

# 統計的モデリング基礎③ ～回帰モデリング～

鹿島久嗣  
(情報学科 計算機科学コース)

# 回帰

# 回帰：

## 片方の変数でもう片方の変数を説明

---

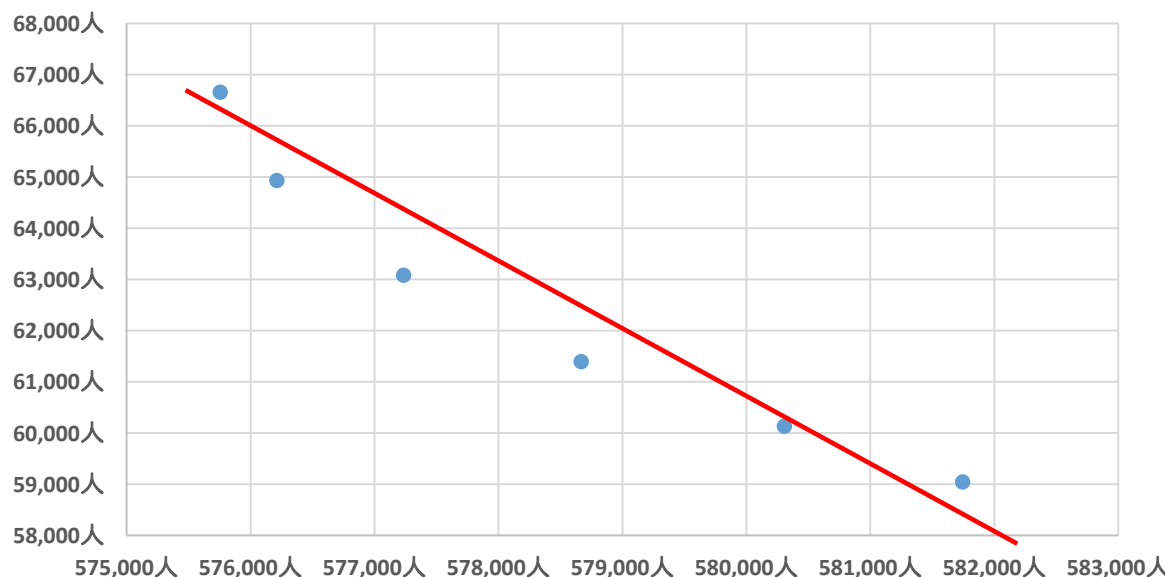
- 相関 (correlation) は二変数  $x, y$  を区別せずに対等に扱う
  - 一方が増えたときに他方が増える (減る) 関係性を調べる
  - 例：身長と体重
- 回帰 (regression) は変数  $x$  で変数  $y$  を説明する
  - 一方から他方が決定される様子や程度を調べる
  - 例：年齢と血圧、所得と消費
  - $x$  を独立変数・説明変数、 $y$  を従属変数・応答変数などによぶ

# 回帰の問題：

片方の変数からもう片方を説明するモデルをデータから推定

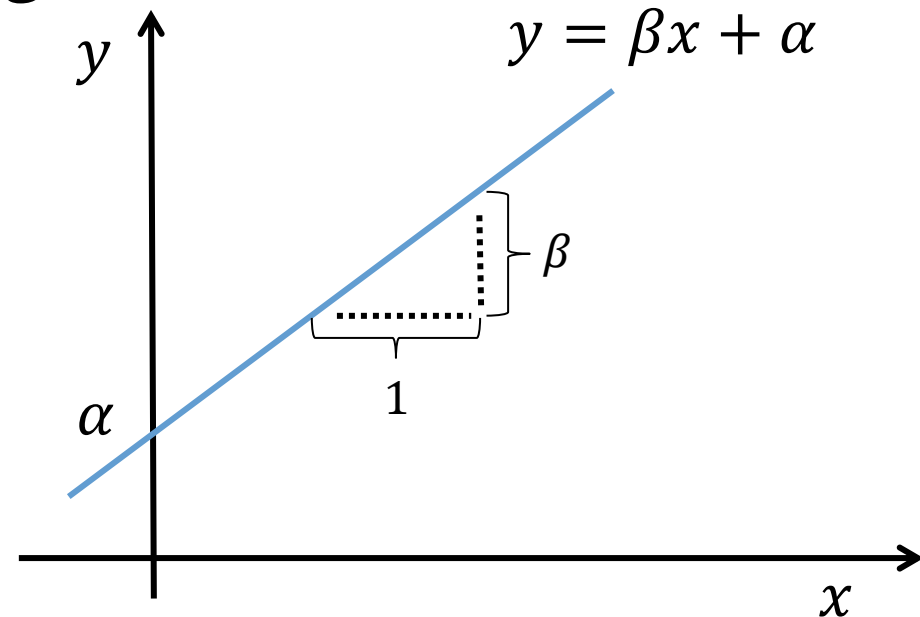
- 2つの変数  $x$  と  $y$  の組について  $N$  組のデータがある
  - $\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$  と  $\mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(n)})$
- $y$  を  $x$  で説明（予測）するモデル  $g$  がほしい
  - 概ね  $y = g(x)$  となる  $g$
  - 例えば直線を  $g$  として仮定
- $g$  の使い道：
  - 予測
  - 因果関係の発見  
(ただし注意が必要)

