2019年6月27日更新

5. 最小二乗法と線形近似

Least-square method & line fitting

東北大学 大学院工学研究科
嶋田 慶太 shimada@m.tohoku.ac.jp

TOHOKU UNIVERSITY

本日の内容

- この授業での列ベクトルと行列の表現
- モデル関数と近似
- 最小二乗法による線形近似
- 具体的な数値を入れた例
- Octave用の例題
- 課題

この授業での列ベクトルと行列の表現

この授業での行列・列ベクトルの表現

列 ベクトル

▶ 波括弧{}で表現

*世界的にこれがスタンダードというわけではないので
もし別の機会で使おうと思ったら，ルールを書いてください．

誤解がない場合は

$$\begin{Bmatrix} x_1 \\ \vdots \\ x_n \end{Bmatrix} \xrightarrow{\text{縮記}} \{x_i\}_{i=[1,\,n]} \xrightarrow{\text{縮記}} \{x_i\}$$

例えば

$\{1\}$　$\{0\}$

は全要素が
1，0の列ベクトル

* $\boldsymbol{X}$のように太字もよく使う．

* 行ベクトルは$\{x_i\}^T$のように列ベクトルの転置で表す

行列

▶ 大文字もしくは角括弧[]で表現

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

例えば

$$[\{x_i\}\{y_i\}]$$

は i 行の成分がx_i, y_iの $n \times 2$ 行列

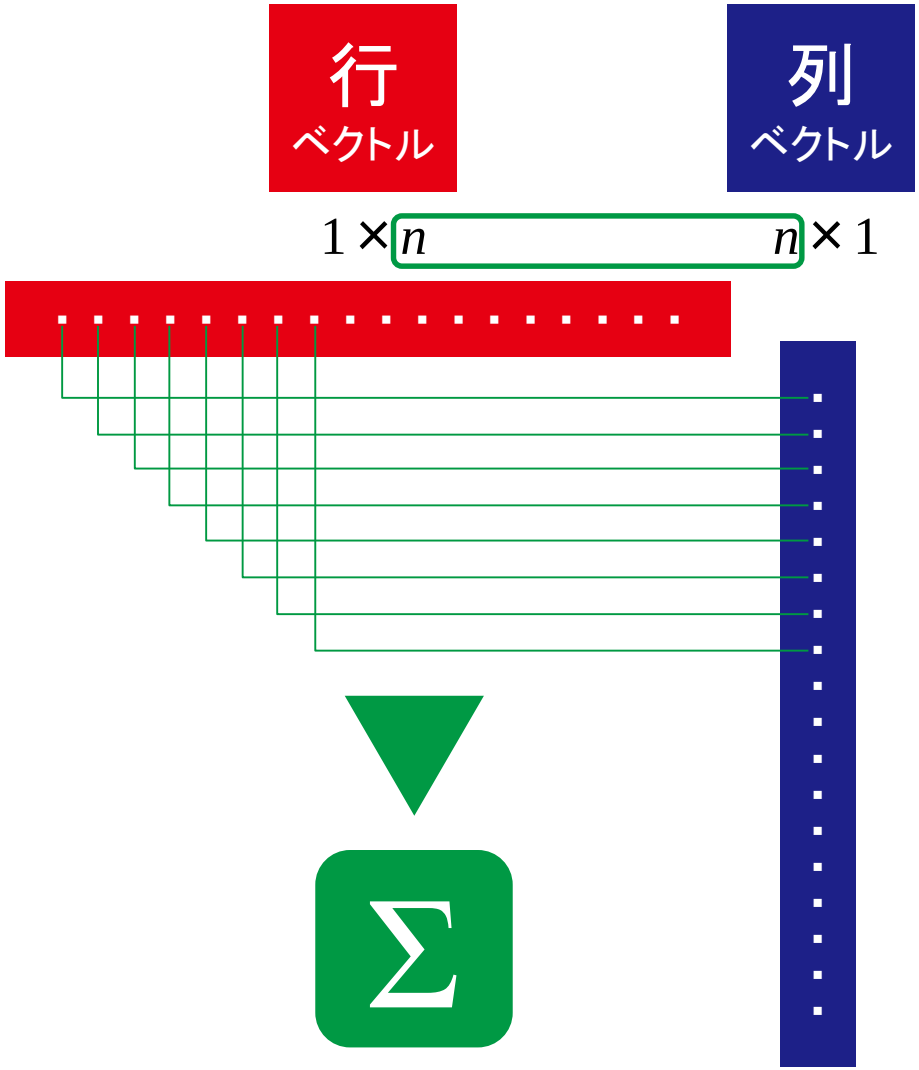
行列・列ベクトルによる総和の書き方

$$\sum_{i=1}^{n} 1 = \{1\}^T\{1\} = n$$

$$\sum_{i=1}^{n} x_i = \{1\}^T\{x_i\} = \{x_i\}^T\{1\}$$

$$\sum_{i=1}^{n} x_i^2 = \{x_i\}^T\{x_i\}$$

$$\sum_{i=1}^{n} x_i y_i = \{y_i\}^T\{x_i\} = \{x_i\}^T\{y_i\}$$



内積を使った書き方.

