

平成24年度 東京大学理学部情報科学科

月曜2限10:30~12:00

理7-214

知識処理論

2012年5月21日(月)

井元清哉

東京大学医科学研究所

ヒトゲノム解析センター

DNA情報解析分野

<http://bonsai.hgc.jp/~imoto>

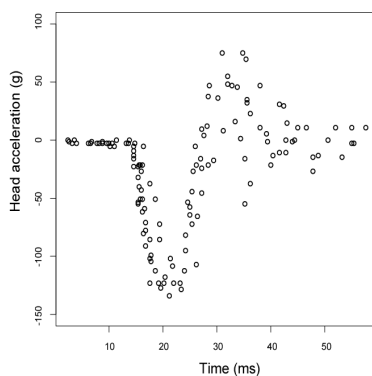
imoto@ims.u-tokyo.ac.jp

線形から非線形へ

データ: $(x_1, y_1), \dots, (x_n, y_n)$; n 個の観測値

線形回帰モデル:

$$y_i = \beta_0 + \beta_1 x_i + e_i \quad E[e_i] = 0, \text{Var}[e_i] = \sigma^2$$



XとYの関係が線形ではない。

$Y = m(x)$ となるような関数 $m(x)$ は?

パラメトリック回帰モデル

Y と X の関係にある関数を仮定し、関数に含まれるパラメータをデータから推定する

例

$$\text{多項式 (1次, 2次, \dots)} \quad y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

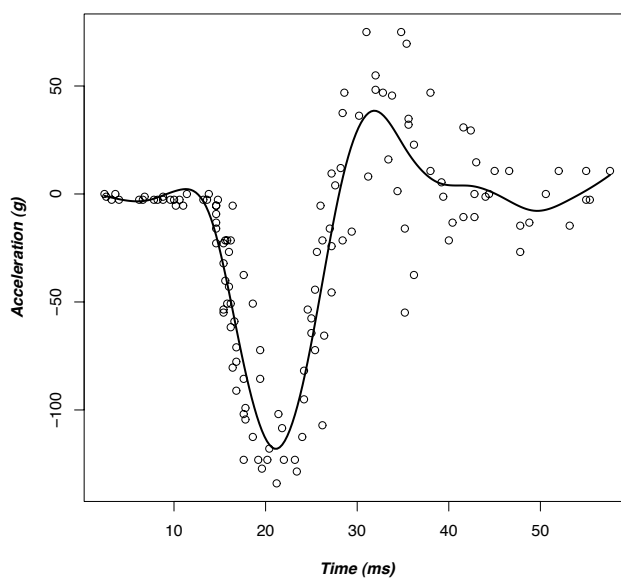
$$\text{対数関数} \quad y = \beta_0 \log(\beta_1 + \beta_2 x) + \varepsilon$$

$$\text{正弦関数} \quad y = \beta_0 \sin(\beta_1 + \beta_2 x) + \varepsilon$$

などなど

Y と X の関係にある程度の知見が必要

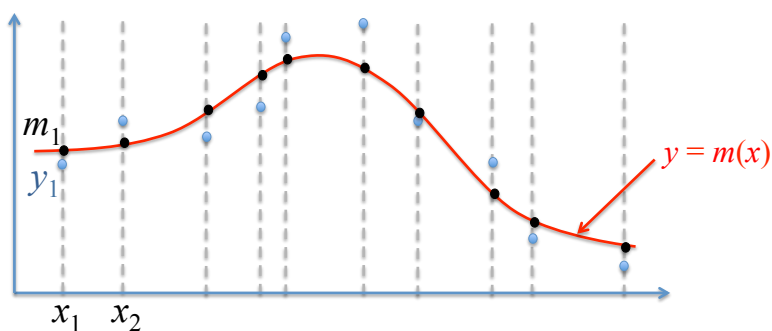
特定の関数型を仮定することなくデータに自分自身の構造を語らせる
⇒ ノンパラメトリック回帰



ノンパラメトリック回帰の手法の例

- スプライン
 - 平滑化スプライン
 - B-スプライン
- カーネル法
 - 局所多項式（局所尤度法）
 - サポートベクトル回帰
- ニューラルネットワーク
 - 階層型ニューラルネットワーク
 - 動径基底関数ネットワーク

