最適化理論

http://www.cs.miyazaki-u.ac.jp/~date/lectures/optimization/

伊達 章

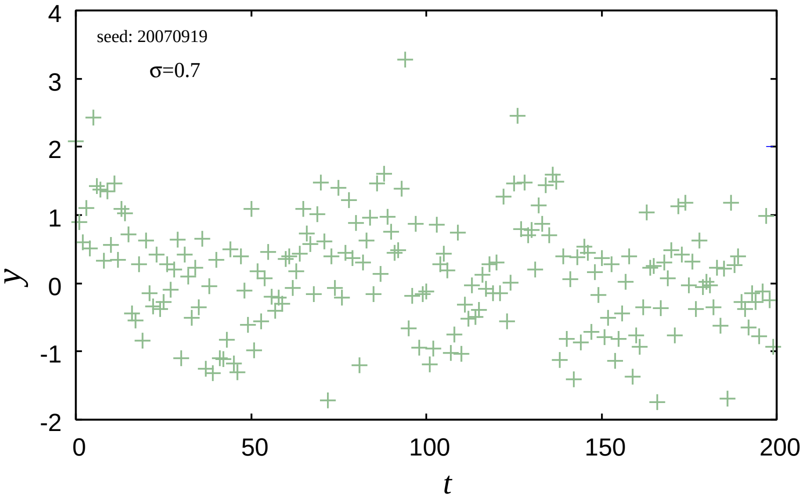
宮崎大学 工学部 情報システム工学科

2016 年 6 月 17 日

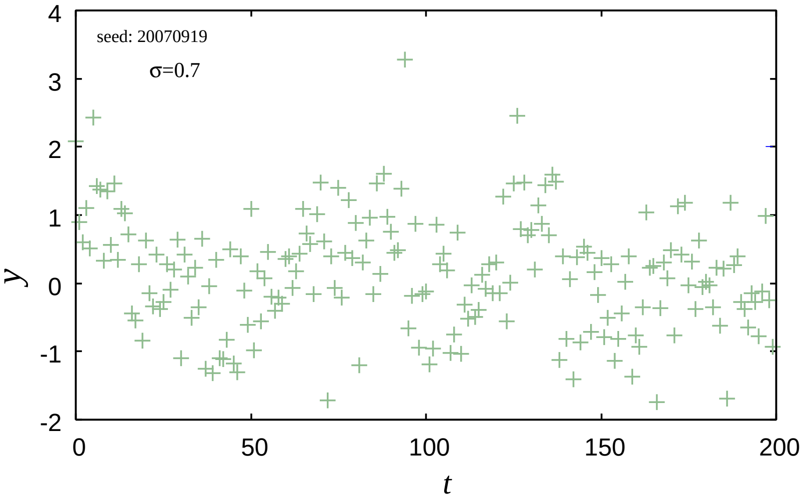
講義のスケジュール（案）

1. 講義の概要
2. 数学的準備：曲線と曲面
3. 数学的準備：1 次形式と 2 次形式
4. 数学的準備：2 次形式の標準形
5. 関数の極値： 関数の勾配と等高線，関数の極値
6. 関数の極値：ラグランジュの未定乗数法
7. 関数の最適化：勾配法・ニュートン法
8. 関数の最適化：共役勾配法
8. 関数の最適化：共役勾配法
9. 統計的最適化：正規分布，最尤推定
10. 動的計画法 （その 1）
11. 動的計画法 （その 2）
12. 最小二乗法：連立一次方程式，特異値分解と一般化逆行列
13. 最小二乗法（その 2）
14. まとめ
15. 定期試験，解説