国家公務員数 vs 特定独立行政法人職員数



# 基本的な回帰モデル： 線形回帰モデル

- 線形モデル： $y = g(x) = \beta x + \alpha$ 
  - $\beta$ ：傾きパラメータ（ $x$ が1増えると、 $y$ が1増える）
  - $\alpha$ ：切片パラメータ
- $x$ と $y$ の間に直線的な関係を仮定する
  - $y$ が $x$ の線形関数に依存



# 回帰モデルのパラメータ推定問題の定式化：

## モデルとデータの食い違いを最小化する最小二乗法

- データ：  $\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$  と  $\mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(n)})$

- モデルの出力する予測値：  $\hat{y}^{(i)} = \beta x^{(i)} + \alpha$

- モデルの予測と実際のデータとの食い違いを定義する：

$$\ell(\alpha, \beta) = \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2 = \sum_{i=1}^n (y^{(i)} - (\beta x^{(i)} + \alpha))^2$$

- 食い違いを二乗誤差で測る

- 最適化問題（最小化）：

$$(\hat{\alpha}, \hat{\beta}) = \operatorname{argmin}_{\alpha, \beta} \ell(\alpha, \beta)$$

# 最小二乗法の解： 二乗誤差を最小化する解

- $\ell(\alpha, \beta)$  を  $\alpha$  と  $\beta$  で偏微分して0とおいて、解くと：

$$\hat{\beta} = \frac{\sum_i (x^{(i)} - \bar{x})(y^{(i)} - \bar{y})}{\sum_i (x^{(i)} - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\alpha} = \bar{y} - \beta \bar{x}$$

- $x$  と  $y$  の共分散： $S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x^{(i)} - \bar{x})(y^{(i)} - \bar{y})$
- $x$  の不偏分散： $S_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x^{(i)} - \bar{x})^2$

# 最小二乗法の性質： 不偏性と推定精度

- いくつかの仮定の下で不偏性をもつ
  - 母集団において  $\epsilon^{(i)} = y^{(i)} - (\beta^* + \alpha^* x^{(i)})$  が同一の分布に従い一定の分散  $\sigma^2$ 、互いに無相関、 $\epsilon_i$  と  $x_i$  が無相関などの仮定
  - 不偏性：  $E[\hat{\beta}] = \beta^*, E[\hat{\alpha}] = \alpha^*$  （標本の取り方についての期待値）
- $\text{Var}[\hat{\beta}] = \frac{\sigma^2}{\sum_{i=1}^n (x^{(i)} - \bar{x})^2}$ ：広範囲の  $x^{(i)}$  があったほうが精度がよい
- $\text{Var}[\hat{\alpha}] = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}}{\sum_{i=1}^n (x^{(i)} - \bar{x})^2} \right)$ ：原点付近の  $x^{(i)}$  があったほうが精度がよい



# 決定係数：

従属変数をモデルがどの程度説明できたかを測る

- 決定係数  $R^2$  : モデルの予測値  $\hat{\mathbf{y}} = (\hat{y}^{(1)}, \hat{y}^{(2)}, \dots, \hat{y}^{(n)})$  とデータ  $\mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(n)})$  との相関係数の2乗

$$R^2 = \frac{(\sum_{i=1}^n (\hat{y}^{(i)} - \bar{y})(y^{(i)} - \bar{y}))^2}{(\sum_{i=1}^n (\hat{y}^{(i)} - \bar{y})^2) (\sum_{i=1}^n (y^{(i)} - \bar{y})^2)} = \frac{\sum_{i=1}^n (\hat{y}^{(i)} - \bar{y})^2}{\sum_{i=1}^n (y^{(i)} - \bar{y})^2}$$

- $\mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(n)})$  の変動（分母）のうち、  
回帰式が説明できる変動（分子）の割合
- 相関係数は  $-1 \leq R \leq 1$  なので、決定係数  $0 \leq R^2 \leq 1$ 
  - 決定係数が1に近いほどデータへのモデルの当てはまりがよい