ノンパラメトリック回帰：基本概念

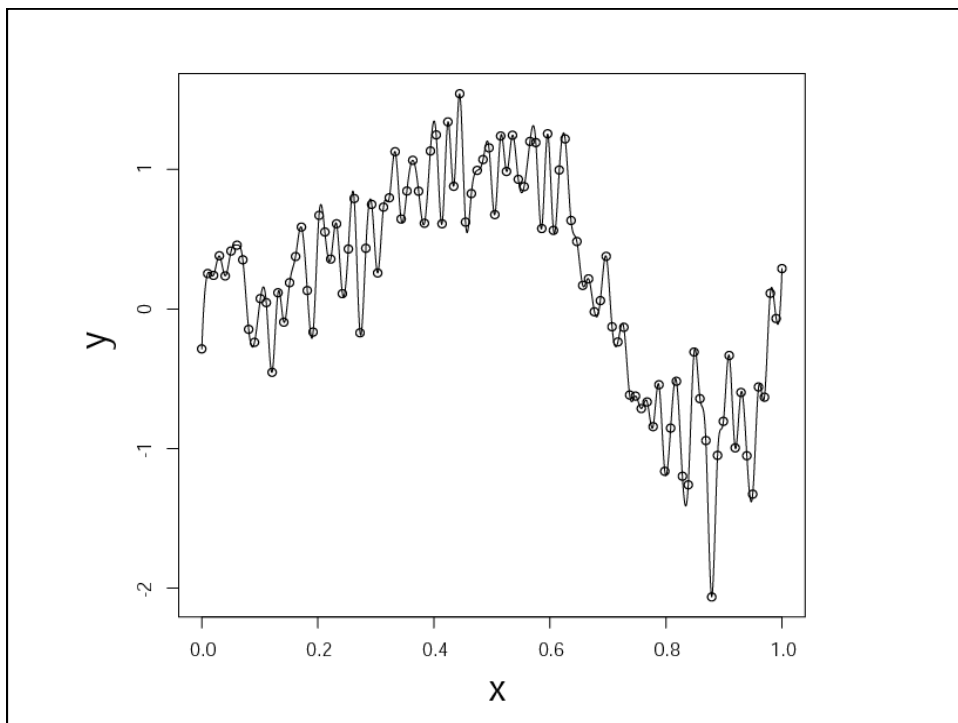
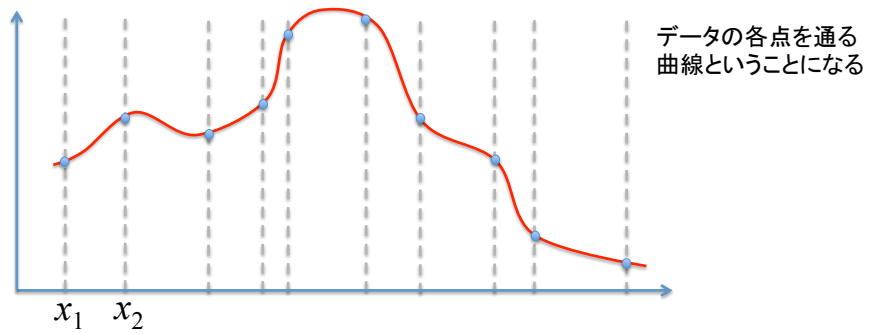


関数 $y = m(x)$ のデータ点での値 (x_i, m_i) の m_i だけを推定の対象とする

もし、 $\mathbf{m} = (m_1, \dots, m_n)^T$ を最小2乗法で推定すると

$$S(\mathbf{m}) = \sum_{i=1}^n (y_i - m_i)^2$$

$$\operatorname{argmin}_{\mathbf{m}} S(\mathbf{m}) = \mathbf{y} = (y_1, \dots, y_n)^T$$



最小2乗法の修正

$\mathbf{m} = (m_1, \dots, m_n)^T$ に対する最小2乗法

$$S(\mathbf{m}) = \sum_{i=1}^n (y_i - m_i)^2$$

に Ridge 推定量や Lasso 推定量、Elastic net で行ったような修正を加えて
うまく \mathbf{m} を推定できないか？

Roughness Penalty

データ (x_i, y_i) に対してある関数 $y = m(x)$ を以下の式を最小化することによって
当てはめる。

$$S_p(\mathbf{m}) = \sum_{i=1}^n \{y_i - m(x_i)\}^2 + \lambda \int_a^b \{m''(t)\}^2 dt$$