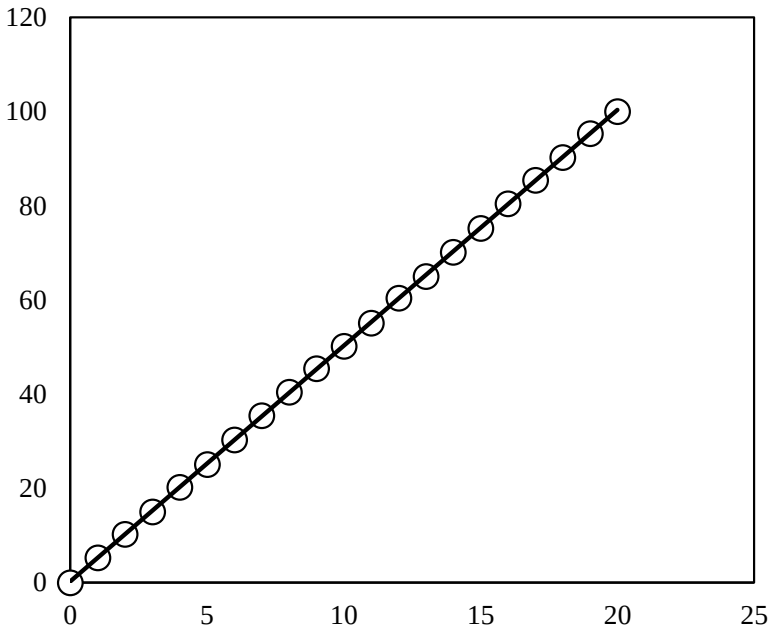
計算の検証


```
N = 50;
x = rand(N,1);
y = rand(N,1);
rows(x)
sum(ones(size(x)))
ones(size(x))'*ones(size(x))
sum(x)
ones(size(x))'*x
sum(x.^2)
x'*x
sum(x.*y)
y'*x
```

モデル関数と近似

入力・出力の情報を整理する

入力	出力
0	0
1	5.24
2	10.15
3	15.13
4	20.5
5	25.43
6	30.31
7	35.01
8	40.23
9	45.22
10	50.22
11	55.27
12	60.37
13	65.3
14	70
15	75.07
16	80.18
17	85.38
18	90.42
19	95.42
20	100.41