# 決定係数： 従属変数をモデルがどの程度説明できたかを測る

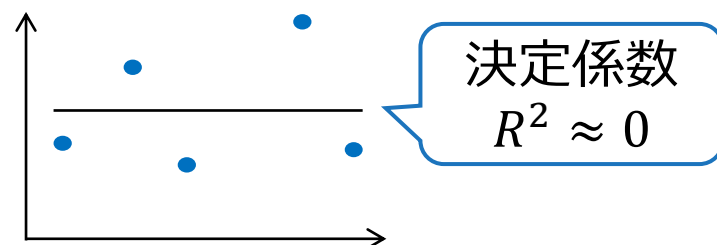
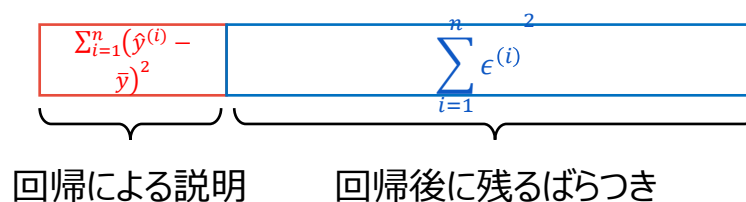
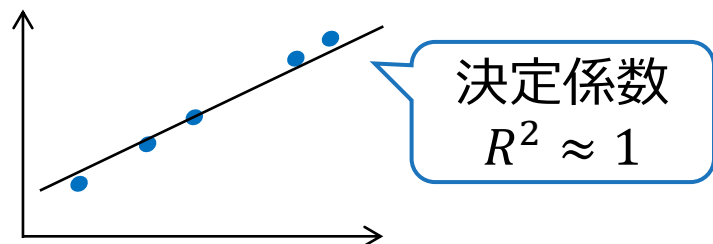
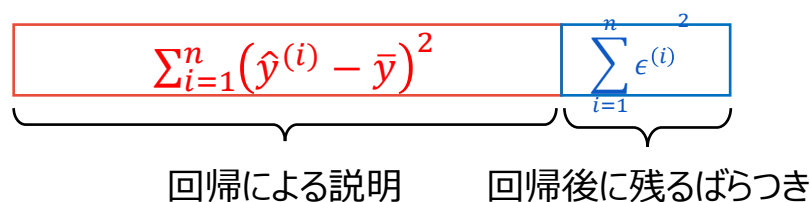
## ■ $y$ の変動の分解：

$$\sum_{i=1}^n (y^{(i)} - \bar{y})^2 = \sum_{i=1}^n (\hat{y}^{(i)} - \bar{y})^2 + \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2$$

$y$  の変動

回帰式の予測  $\hat{y}^{(i)}$   
が説明できる変動

残差の平方和  
 $\sum_{i=1}^n \epsilon^{(i)2}$



# 課題： 回帰モデリングを試してみよう！

- 自分でデータを見つけよう！
  - 従属変数と独立変数を決めよう！
  - データ： $\mathbf{x} = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$  と  $\mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(n)})$
- 回帰モデルを推定してみよう！： $\hat{y}^{(i)} = \beta x^{(i)} + \alpha$

$$\hat{\beta} = \frac{\sum_i (x^{(i)} - \bar{x})(y^{(i)} - \bar{y})}{\sum_i (x^{(i)} - \bar{x})^2}, \quad \hat{\alpha} = \bar{y} - \beta \bar{x}$$