スプライン曲線

データ $(x_i, y_i), i = 1, \dots, n, (n \geq 2)$

$$a < x_1 < x_2 < \dots < x_n < b$$

条件 $y_i = w(x_i)$ を満たす関数 w が

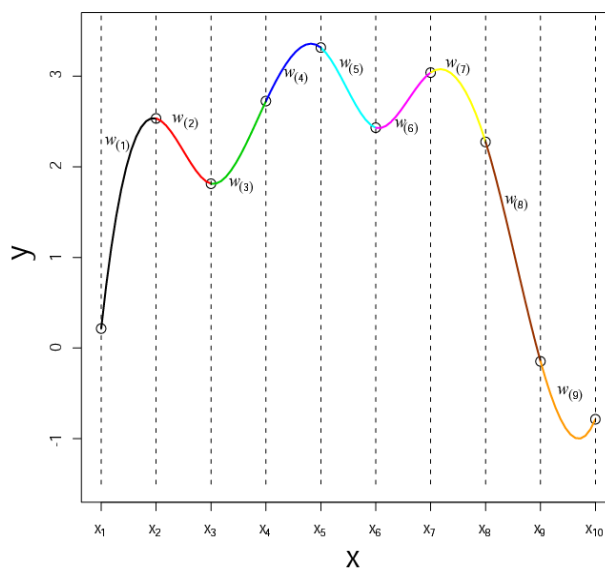
(I) 各小区間 $(a, x_1), (x_1, x_2), \dots, (x_n, b)$ で次数 r の多項式

(II) 区間 $[a, b]$ において, $r-1$ 階までの導関数が連続
を満たすとき, w を次数 r のスプライン関数と呼ぶ

2次導関数が小区間 $[a, x_1], [x_n, b]$ で 0 の3次スプラインを
特に, **自然3次スプライン**と呼ぶ

点 x_1, x_2, \dots, x_n は節点と呼ばれる

自然3次スプライン



スプライン: 表現の一意性

問題

n ($n \geq 2$) 個の実数値の組 (x_i, y_i) , $i = 1, \dots, n$ に対して、条件 $y_i = w(x_i)$ を満たす自然3次スプラインは一意に定まることを示せ。ただし、 $a < x_1 < \dots < x_n < b$ とする。

スプライン曲線: パラメータ化

X と Y に関して観測されたデータを (x_i, y_i) , $i = 1, \dots, n$ とする。

このとき、モデルを $y_i = w(x_i) + \varepsilon_i$ とする。

無限次元パラメータ

各点 x_i に対する $w(x_i)$ を改めて w_i とパラメタライズする。

すなわち、パラメータは $w = (w_1, \dots, w_i, \dots, w_n)^T$ となる。

無限次元パラメータから有限次元パラメータへ

スプライン曲線の推定

データ $(x_i, y_i), i = 1, \dots, n$

モデル $y_i = w_i + \varepsilon_i$ パラメータ $w = (w_1, \dots, w_n)^T$

パラメータ w の推定

最小2乗法

$$S(w) = \sum_{i=1}^n (y_i - w_i)^2$$

S が最小となるようにパラメータを推定

$w_i = y_i$ のとき $S = 0$ となる

スプライン曲線: 最小2乗推定量

データ $(x_i, y_i), i = 1, \dots, n$ に対して、モデル $y_i = w_i + \varepsilon_i$ を想定し、最小2乗法によって推定した推定値

$$\hat{w}_i = y_i$$

は、データに含まれているノイズもモデルに取り込んでいる。

過適合 (Over-fitting)

過学習 (Over-learning)

将来の(未知の)データに対する予測能力の低下

最小2乗法を修正した推定方式が必要

平滑化スプライン

罰則付き最小2乗法 (Penalized least square method)

$$S_p(\mathbf{w}) = \sum_{i=1}^n (y_i - w_i)^2 + \lambda \int_a^b \{w''(t)\}^2 dt$$

残差2乗和

関数 w のデータへの
当てはまりを表す

Roughness penalty

関数 w の複雑さを表す

λ : 関数 w の「データへの当てはまり」と「複雑さ」
のバランスをコントロールするパラメータ
平滑化パラメータと呼ばれる

平滑化パラメータの効果

$$S_p(\mathbf{w}) = \sum_{i=1}^n (y_i - w_i)^2 + \lambda \int_a^b \{w''(t)\}^2 dt \quad \arg \min S_\lambda(\mathbf{w}) = \hat{\mathbf{w}}$$

(a) λ が小さい値の場合

$S_p(\mathbf{w})$ の最小化は主に第一項 $\sum_{i=1}^n (y_i - w_i)^2$ の最小化と等価

➡ 最小2乗推定量 $\hat{w}_i = y_i$

(b) λ が大きい値の場合

$S_p(\mathbf{w})$ の最小化は主に第二項 $\int_a^b \{w''(t)\}^2 dt$ の最小化と等価

➡ 線形回帰

2階微分: 自然3次スプライン

補題1

関数 w を節点 x_1, x_2, \dots, x_n を持つ自然3次スプラインとする。ただし、 $x_1 < \dots < x_n$ とする。また、 w の2階微分 $\tau_i = w''(x_i)$ で定義されるベクトルを $\tau = (\tau_2, \dots, \tau_{n-1})^T$ とする。ただし、 $\tau_1 = \tau_n = 0$ である。このとき、 $n \times (n-2)$ 行列 Q と $(n-2) \times (n-2)$ 対称行列 R