$$y = 5x$$

といえると情報が整理できる

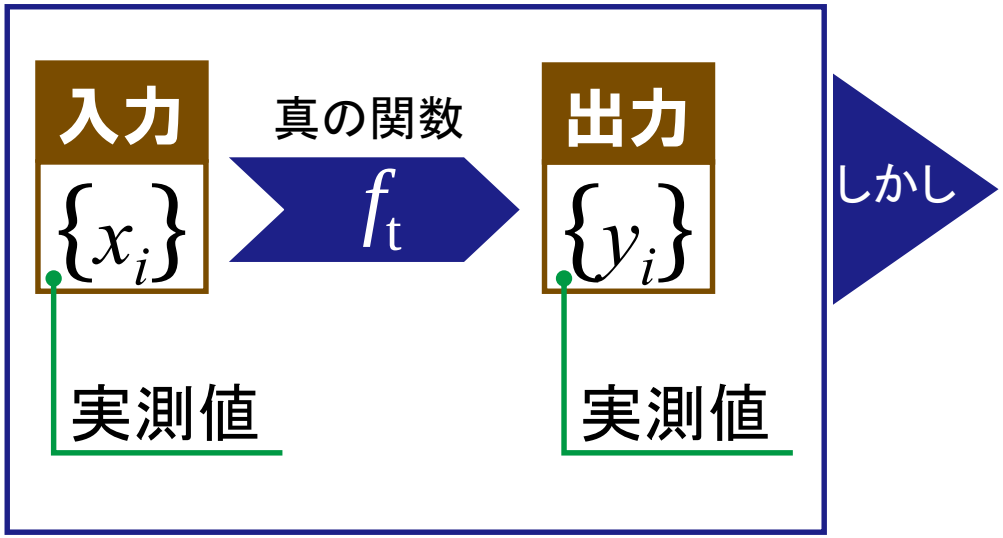
この時にデータ（入力・出力）
に基づいて数式を得る方法を考える

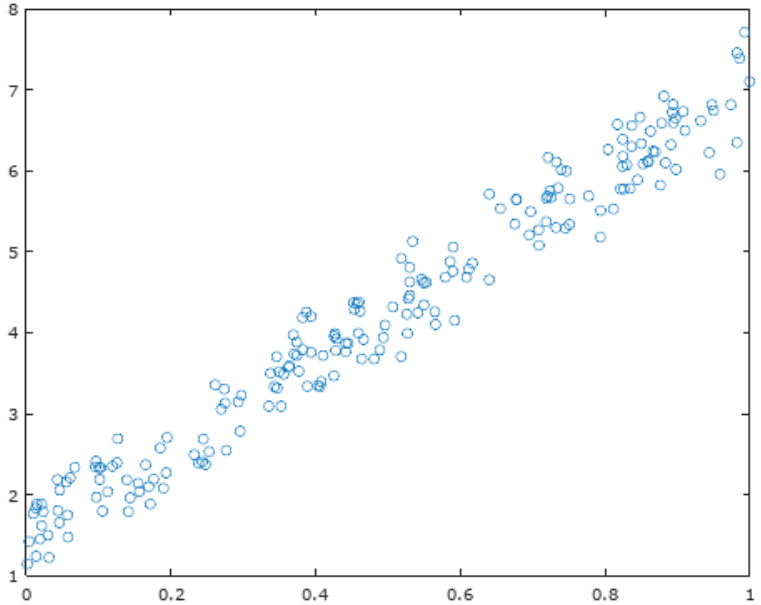
実験データ：入力と出力の関係

一般的な実験の目的

入力 x に対する出力 y を調べ，関係を見つける

例えば，
入力▸電流I
出力▸電圧V
オームの法則 $V = RI$ が見出されたように．



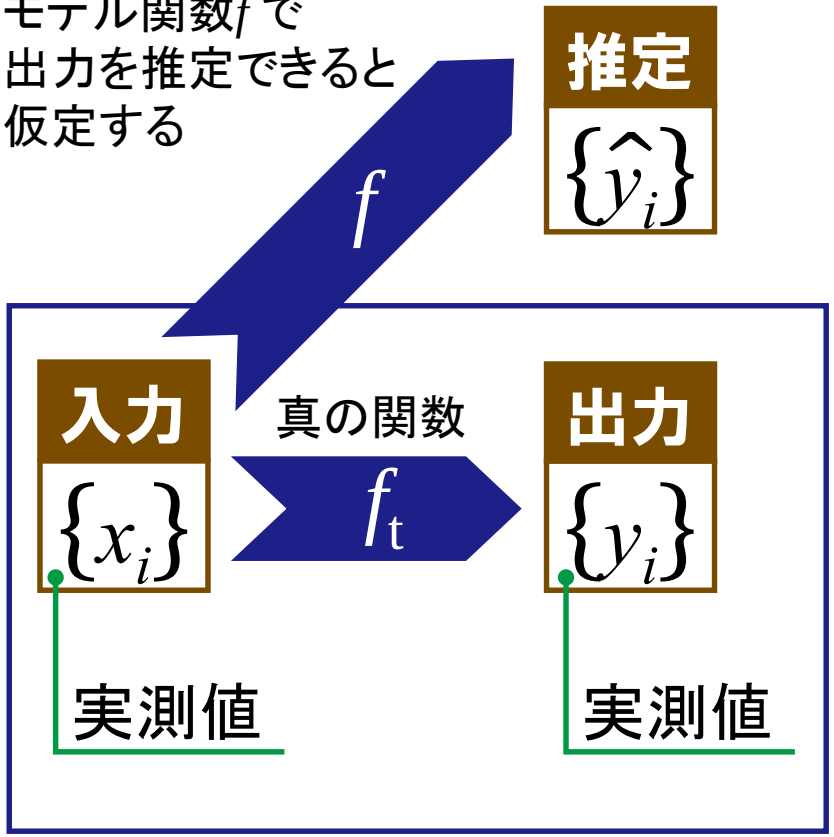


入力・出力の対応を
真の関数が記述するはず

現実には種々の要因によりばらつく

モデル関数：入力と出力を関数で近似

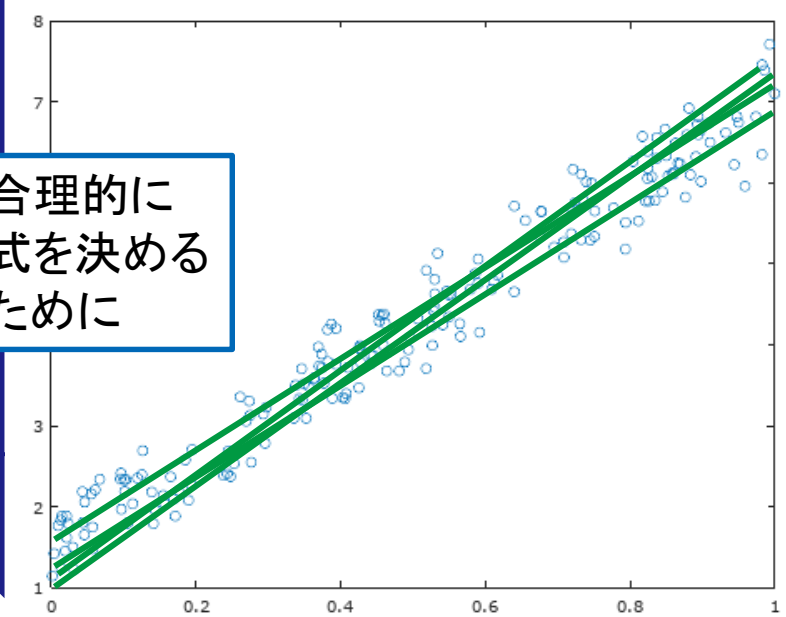
モデル関数fで
出力を推定できると
仮定する



例えばモデル関数を一次関数とすると，

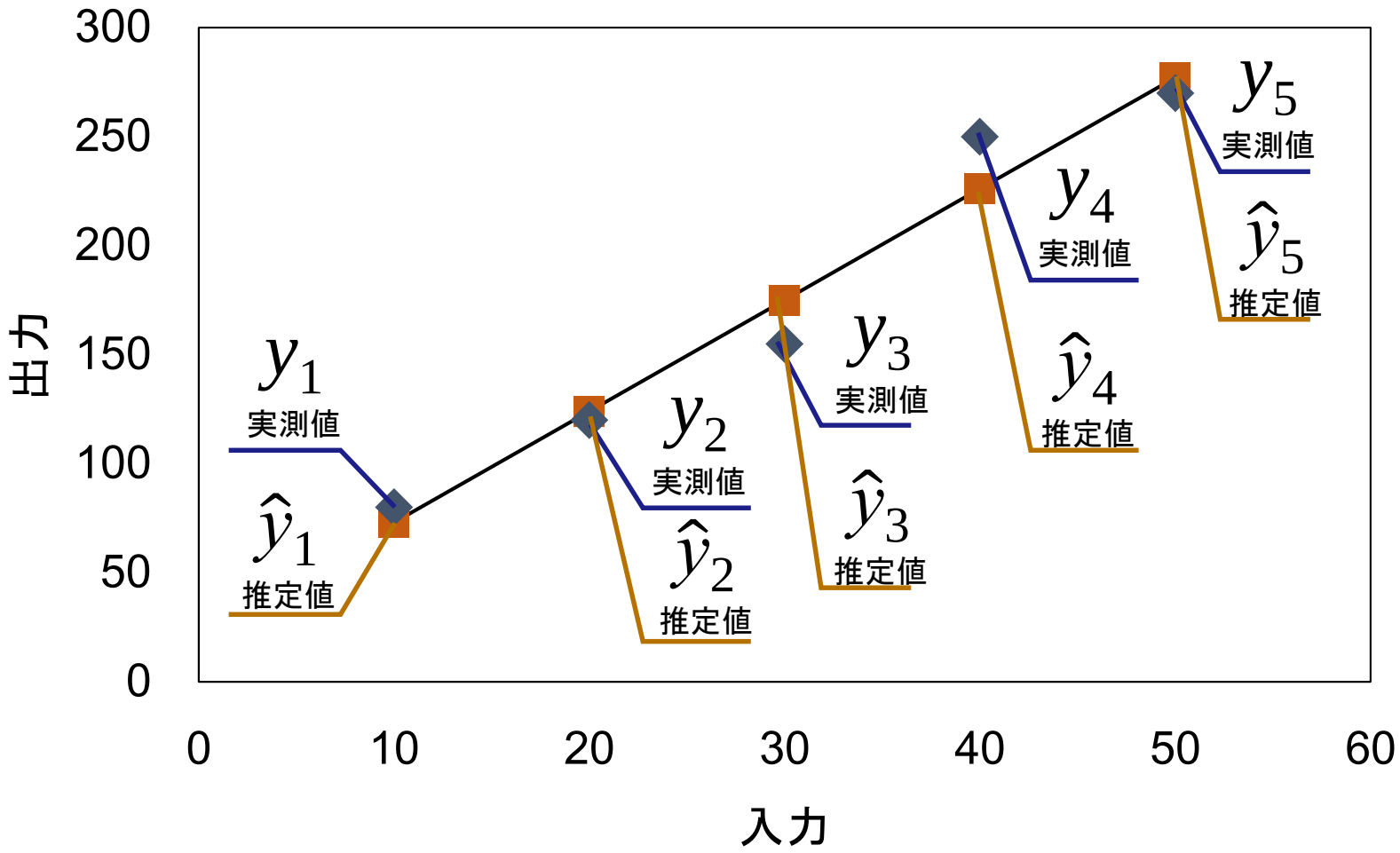
? それらしい線は
沢山引けてしまう

合理的に
式を決める
ために



残差 $\varepsilon_i = y_i - \hat{y}_i$ の2乗和を最小にする.

具体的には



実測値 y_i と推定値 $\hat{y}_i$ は点と点の比較（線はそのあとに引ける）
推定は基本的に入力値のみを使う.

最小二乗法による線形近似

残差ベクトルと残差の二乗和

残差ベクトル

$$\{\varepsilon_i\} = \{y_i\} - \{\hat{y}_i\}$$

線形近似
なので

$$= \{y_i\} - \{ax_i + b\}$$

係数部分と
分離して

$$= \{y_i\} - [\{x_i\}\{1\}] \begin{Bmatrix} a \\ b \end{Bmatrix}$$

残差の二乗和

$$E = \sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2$$

$$= \{\varepsilon_i\}^T \{\varepsilon_i\}$$

残差の二乗和を最小にする

$$E = \{\varepsilon_i\}^T \{\varepsilon_i\}$$

$$= \left[\underbrace{\{y_i\}}_{\text{既知}} - \underbrace{[\{x_i\}\{1\}]}_{\text{既知}} \underbrace{\begin{Bmatrix} a \\ b \end{Bmatrix}}_{\text{未知}}\right]^T \left[\underbrace{\{y_i\}}_{\text{既知}} - \underbrace{[\{x_i\}\{1\}]}_{\text{既知}} \underbrace{\begin{Bmatrix} a \\ b \end{Bmatrix}}_{\text{未知}}\right]$$

既知部分: 実験や統計データを想定

a, b: 未知値 ◀ 変化させて E を最小にする

偏微分して
極値を出す

$$\frac{\partial E}{\partial a} = 0, \quad \frac{\partial E}{\partial b} = 0$$

係数についての偏微分 (1)