- 決定係数を計算、データと回帰モデルをプロットしてみよう！
- 推定に使用しないデータに対しても、予測を評価してみよう



# 重回帰

# 重回帰： 複数の独立変数を用いて予測

- (単) 回帰では、ひとつの独立変数から予測を行う

$$g(x) = \beta x + \alpha$$

- 例：年齢から年収を予測する

$$(\text{年収}) = \beta \times (\text{年齢}) + \alpha$$

- 重回帰では複数の ( $m$ 個の) 独立変数を用いる

$$g(x) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \alpha$$

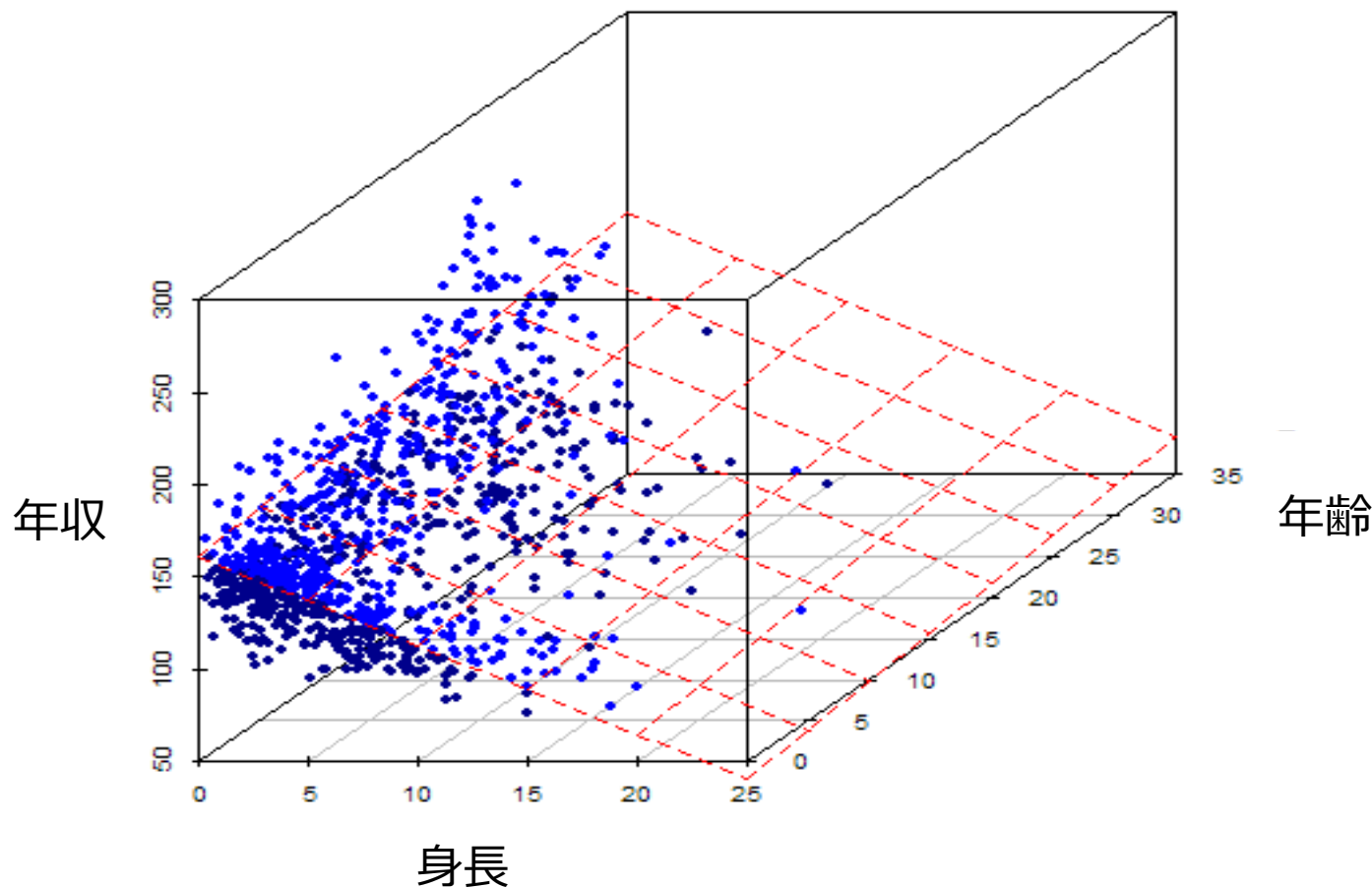
- 例：年齢と身長から年収を予測する

$$(\text{年収}) = \beta_{(\text{年齢})} \times (\text{年齢}) + \beta_{(\text{身長})} \times (\text{身長}) + \alpha$$

# 重回帰のイメージ：

(超) 平面でデータに当てはめる

- 単回帰では直線で近似、重回帰では (超) 平面で近似



# 重回帰モデルの推定問題：

## 最小二乗法によってパラメータを推定する

- 単回帰と同じく、モデルの予測と実際のデータとの食い違いを二乗誤差で測る

$$\begin{aligned}\ell(\alpha, \{\beta_i\}_{i=1}^m) &= \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2 \\ &= \sum_{i=1}^n \left( y^{(i)} - \left( \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)} + \cdots + \beta_m x_m^{(i)} + \alpha \right) \right)^2\end{aligned}$$

- 最適化問題（最小化）を解いてパラメータ推定値を求める：

$$(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m) = \operatorname{argmin}_{\alpha, \{\beta_i\}_{i=1}^m} \ell(\alpha, \{\beta_i\}_{i=1}^m)$$