$$Q = \begin{pmatrix} \frac{1}{h_1} & 0 & 0 & \cdots & 0 \\ -\frac{1}{h_1} - \frac{1}{h_2} & \frac{1}{h_2} & 0 & \cdots & 0 \\ \frac{1}{h_2} & -\frac{1}{h_2} - \frac{1}{h_3} & \frac{1}{h_3} & \cdots & 0 \\ 0 & \frac{1}{h_3} & -\frac{1}{h_3} - \frac{1}{h_4} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \frac{1}{h_{n-2}} & -\frac{1}{h_{n-2}} - \frac{1}{h_{n-1}} \\ 0 & \cdots & \cdots & 0 & \frac{1}{h_{n-1}} \end{pmatrix}$$

2階微分: 自然3次スプライン

補題1 (続き)

$$R = \begin{pmatrix} \frac{h_1+h_2}{3} & \frac{h_2}{6} & 0 & 0 & \cdots & 0 \\ \frac{h_2}{6} & \frac{h_2+h_3}{3} & \frac{h_3}{6} & 0 & \cdots & 0 \\ 0 & \frac{h_3}{6} & \frac{h_3+h_4}{3} & \frac{h_4}{6} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \frac{h_{n-3}+h_{n-2}}{3} & \frac{h_{n-2}}{6} \\ 0 & 0 & \cdots & \cdots & \frac{h_{n-2}}{6} & \frac{h_{n-2}+h_{n-1}}{3} \end{pmatrix}$$

を用いて

$$Q^T \mathbf{w} = R \boldsymbol{\tau}$$

が成り立つ。ただし、 $h_i = x_{i+1} - x_i$ である。

Roughness Penalty の表現

定理1

補題1の状況下で

$$\int_a^b \{w''(t)\}^2 dt = \mathbf{w}^T \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T \mathbf{w}$$

補題1の証明

Proof. Consider a natural cubic spline $w(\cdot)$ on the interval $[x_j, x_{j+1}]$, $j = 1, \dots, n-1$, and define

$$w(x_j) = w_j, \quad w''(x_j) = \tau_j$$

and $h_j = x_{j+1} - x_j$. On the interval $[x_j, x_{j+1}]$, $w''(x)$ is linear, since $w(x)$ is a cubic polynomial. Therefore $w''(x)$ can be expressed as $w''(x) = D_j x + C_j$ ($x_j \leq x \leq x_{j+1}$), where D_j and C_j are constants satisfying

$$\begin{cases} \lim_{\delta \downarrow 0} w''(x_j + \delta) = D_j x_j + C_j = \tau_j, \\ \lim_{\delta \downarrow 0} w''(x_{j+1} - \delta) = D_j x_{j+1} + C_j = \tau_{j+1}. \end{cases}$$

Then D_j and C_j are given by

$$D_j = \frac{\tau_{j+1} - \tau_j}{h_j}, \quad C_j = \frac{x_{j+1}\tau_j - x_j\tau_{j+1}}{h_j},$$

respectively. On the other hand, D_j and C_j relate to d_j and c_j in (2.1) such as

$$d_j = \frac{D_j}{6}, \quad (2.6)$$

$$c_j = \frac{D_j x_j + C_j}{2}. \quad (2.7)$$

From the first equation of (2.2), it is immediately clear that

$$b_j = \frac{a_{j+1} - a_j}{h_j} - c_j h_j - d_j h_j^2. \quad (2.8)$$

From (2.6), (2.7), (2.8), $a_j = w_j$ and (2.1), the natural cubic spline $w(x)$ on $[x_j, x_{j+1}]$ can be rewritten as

$$w_{(j)}(x) = \frac{w_j(x_{j+1} - x) + w_{j+1}(x - x_j)}{h_j} - \frac{1}{6}(x_{j+1} - x)(x - x_j) \left\{ \left(1 + \frac{x_{j+1} - x}{h_j}\right) \tau_j + \left(1 + \frac{x - x_j}{h_j}\right) \tau_{j+1} \right\} \quad (2.9)$$

Then $w'_{(j)}$ is given by

$$w'_{(j)}(x) = \frac{w_{j+1} - w_j}{h_j} - \frac{(x_{j+1} - x) - (x - x_j)}{6} \left\{ \left(1 + \frac{x_{j+1} - x}{h_j}\right) \tau_j + \left(1 + \frac{x - x_j}{h_j}\right) \tau_{j+1} \right\} - \frac{(x_{j+1} - x)(x - x_j)}{6} \frac{\tau_{j+1} - \tau_j}{h_j}.$$