条件 (1)

$$\frac{\partial E}{\partial a} = \frac{\partial}{\partial a}\left(\{\varepsilon_i\}^T\{\varepsilon_i\}\right) = \overbrace{\left(\frac{\partial}{\partial a}\{\varepsilon_i\}^T\right)}^{-\{x_i\}^T}\{\varepsilon_i\} + \{\varepsilon_i\}^T\overbrace{\left(\frac{\partial}{\partial a}\{\varepsilon_i\}\right)}^{-\{x_i\}} = 0$$

- Eの定義により分解
- ライプニッツルールにより分解

残差ベクトルとその微分

$$\{\varepsilon_i\} = \{y_i\} - [\{x_i\}\{1\}]\begin{Bmatrix} a \\ b \end{Bmatrix}$$

$$\frac{\partial}{\partial a}\{\varepsilon_i\} = -[\{x_i\}\{1\}]\begin{Bmatrix} 1 \\ 0 \end{Bmatrix} = -\{x_i\}$$

$$\{x_i\}^T\{\varepsilon_i\} = 0$$

$\{x_i\}$と$\{\varepsilon_i\}$が直交

係数についての偏微分 (2)

条件 (2)

$$\frac{\partial E}{\partial b} = \frac{\partial}{\partial b}(\{\varepsilon_i\}^T\{\varepsilon_i\}) = \underbrace{\left(\frac{\partial}{\partial b}\{\varepsilon_i\}^T\right)}_{-\{1\}^T}\{\varepsilon_i\} + \{\varepsilon_i\}^T\underbrace{\left(\frac{\partial}{\partial b}\{\varepsilon_i\}\right)}_{-\{1\}} = 0$$

- Eの定義により分解
- ライプニッツルールにより分解

残差ベクトルとその微分

$$\{\varepsilon_i\} = \{y_i\} - [\{x_i\}\{1\}]\begin{Bmatrix} a \\ b \end{Bmatrix}$$

$$\frac{\partial}{\partial b}\{\varepsilon_i\} = -[\{x_i\}\{1\}]\begin{Bmatrix} 0 \\ 1 \end{Bmatrix} = -\{1\}$$

$$\{1\}^T\{\varepsilon_i\} = 0$$

$\{1\}$と$\{\varepsilon_i\}$が直交

$$[\{x_i\}\{1\}]^T\{\varepsilon_i\} = 0$$

まとめると

得られた条件

$$[\{x_i\}\{1\}]^T\{\varepsilon_i\} = 0$$

X と置いて

残差ベクトルの定義

$$\{\varepsilon_i\} = \{y_i\} - [\{x_i\}\{1\}]\begin{Bmatrix} a \\ b \end{Bmatrix}$$

X と置いて

代入

$$X^T\left(\{y_i\} - X\begin{Bmatrix} a \\ b \end{Bmatrix}\right) = \{0\}$$

変形して

$$X^T X\begin{Bmatrix} a \\ b \end{Bmatrix} = X^T\{y_i\}$$

$$\begin{Bmatrix} a \\ b \end{Bmatrix} = (X^T X)^{-1} X^T\{y_i\}$$

▲ 擬似逆行列X^+

```
pinv(x.^[1,0])
```

もしくは

```
pinv([x,ones(size(x))])
```

最小二乗法のまとめ

最小二乗法は

$\{x_i\}$ と $\{y_i\}$ の関係を $f(x)$ で近似する方法.

実際の入力　実際の出力　モデル関数

$\{\hat{y}_i\} = \{f(x_i)\}$ と $\{y_i\}$ の差 $\{\varepsilon_i\} = \{y_i\} - \{\hat{y}_i\}$

推定値　実際の出力　残差ベクトル

の二乗総和 $E = \{\varepsilon_i\}^T\{\varepsilon_i\}$ が最小になるように f を決める.

残差平方和

線形近似であれば,
偏微分を使い, 整理し, $\begin{Bmatrix} a \\ b \end{Bmatrix} = (X^T X)^{-1} X^T \{y_i\}$ として係数が得られる

▸ 実は線形近似以外でも使える.

具体的な数値を入れた例

具体的な数値を入れた例

x	y
50	105
100	206
150	310
200	380
250	511

というデータがあったとき，

$$\begin{pmatrix} 50 & 1 \\ 100 & 1 \\ 150 & 1 \\ 200 & 1 \\ 250 & 1 \end{pmatrix} \begin{Bmatrix} a \\ b \end{Bmatrix} = \begin{Bmatrix} 105 \\ 206 \\ 310 \\ 380 \\ 511 \end{Bmatrix}$$

を満たせる a, b は 存在しない（変数2に対し式が5できる）

なので残差

$$\begin{Bmatrix} 105 \\ 206 \\ 310 \\ 380 \\ 511 \end{Bmatrix} - \begin{pmatrix} 50 & 1 \\ 100 & 1 \\ 150 & 1 \\ 200 & 1 \\ 250 & 1 \end{pmatrix} \begin{Bmatrix} a \\ b \end{Bmatrix}$$