- すべてのパラメータについて偏微分して0とおき連立方程式を得る

# 行列とベクトルを用いた表記：

## 行列とベクトルを用いて書き換えると便利

- モデル：  $y = \boldsymbol{\beta}^\top \mathbf{x}$ 
  - パラメータ：  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m, \alpha)^\top$
  - 独立変数：  $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_m^{(i)}, 1)^\top$  } 最後の次元は切片部分に相当
- 目的関数：  $\ell(\boldsymbol{\beta}) = \sum_{i=1}^n (y^{(i)} - \hat{y}^{(i)})^2$   
 $= \sum_{i=1}^n (y^{(i)} - \boldsymbol{\beta}^\top \mathbf{x}^{(i)})^2 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$ 
  - 計画行列：  $\mathbf{X} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)})^\top$
  - 従属変数：  $\mathbf{y} = (y^{(1)}, y^{(2)}, \dots, y^{(N)})^\top$



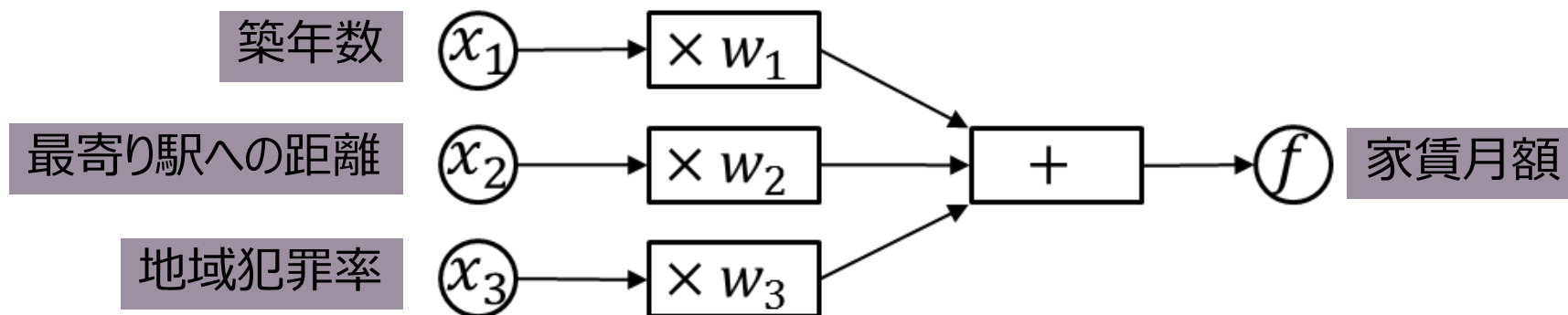
# 例： 家賃予測

- 計画行列：4件の賃貸住宅

$$\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}]^T = \left[ \begin{pmatrix} 15 \\ 10 \\ 1.0 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 0.1 \end{pmatrix}, \begin{pmatrix} 35 \\ 5 \\ 7.0 \end{pmatrix}, \begin{pmatrix} 40 \\ 70 \\ 1.0 \end{pmatrix} \right]^T$$

- 独立変数（ベクトル）：4件分の家賃

$$\mathbf{y} = (y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)})^T = (140, 85, 220, 115)^T$$



# 重回帰モデルの解： 解析解が得られる

- 目的関数： $L(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$
- 解： $\boldsymbol{\beta}^* = \operatorname{argmin}_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ 

ただし、本当に（数値的に）解くときには連立方程式のほうを解く
- ただし、解が存在するためには $\mathbf{X}^\top \mathbf{X}$ が正則である必要
  - モデルの次元数 $m$ よりもデータ数 $n$ が大きい場合はおおむね成立
- 正則化：正則でない場合には $\mathbf{X}^\top \mathbf{X}$ の対角成分に正の定数 $\lambda > 0$ を加えて正則にする
  - 新たな解： $\boldsymbol{\beta}^* = \operatorname{argmin}_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}$
  - 目的関数に戻すと： $L(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2$

パラメータのノルムに関する  
ペナルティ項

## 多重共線性：

### 独立変数間に強い相関がある場合には注意

- 重回帰モデルにおいて、独立変数間に強い相関がある場合には推定されたパラメータの分散が大きくなり、信頼性が下がる
  - どちらでも説明できるので、パラメータの重みを奪い合う
  - 例：年齢と勤続年数など
- 予測には影響しないが、得られたモデル（パラメータ）を解釈したい場合には注意を要する
  - 相関が強い場合には、片方ずつ用いた結果を調べるなどを行う

# 質的変数の取り扱い

# 質的変数の扱い： ダミー変数の利用

- 独立変数が質的変数（記号を値としてとる）の場合
  - 例：{右, 左}、{京都, 大阪, 東京}
- ダミー変数：{0,1}の2値をとる変数
  - {右, 左}を{0,1}として表現
  - 3値以上の場合には、選択肢数-1個のダミー変数を用いる：  
京都 = (1,0)、大阪 = (0,1)、東京 = (0,0)