観測データ（時系列）

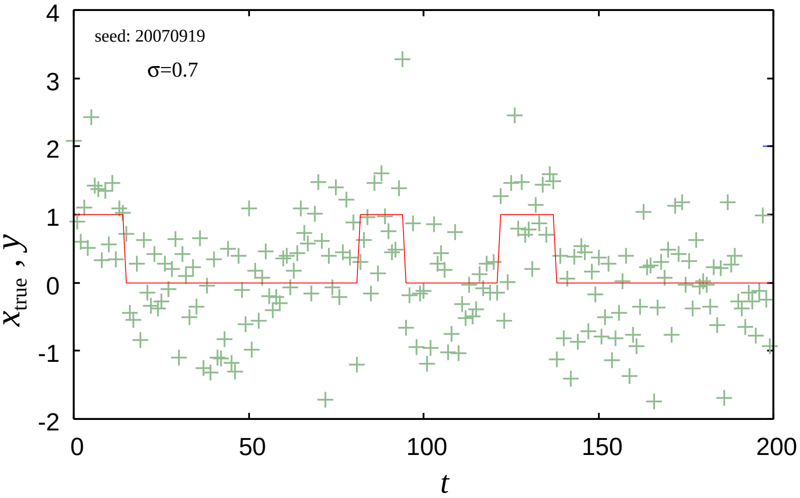


観測データ（時系列）



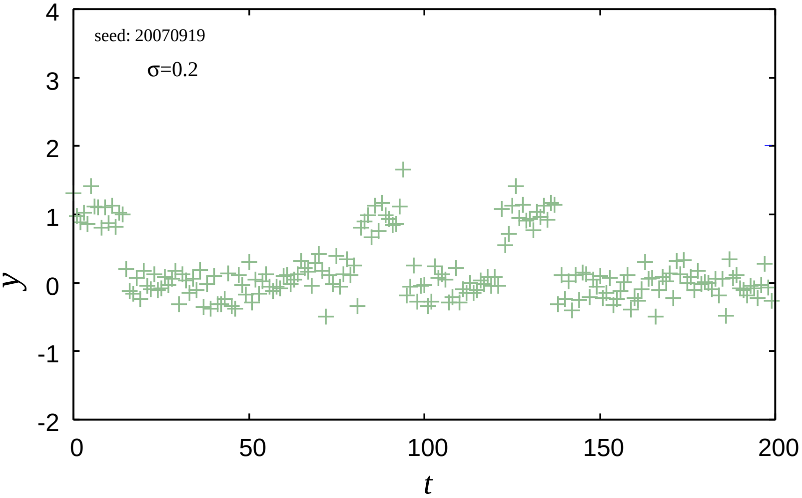
もとの信号は 0 か 1．復元したい！

観測データ（時系列）



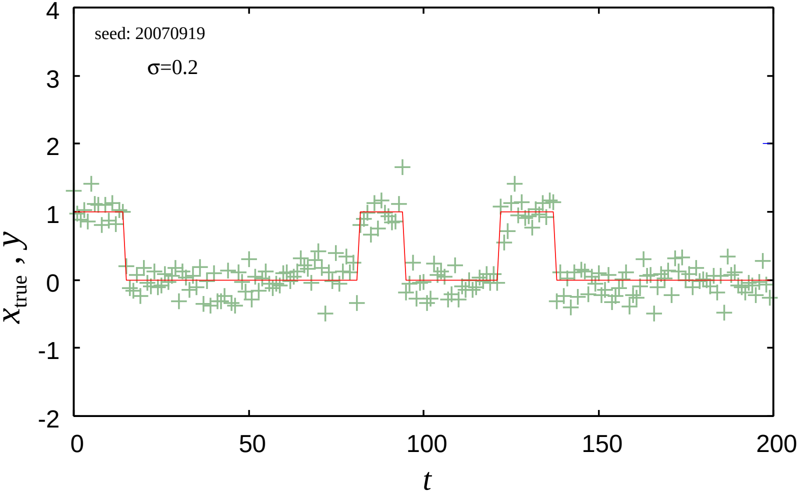
もとの信号は 0 か 1． 復元したい！

観測データ（時系列）



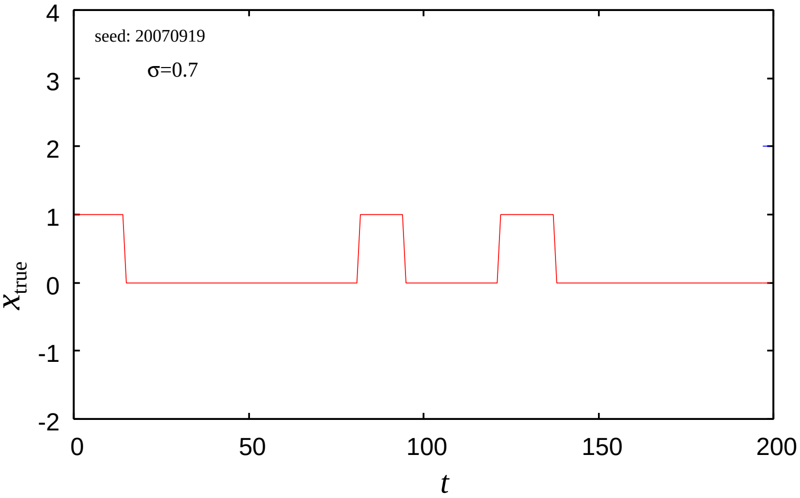
こちらのほうが復元しやすい

観測データ（時系列）