の二乗を考える

具体的な数値を入れた例

$$
\begin{aligned}
E = \quad & \{105 - (50a + b)\}^2 \\
& +\{206 - (100a + b)\}^2 \\
& +\{310 - (150a + b)\}^2 \\
& +\{380 - (200a + b)\}^2 \\
& +\{511 - (250a + b)\}^2
\end{aligned}
$$

a に注目すれば
下に凸型の2次関数

b に注目しても
下に凸型の2次関数

残差の二乗和を最小にする.

$$
\begin{aligned}
\frac{\partial E}{\partial a} = \quad & -2 \cdot 50 \{105 - (50a + b)\} = 0 \\
& -2 \cdot 100 \{206 - (100a + b)\} \\
& -2 \cdot 150 \{310 - (150a + b)\} \\
& -2 \cdot 200 \{380 - (200a + b)\} \\
& -2 \cdot 250 \{511 - (250a + b)\}
\end{aligned}
$$

具体的な数値を入れた例

$$
\begin{aligned}
-2\cdot 50\,\{105 &- (50a+b)\}\\
-2\cdot 100\,\{206 &- (100a+b)\}\\
-2\cdot 150\,\{310 &- (150a+b)\}\\
-2\cdot 200\,\{380 &- (200a+b)\}\\
-2\cdot 250\,\{511 &- (250a+b)\}
\end{aligned}
= 0
$$

▼ よく見比べてこの式に変形できることを確認して

$$
\begin{Bmatrix} 50 & 100 & 150 & 200 & 250 \end{Bmatrix}
\left(
\begin{Bmatrix} 105 \\ 206 \\ 310 \\ 380 \\ 511 \end{Bmatrix}
-
\begin{pmatrix} 50 & 1 \\ 100 & 1 \\ 150 & 1 \\ 200 & 1 \\ 250 & 1 \end{pmatrix}
\begin{Bmatrix} a \\ b \end{Bmatrix}
\right) = 0
$$

具体的な数値を入れた例

b でも同様に

$$\frac{\partial E}{\partial b} = \begin{array}{l} -2\cdot 1\{105 - (50a + b)\} \\ -2\cdot 1\{206 - (100a + b)\} \\ -2\cdot 1\{310 - (150a + b)\} \\ -2\cdot 1\{380 - (200a + b)\} \\ -2\cdot 1\{511 - (250a + b)\} \end{array} = 0$$

$$\{1\ 1\ 1\ 1\ 1\}\left(\begin{Bmatrix} 105 \\ 206 \\ 310 \\ 380 \\ 511 \end{Bmatrix} - \begin{pmatrix} 50 & 1 \\ 100 & 1 \\ 150 & 1 \\ 200 & 1 \\ 250 & 1 \end{pmatrix}\begin{Bmatrix} a \\ b \end{Bmatrix}\right) = 0$$

具体的な数値を入れた例

$$\begin{bmatrix} 50 & 100 & 150 & 200 & 250 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \left(\begin{Bmatrix} 105 \\ 206 \\ 310 \\ 380 \\ 511 \end{Bmatrix} - \begin{bmatrix} 50 & 1 \\ 100 & 1 \\ 150 & 1 \\ 200 & 1 \\ 250 & 1 \end{bmatrix} \begin{Bmatrix} a \\ b \end{Bmatrix} \right) = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

変形して

$$\begin{bmatrix} 50 & 100 & 150 & 200 & 250 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 50 & 1 \\ 100 & 1 \\ 150 & 1 \\ 200 & 1 \\ 250 & 1 \end{bmatrix} \begin{Bmatrix} a \\ b \end{Bmatrix} = \begin{bmatrix} 50 & 100 & 150 & 200 & 250 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{Bmatrix} 105 \\ 206 \\ 310 \\ 380 \\ 511 \end{Bmatrix}$$

具体的な数値を入れた例

$$\begin{Bmatrix} a \\ b \end{Bmatrix} = \left(\begin{pmatrix} 50 & 100 & 150 & 200 & 250 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 50 & 1 \\ 100 & 1 \\ 150 & 1 \\ 200 & 1 \\ 250 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 50 & 100 & 150 & 200 & 250 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{Bmatrix} 105 \\ 206 \\ 310 \\ 380 \\ 511 \end{Bmatrix}$$

例：他の関数をモデル関数とした場合

それぞれの残差ベクトル $\{\varepsilon_i\}$

(1) 比例の式（$y = ax$）の場合

$$\{\varepsilon_i\} = \{y_i\} - a\{x_i\}$$

(2) 2次関数（$y = ax^2 + bx + c$）の場合

$$\{\varepsilon_i\} = \{y_i\} - \underbrace{[\{x^2{}_i\}\{x_i\}\{1\}]}_{X\text{と置いて}} \begin{Bmatrix} a \\ b \\ c \end{Bmatrix}$$

(3) 正弦関数（$y = a\sin x$）の場合

$$\{\varepsilon_i\} = \{y_i\} - a\underbrace{\{\sin(x_i)\}}_{X\text{と置いて}}$$

スライド9-11の1次関数と同様にたどると，

$$a = (\{x_i\}^T\{x_i\})^{-1}\{x_i\}^T\{y_i\}$$

$$\begin{Bmatrix} a \\ b \\ c \end{Bmatrix} = (X^TX)^{-1}X^T\{y_i\}$$

$$a = (X^TX)^{-1}X^T\{y_i\}$$

一般的な形が見えてきたのでは…?