東京をベースラインとして各地域の差分を示す

- 例：年齢と性別から年収を予測する

$$(\text{年収}) = \beta_1 \times (\text{年齢}) + \beta_2 \times (\text{性別}) + \alpha$$

- 性別が男性であるか {0(No), 1(Yes)} のダミー変数

## 従属変数が質的変数の場合：

ダミー変数を従属変数として回帰を適用（が、やや不適）

- 従属変数が質的変数の場合
  - 例：年収と年齢から性別を当てる
- 従属変数をダミー変数として回帰を適用する
  - 例：(性別) =  $\beta_1 \times (\text{年齢}) + \beta_2 \times (\text{年収}) + \alpha$
- 回帰モデルの適用は厳密にはちょっと変
  - 回帰モデルは連続値を出力するが、本来、性別にあたるダミー変数は $\{0,1\}$ のいずれかの値のみをとる
  - 最小二乗法が仮定している均一分散性が成立しない
    - 「効率性」が満たされないため推定値のバラつきが大きい

# 質的変数・非線形回帰

# 非線形回帰：

## 線形回帰に非線形性を導入する

---

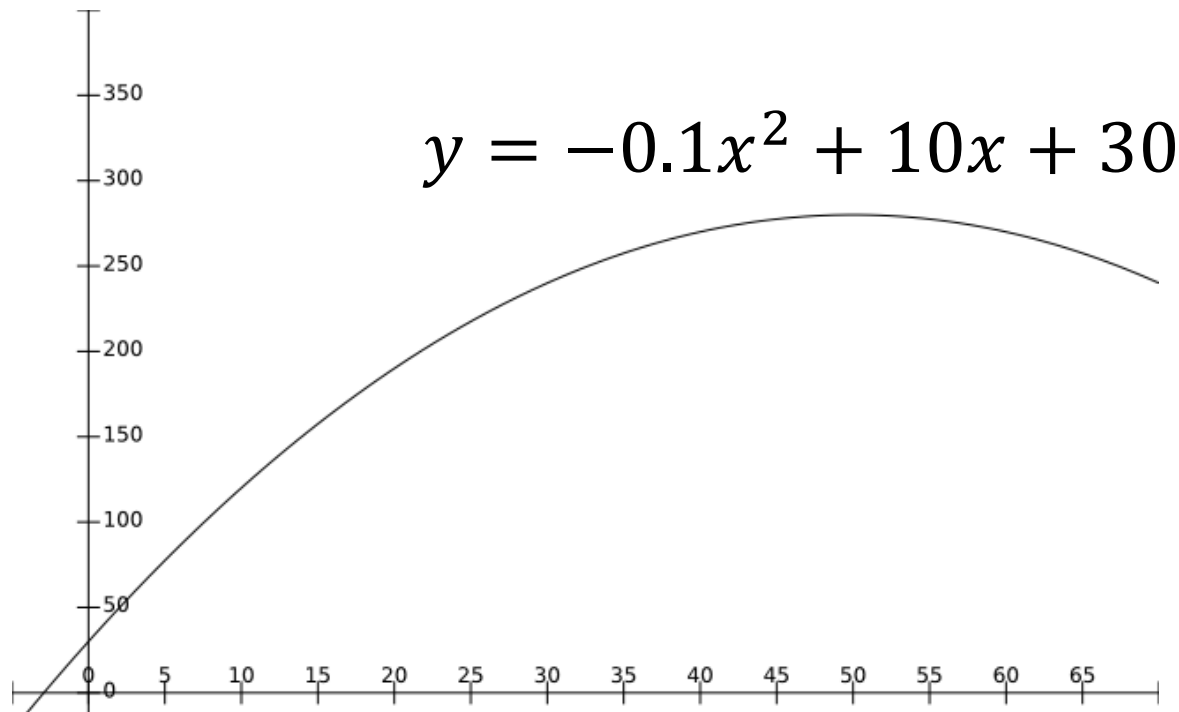
- ここまでは線形モデルを仮定してきた： $y = \boldsymbol{\beta}^\top \mathbf{x}$ 
  - パラメータ： $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m, \alpha)^\top$
  - 独立変数： $\mathbf{x} = (x_1, x_2, \dots, x_m, 1)^\top$
  - シンプルで安定して扱いやすい
- 線形モデルに非線形性を導入するにはどうしたらよいか？
  1. 変数変換（例： $x \rightarrow \log x$ ）
  2. 交差項（例： $x_1, x_2 \rightarrow x_1 x_2$ ）
  3. カーネル法