Hence we have

$$\lim_{\delta \downarrow 0} w'_{(j-1)}(x_j - \delta) = \frac{w_j - w_{j-1}}{h_{j-1}} + \frac{1}{6} h_{j-1} (\tau_{j-1} + 2\tau_j),$$

$$\lim_{\delta \downarrow 0} w'_{(j)}(x_j + \delta) = \frac{w_{j+1} - w_j}{h_j} + \frac{1}{6} h_j (2\tau_j + \tau_{j+1}).$$

From the continuous conditions, the first derivatives of $w_{(j-1)}$ and $w_{(j)}$ at x_j are satisfied $\lim_{\delta \downarrow 0} w'_{(j-1)}(x_j - \delta) = \lim_{\delta \downarrow 0} w'_{(j)}(x_j + \delta)$ and we have the following equation

$$\frac{w_{j+1} - w_j}{h_j} - \frac{w_j - w_{j-1}}{h_{j-1}} = \frac{1}{6} h_{j-1} \tau_{j-1} + \frac{1}{3} (h_{j-1} - h_j) \tau_j + \frac{1}{6} h_j \tau_{j+1}, \quad j = 2, \dots, n-1.$$

These equations implies $\mathbf{Q}^T \mathbf{w} = \mathbf{R} \boldsymbol{\tau}$. This completes the proof of the theorem. \square

定理1の証明

Proof. Since $w(\cdot)$ is a natural cubic spline, it is immediately clear that $w''(a) = w''(b) = 0$ and $w'''(x)$ is constant on each interval (x_j, x_{j+1}) and zero outside $[x_1, x_n]$. Following the formula of integration by parts, we obtain

$$\begin{aligned} \int_a^b \{w''(x)\}^2 dx &= - \int_a^b w'''(x)w'(x)dx \\ &= - \sum_{j=1}^{n-1} \lim_{\delta \downarrow 0} w'''(x_j + \delta) \int_{x_j}^{x_{j+1}} w'(x)dx. \end{aligned}$$

In the proof of Theorem 2.1, we have

$$\lim_{\delta \downarrow 0} w'''(x_j + \delta) = D_j = \frac{\tau_{j+1} - \tau_j}{h_j}.$$

Then the roughness penalty is

$$\int_a^b \{w''(x)\}^2 dx = \sum_{j=1}^{n-1} \frac{\tau_{j+1} - \tau_j}{h_j} (w_{j+1} - w_j). \quad (2.10)$$

Since $\tau_1 = \tau_n = 0$, it follows by rearranging (2.10) that

$$\begin{aligned} \int_a^b \{w''(x)\}^2 dx &= \sum_{j=2}^{n-1} \tau_j \left(\frac{w_{j+1} - w_j}{h_j} - \frac{w_j - w_{j-1}}{h_{j-1}} \right) \\ &= \boldsymbol{\tau}^T \mathbf{Q}^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T \mathbf{w}, \end{aligned}$$

making use of the property $\boldsymbol{\tau} = \mathbf{R}^{-1} \mathbf{Q}^T \mathbf{w}$. This completes the proof of Theorem 2.2. \square

平滑化スプライン

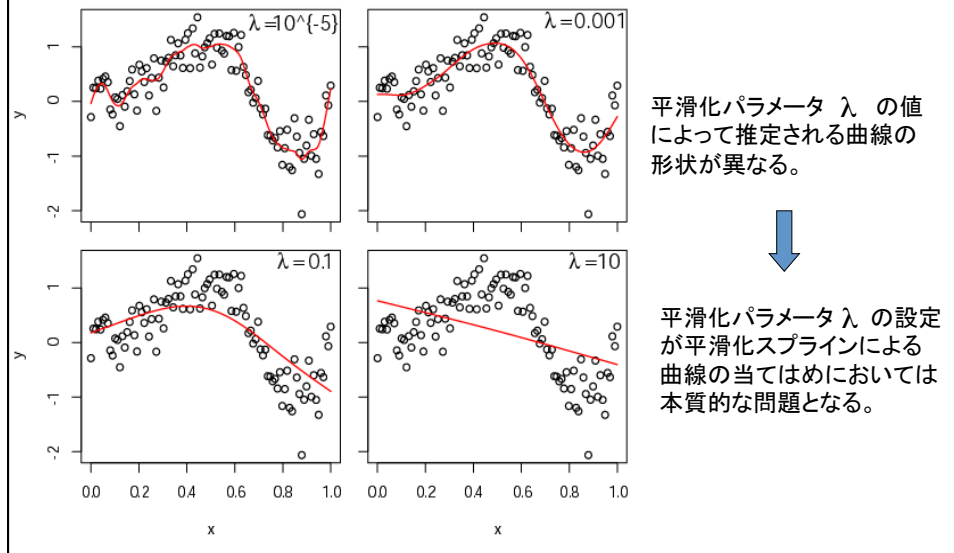
$$\begin{aligned} S_p(\mathbf{w}) &= \sum_{i=1}^n (y_i - w_i)^2 + \lambda \int_a^b \{w''(t)\}^2 dt \\ \arg \min S_p(\mathbf{w}) &= \hat{\mathbf{w}} \end{aligned}$$