例題

例題

就労年数と平均給与の関係について

Years of service	<5	<10	<15	<20	<25	<30	<35
Salary (mil. JPY)	370.8	459.4	533.8	597.7	669.7	719.7	753.8

```
years=[5:5:35]';
income=[371;460;534;598;670;720;754];
plot(years,income,"o")
set(gca,"fontsize",14)
X=years.^[1,0];
p=pinv(X)*income;
hold on
plot([0:1:40],p(1)*[0:1:40]+p(2))
axis([0,40,0,900])
```

◀ データ格納（最初から列ベクトル）

上はプライム記号，
下はセミコロン区切りに気を付ける

```
year=
     5
    10
    15
    20
    25
    30
    35
```

```
income=
   371
   460
   534
   598
   670
   720
   754
```

例題

就労年数と平均給与の関係について

Years of service	<5	<10	<15	<20	<25	<30	<35
Salary (mil. JPY)	370.8	459.4	533.8	597.7	669.7	719.7	753.8

```
years=[5:5:35]';
income=[371;460;534;598;670;720;754];
plot(years,income,"o")
set(gca,"fontsize",14)
X=years.^[1,0];
p=pinv(X)*income;
hold on
plot([0:1:40],p(1)*[0:1:40]+p(2))
axis([0,40,0,900])
```

◀ オリジナルデータからのグラフ作成

plot: x軸，y軸，プロットの種類．
第三引数のoはプロットの形
その他は"x"や"@"などいろいろ

gca: graphic current axes

例題

就労年数と平均給与の関係について

Years of service	<5	<10	<15	<20	<25	<30	<35
Salary (mil. JPY)	370.8	459.4	533.8	597.7	669.7	719.7	753.8

```
years=[5:5:35]';
income=[371;460;534;598;670;720;754];
plot(years,income,"o")
set(gca,"fontsize",14)
X=years.^[1,0];
p=pinv(X)*income;
hold on
plot([0:1:40],p(1)*[0:1:40]+p(2))
axis([0,40,0,900])
```

years（横軸）から係数行列を作成

要素ごと計算なので，

$$X = \begin{array}{cc} \texttt{year.^1} & \texttt{year.^0} \\ \end{array} \begin{pmatrix} 5 & 1 \\ 10 & 1 \\ 15 & 1 \\ 20 & 1 \\ 25 & 1 \\ 30 & 1 \\ 35 & 1 \end{pmatrix}$$

例題

就労年数と平均給与の関係について

Years of service	<5	<10	<15	<20	<25	<30	<35
Salary (mil. JPY)	370.8	459.4	533.8	597.7	669.7	719.7	753.8

```
years=[5:5:35]';
income=[371;460;534;598;670;720;754];
plot(years,income,"o")
set(gca,"fontsize",14)
X=years.^[1,0];
p=pinv(X)*income;
hold on
plot([0:40],p(1)*[0:40]+p(2))
axis([0,40,0,900])
```

hold on で前のグラフを保持
（これがないと，次の plot の指示で
前のグラフは消える）

近似直線を上書き．
なお plot の第三引数は省略

x軸とy軸の範囲を指定

例題

就労年数と平均給与の関係について

Years of service	<5	<10	<15	<20	<25	<30	<35
Salary (mil. JPY)	370.8	459.4	533.8	597.7	669.7	719.7	753.8

別解

```
years=[5:5:35]';
income=[371;460;534;598;670;720;754];
X=years.^[1,0];
p=pinv(X)*income;
plot(years,income,"o",[0:40], [0:40]'.^[1,0]*p)
set(gca,"fontsize",14)
axis([0,40,0,900])
```