# 変数変換： 簡単に非線形性を導入する方法

- 独立変数に対して非線形の変換を適用する：

$$x \rightarrow \log x, e^x, x^2, \frac{1}{x}, \dots$$



# 変数の対数変換： 傾きパラメータ $\beta$ の意味が異なる

- $y = \beta x + \alpha$  の独立変数 ( $x$ ) と従属変数 ( $y$ ) は対数変換して用いられることがある
- 変換と係数の意味

		従属変数	
		$y$	$\log y$
独立変数	$x$	$y = \beta x + \alpha$ $x$ が1単位増加すると $y$ が $\beta$ 単位増加する	$\log y = \beta x + \alpha$ $x$ が1単位増加すると $y$ が $1 + \beta$ 倍になる
	$\log x$	$y = \beta \log x + \alpha$ $x$ を2倍すると $y$ が $\beta$ 単位増加する	$\log y = \beta \log x + \alpha$ $x$ を2倍すると $y$ が $1 + \beta$ 倍になる

# 交差項： 変数の組み合わせを導入

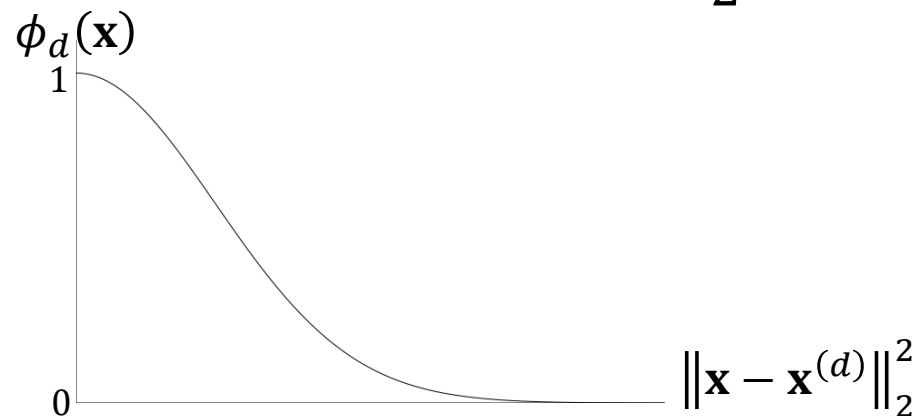
- もともとの独立変数  $x_1, x_2, \dots, x_m$  に加えて、2変数の交差項  $\{x_d x_{d'}\}_{d,d'}$  を用いる
  - ダミー変数の交差項は2変数のANDに相当
- すべての交差項を採用すると行列パラメータ  $\mathbf{B}$  を導入して  $y = \mathbf{x}^\top \mathbf{B}^\top \mathbf{x}$  と書くことができる

$$y = \text{Trace} \left( \underbrace{\begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,m} \\ \vdots & \ddots & \vdots \\ \beta_{m,1} & \cdots & \beta_{m,m} \end{bmatrix}}_{\mathbf{B}}^\top \underbrace{\begin{bmatrix} x_1^2 & x_1 x_2 & \cdots & x_1 x_m \\ x_2 x_1 & x_2^2 & \cdots & x_2 x_m \\ \vdots & \vdots & \ddots & \vdots \\ x_m x_1 & x_m x_2 & \cdots & x_m^2 \end{bmatrix}}_{\mathbf{x} \mathbf{x}^\top} \right)$$

# カーネル回帰：

## カーネル関数を用いた非線形性の導入

- 前述の変数変換アプローチを一般化する
- 線形モデル  $y = \boldsymbol{\beta}^\top \mathbf{x}$  において、 $d$  番目の独立変数  $x_d$  を「カーネル関数」をもちいた基底  $\phi_d(\mathbf{x})$  で与える
  - カーネル関数  $\phi_d(\mathbf{x})$ ：  
独立変数  $\mathbf{x}$  に何らかの非線形変換を適用したもの
- カーネルの例：ガウスカーネル  $\phi_d(\mathbf{x}) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}^{(d)}\|_2^2)$ 
  - 要するに、 $d$  番目のデータとの「類似度」のようなもの