$$\begin{aligned} S_\lambda(\mathbf{w}) &= (\mathbf{y} - \mathbf{w})^T (\mathbf{y} - \mathbf{w}) + \lambda \mathbf{w}^T \mathbf{K} \mathbf{w} \\ &= \mathbf{y}^T \mathbf{y} - 2\mathbf{w}^T \mathbf{y} + \mathbf{w}^T (\mathbf{I} + \lambda \mathbf{K}) \mathbf{w} \end{aligned}$$

$$\frac{\partial S_\lambda(\mathbf{w})}{\partial \mathbf{w}} = -2\mathbf{y} + 2(\mathbf{I} + \lambda \mathbf{K})\mathbf{w} \Rightarrow \frac{\partial S_\lambda(\mathbf{w})}{\partial \mathbf{w}} = 0 \quad \text{よリ}$$

$$\hat{\mathbf{w}} = (\mathbf{I} + \lambda \mathbf{K})^{-1} \mathbf{y}$$

例：平滑化スプライン



平滑化パラメータの選択 (1)

交差検証法 (Cross-validation)

今 i 番目のデータの組 (x_i, y_i) を除いた $n-1$ 組のデータによって推定された自然3次スプラインを $\hat{w}^{(-i)}$ と表す。
このとき、交差検証法基準は

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{w}^{(-i)}(x_i)\}^2$$

と表され、CVが最小となる平滑化パラメータ λ の値を最適値として選択する。

平滑化パラメータの選択 (2)

$$S_p(\mathbf{w}) = \sum_{i=1}^n (y_i - w_i)^2 + \lambda \int_a^b \{w''(t)\}^2 dt$$

$$\operatorname{argmin} S_p(\mathbf{w}) = \hat{\mathbf{w}} = (\mathbf{I} + \lambda \mathbf{K})^{-1} \mathbf{y} = \mathbf{H}_\lambda \mathbf{y}$$

定理2

$$\operatorname{CV}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{w}(x_i)}{1 - (\mathbf{H}_\lambda)_{ii}} \right\}^2$$

平滑化パラメータの選択 (3)

一般化交差検証法 (Generalized Cross-validation)

Craven and Wahba (1979)

$$\operatorname{GCV}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{w}(x_i)}{1 - \operatorname{tr}(\mathbf{H}_\lambda) / n} \right\}^2$$

$$\operatorname{GCV}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left\{ \left(\frac{1 - (\mathbf{H}_\lambda)_{ii}}{1 - \operatorname{tr}(\mathbf{H}_\lambda) / n} \right) \{y_i - \hat{w}^{(-i)}(x_i)\} \right\}^2$$

線形重回帰モデルを非線形へ

データ: $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$

線形重回帰モデル:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

$w_1(x_{i1})$ $w_2(x_{i2})$ $w_p(x_{ip})$

スプライン加法モデル

$$y_i = \beta_0 + w_1(x_{i1}) + w_2(x_{i2}) + \dots + w_p(x_{ip}) + \varepsilon_i$$

Roughness Penalty: スプライン加法モデル

スプライン加法モデル

$$y_i = \beta_0 + w_1(x_{i1}) + w_2(x_{i2}) + \dots + w_p(x_{ip}) + \varepsilon_i$$

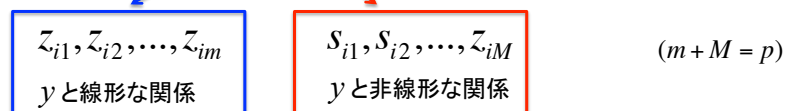
罰則付き最小2乗法

$$S_A(\mathbf{w}_1, \dots, \mathbf{w}_p) = \sum_{i=1}^n \{y_i - m(x_i)\}^2 + \sum_{j=1}^p \lambda_j \int_{a_j}^{b_j} \{w_j''(t)\}^2 dt$$

変数の数 (p) が大きいとパラメータ数が激増。
 モデルの解釈も困難に。
 平滑化パラメータが p 個。 R^p での最適化が必要。

セミパラメトリックモデル

データ: $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$



$$y_i = \beta_0 + \beta_1 z_{i1} + \dots + \beta_m z_{im} + w_1(s_{i1}) + \dots + w_M(s_{iM}) + \varepsilon_i$$

