $X \in M_{nm}(\mathbb{R})$ を一般逆行列といい A^- で表す

- ★ A^- は必ず存在し，一般的には A^- は一意ではなく複数存在する
- ★ 連立一次方程式 $Ax = b$ の解の一つは，存在するならば $x = A^-b$ と書ける
- ★ 連立一次方程式 $Ax = b$ の解は，存在するならば，任意のベクトル y を用いて $x = A^-b + (I - A^-A)y$ と書ける
- ★ 連立一次方程式 $Ax = b$ は $(I - AA^-)b = 0$ ならば解が存在する

Moore–Penroseの一般逆行列

★ 正則でなくても，長方行列でも良い行列 $A \in M_{mn}(\mathbb{R})$ に対して，$AXA = A, XAX = X, (AX)^T = AX, (XA)^T = XA$ を満たす行列 $X \in M_{nm}(\mathbb{R})$ をMoore–Penroseの一般逆行列といい A^+ で表す

- ★ A^+ は必ず存在し，一意である
- ★ 連立一次方程式 $Ax = b$ の解が存在するならば，その中で $\|x\|_2$ が最小となるものは $x = A^+b$ となる
- ★ 連立一次方程式 $Ax = b$ の解が存在しなければ，$\|Ax - b\|_2$ が最小とするのは $x = A^+b$ となる