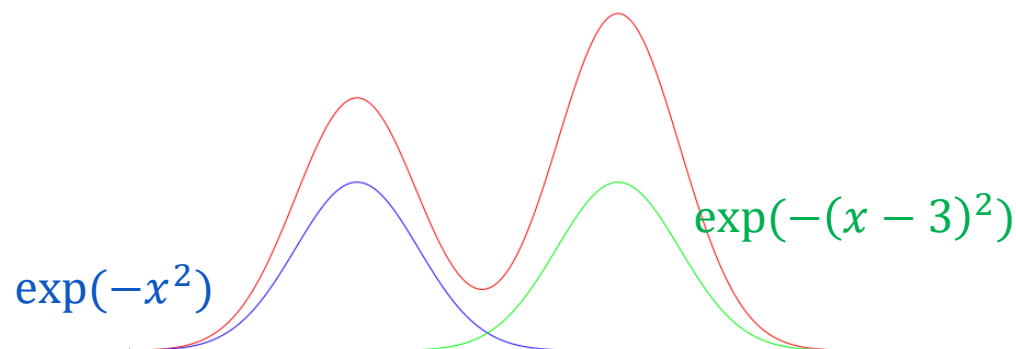
# カーネル回帰： カーネル関数を用いた非線形性の導入

## ■ カーネル回帰モデル：

$$y = \boldsymbol{\beta}^\top \boldsymbol{\Phi}(\mathbf{x}) = \beta_1 \phi_1(\mathbf{x}) + \beta_2 \phi_2(\mathbf{x}) + \cdots + \beta_n \phi_n(\mathbf{x}) + \alpha$$

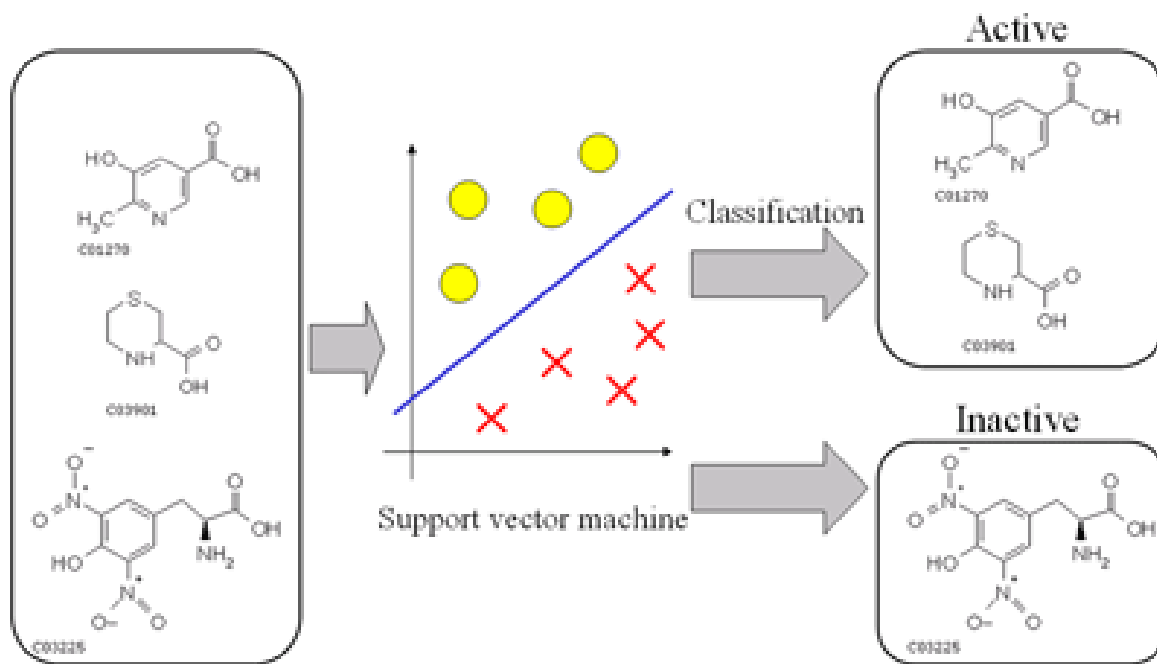
- モデルの次元数 $n$ は、もとの $\mathbf{x}$ の次元数 $m$ とは異なることに注意
  - 通常はモデルの次元数 $n$  = データサイズにとる
  - $\phi_d(\mathbf{x})$ は $\mathbf{x}$ と $\mathbf{x}^{(d)}$ の類似度を表すカーネル関数

## ■ $n = 2, m = 1$ の例： $y = 1.5 \exp(-x^2) + 2 \exp(-(x - 3)^2)$



# さまざまなカーネル関数： カーネル関数を変えれば様々なデータに対応可能

- カーネル回帰はカーネル関数の定義を変えることで、任意の対象を扱うことができる
  - 独立変数がベクトル  $\mathbf{x} = (x_1, x_2, \dots, x_m, 1)^T$  である必要すらない
- カーネル関数によって、系列、木、グラフなども扱うことができる



# まとめ： 回帰モデリング

---

- 回帰では、（1個ないし複数の）独立変数から従属変数を説明・予測するモデルを作る
- 線形回帰モデル：独立変数が線形に効くモデル
- 最小二乗法によって回帰モデルのパラメータが求まる
- モデルの当てはまりは決定係数によって測る
- 変数変換や交差項などによって非線形性を導入できる