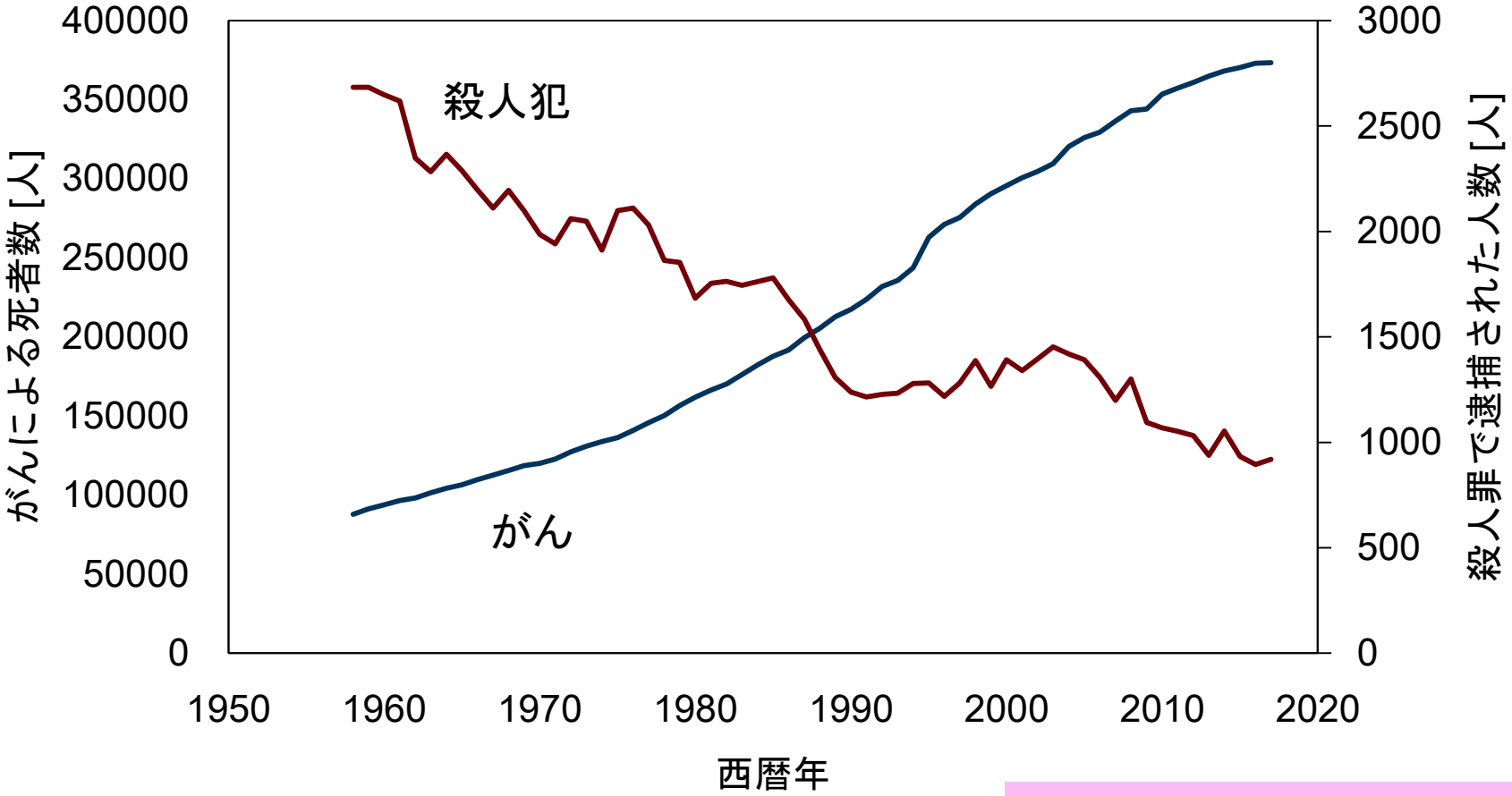
スクリプトファイルの場合，「hold on」せず，また足し算もまとめて

```
plot(years,income,"o", [0:40],[0:40]'.^[1,0]*p)
```

としてもよい．

課題

課題　がん死亡者数と殺人罪による逮捕者数



出典: がんに関する統計データ，国立がん研究センター
平成30年版犯罪白書，警察庁

data05.csv
というファイルをダウンロードし，
殺人犯数とがんによる死者数の
相関関係を示せ.

課題について

課題1

最小二乗法による線形フィッティングを行い，
結果を描画するスクリプトファイルについて作成せよ.

補足1: データは授業HPからダウンロードできる.

補足2: データのロードは，作業ディレクトリにファイルを置き

```
>> x=csvread("data05.csv");
```

で行う. ▸ xにcsvファイルのデータが格納される

1列目は年数（今回は使わない），2列目ががんによる死者数，3列目が殺人罪による逮捕者数

課題2

最小二乗法により $a_k x^k + a_1 x + a_0$ という形式で
フィッティングを行い，
結果を描画するスクリプトファイルについて作成せよ.
ここで k は各自の学籍番号の整数部の1/1000とする.
（すなわちB7TB1357であれば $k = 1.357$）

補足

この課題で知ってほしいことは
殺人罪で検挙される数が減ればがんによる死者数が増える
ということではない.（年数の変化に相関がある2つを比べただけ.）

知ってほしいことは,
依存性がない事柄でも相関は取れることがある.
ということである.

統計は時として詐欺師に都合のよい学問となる.

"There are three kinds of lies:
lies, damned lies, and statistics."

Appendix

おまけ　多項式近似に拡張すると

$$E = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{m} a_j x^{j-1}\right)^2$$

Octaveでは
x.^[m-1:-1:0]
と容易に表現できる.

要素を (*m*-1)乗　要素を (*m*-2)乗　1乗　0乗

残差ベクトル

$$\{\varepsilon_i\} = \{y_i\} - \left[\{x_i^{m-1}\}\{x_i^{m-2}\}\ldots\{x_i\}\{1\}\right]\{a_j\}$$

*n*次列ベクトル　*m*次列ベクトル

$$= \{\varepsilon_i\}^T\{\varepsilon_i\}$$

$$\frac{\partial \mathrm{E}}{\partial a_j} = 0\ (j = 1 \cdots m)$$

$$\boxed{\{a_j\} = (X^T X)^{-1} X^T \{y_i\}}$$

さらなる一般化は各自調べてみよう.

おまけ　Octaveには多項式近似関数がある

▶ x y を入出力の列ベクトルとして，

```
polyfit(x, y, n)
```

1行(n+1)列の計算結果を出力

ans = ○ ○ ・・・・ ○

x^nの係数　x^{n-1}の係数　定数項

```
pinv(x.^[n:-1:0])*y
```

ans =

○ ◀ x^nの係数

○ ◀ x^{n-1}の係数

・

・

○ ◀ 定数項