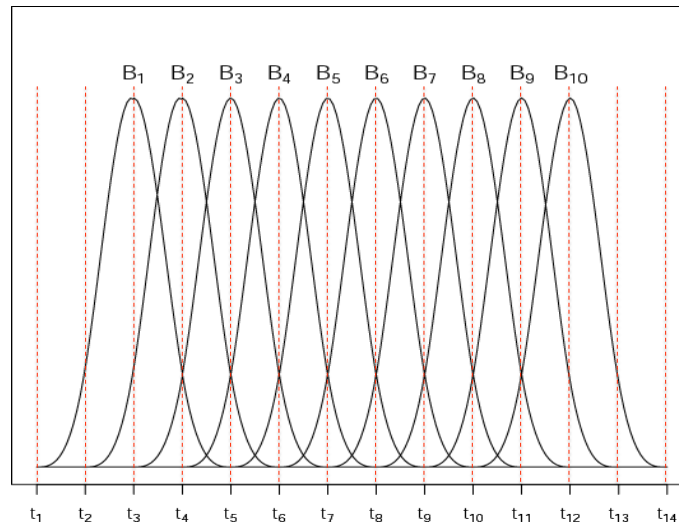
線形で十分な変数は線形の方が良い。
どの変数を非線形にするか？

B-スプライン (1)

B-スプライン $B_j(t; r)$ は以下の性質を満たす。
ただし、 t_i ($t_1 < \dots < t_{m+r+1}$) は節点。

- (I) 区間 (t_j, t_{j+r+1}) において $B_j(t; r)$ は正。それ以外では 0
- (II) $B_j(t; r)$ は $r+1$ 個の r 次多項式からなる
- (III) $r-1$ 階までの導関数が連続

B-スプライン (2)



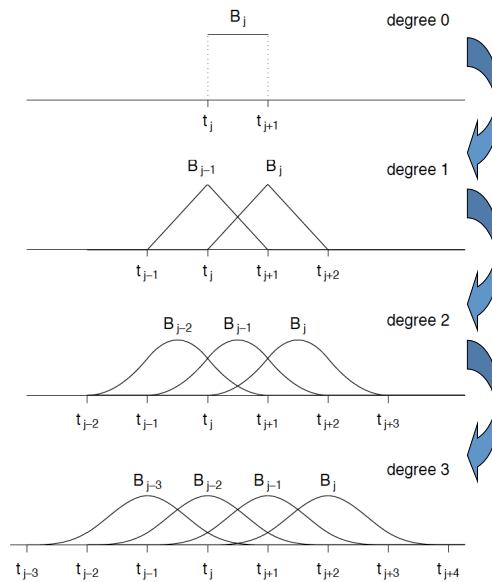
B-スプライン (3)

再帰計算により高次の B-スプラインが求まる (De Boor のアルゴリズム (1978))

$$B_j(x; 0) = \begin{cases} 1, & t_j \leq x < t_{j+1} \\ 0, & \text{otherwise} \end{cases} \quad \text{0 次 B-スプライン}$$

$$B_j(x; r) = \frac{x - t_j}{t_{j+r} - t_j} B_j(x; r-1) + \frac{t_{j+r+1} - x}{t_{j+r+1} - t_{j+1}} B_{j+1}(x; r-1)$$

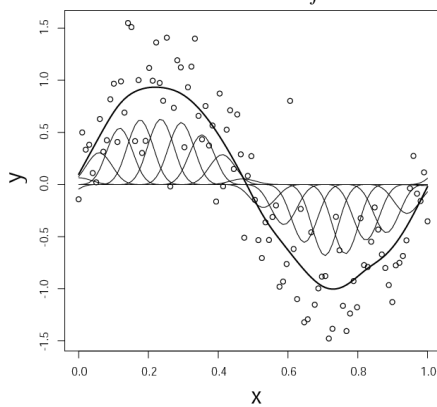
再帰計算



B-スプラインによる曲線の当てはめ

B-スプラインによる曲線の構成

$$w(x) = \sum_{j=1}^m \gamma_j B_j(x) = \boldsymbol{\gamma}^T \mathbf{b}(x)$$



$$\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)^T$$

$$\mathbf{b}(x) = (B_1(x), \dots, B_m(x))^T$$

スプラインと比較すると

	スプライン	B-スプライン
パラメータ数	n	m

B-スプライン回帰モデル

データ $(x_i, y_i), i = 1, \dots, n$

モデル $y_i = \boldsymbol{\gamma}^T \mathbf{b}(x_i) + \varepsilon_i$

ノイズ $\varepsilon_1, \dots, \varepsilon_n$ が互いに独立に平均0, 分散 σ^2 の正規分布に従うとすると

$$f(\mathbf{y} | \mathbf{x}, \boldsymbol{\gamma}, \sigma^2) = \prod_{i=1}^n f(y_i | x_i, \boldsymbol{\gamma}, \sigma^2)$$

$$f(y_i | x_i, \boldsymbol{\gamma}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\{y_i - \boldsymbol{\gamma}^T \mathbf{b}(x_i)\}^2}{2\sigma^2}\right]$$

B-スプライン回帰モデルの推定

罰則付き最小2乗法

$$S_B(\boldsymbol{\gamma}) = \sum_{i=1}^n \{y_i - \boldsymbol{\gamma}^T \mathbf{b}(x_i)\}^2 + \lambda \sum_{k=2}^m (\gamma_k - 2\gamma_{k-1} + \gamma_{k-2})^2$$

$$\hat{\boldsymbol{\gamma}} = \operatorname{argmin}\{S_B(\boldsymbol{\gamma})\}$$

$$\int_a^b \left\{ \sum_{j=1}^m \gamma_j B_j''(t; 3) \right\}^2 dt \Leftrightarrow \sum_{k=2}^m (\gamma_k - 2\gamma_{k-1} + \gamma_{k-2})^2$$

等価自由度

$$\hat{\boldsymbol{y}} = \mathbf{B}(\mathbf{B}^T \mathbf{B} + \lambda \mathbf{K})^{-1} \mathbf{B}^T \mathbf{y} = \mathbf{H}_\lambda \mathbf{y}$$

$df(\lambda) = \text{tr}(\mathbf{H}_\lambda)$ は等価自由度 (Equivalent degree of freedom) と呼ばれる

パラメトリック回帰モデル $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\theta} \in \mathbb{R}^p$, $\boldsymbol{\varepsilon} \sim N_p(\mathbf{0}, \sigma^2 \mathbf{I})$ だと

$\hat{\boldsymbol{\theta}}^{ML} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ より $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ なので

$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ となる (λ には依存しない)

$$\text{tr}\{\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T\} = p \quad = \text{パラメータ数}$$

$$\text{AIC} = -2 \log f(\mathbf{y} | \mathbf{x}, \hat{\boldsymbol{\theta}}^{ML}, \hat{\sigma}^2) + 2(p+1)$$

平滑化パラメータの選択

データ $(x_i, y_i), i = 1, \dots, n$

モデル $y_i = \boldsymbol{\gamma}^T \mathbf{b}(x_i) + \varepsilon_i$, $\varepsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$

修正AIC (Modified AIC) Eilers and Marx (1996)

$$\begin{aligned} \text{AIC}_m(\lambda) &= -2 \sum_{i=1}^n \log f(y_i | x_i, \hat{\boldsymbol{\gamma}}, \hat{\sigma}^2) + 2\{\text{tr}(\mathbf{H}_\lambda) + 1\} \\ &= n \log(2\pi \hat{\sigma}^2) + n + 2\{\text{tr}(\mathbf{H}_\lambda) + 1\} \end{aligned}$$

ただし

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{\boldsymbol{\gamma}}^T \mathbf{b}(x_i)\}^2$$

平滑化パラメータの選択

パラメトリック回帰モデル

$$y = X\theta + \varepsilon, \quad \theta \in \mathbb{R}^p, \quad \varepsilon \sim N_p(\mathbf{0}, \sigma^2 I)$$

データを生成した真の分布が正規分布でモデルに含まれるとき

$$AIC_C = -2 \log f(y | x, \hat{\theta}, \sigma^2) + \frac{2n(p+1)}{n-p-2} \quad \text{Sugiura (1978)}$$

修正 AIC Hurvich, Simonoff and Tsai (1998)

$$AIC_C(\lambda) = n \log(2\pi\hat{\sigma}^2) + n + \frac{2n \{ \text{tr}(H_\lambda) + 1 \}}{n - \text{tr}(H_\lambda) - 2}$$

様々な基準の性能比較

	$SPIC_N$	CV	GCV	$ABIC_N$	AIC_m^*	AIC_C
$\sigma/R_y = 0.05$						
$\varepsilon = 1.0$						
MEAN $\times 10^6$	9.102	13.66	15.78	30.70	14.01	18.80
SD $\times 10^5$	1.263	1.856	1.847	1.107	1.689	2.064
ASE $\times 10^3$	3.523	3.577	3.581	3.721	3.559	3.614
PASE $\times 10^2$	3.819	3.826	3.824	3.838	3.821	3.826
$\varepsilon = 0.9$						
MEAN $\times 10^5$	1.095	1.558	1.964	3.807	1.746	2.439
SD $\times 10^5$	1.268	1.838	2.175	1.415	1.929	2.730
ASE $\times 10^3$	4.085	4.136	4.138	4.297	4.117	4.184
PASE $\times 10^2$	4.498	4.503	4.503	4.523	4.500	4.508