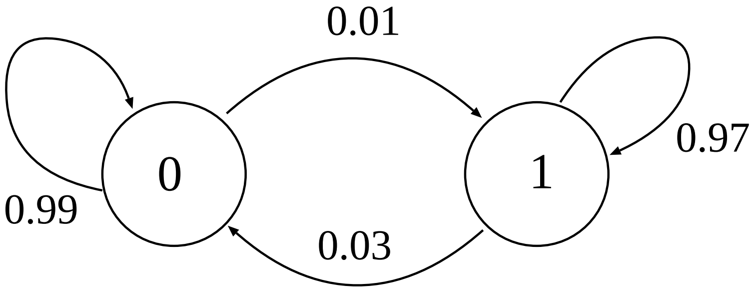
こちらのほうが復元しやすい

もとの信号 $x_{\mathbf{true}}$

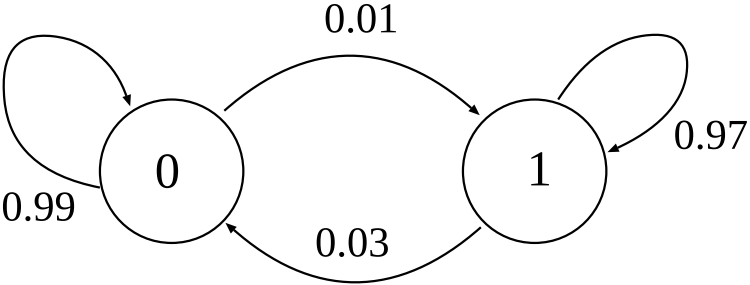


マルコフ的情報源

マルコフ的情報源

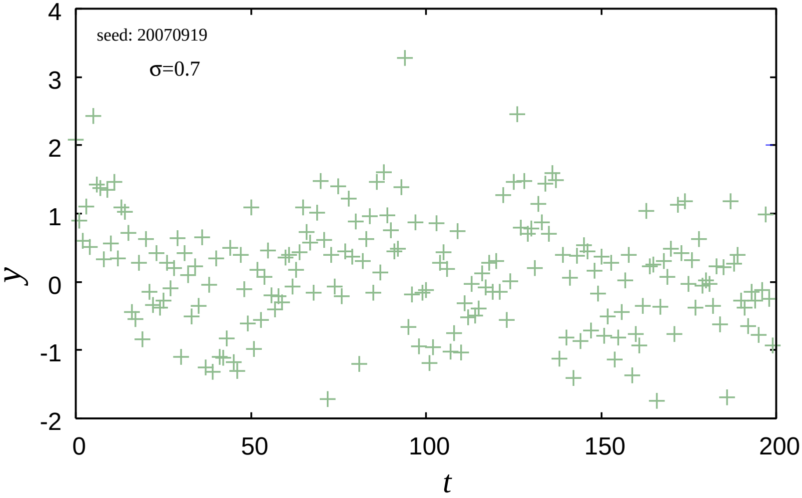


マルコフ的情報源



000000001111111111100000000

観測データ（時系列）



多段決定問題

多変数関数 $f(\boldsymbol{x})$ の最大化

$$J = f(x_1, x_2, \cdots, x_n) \to \max$$

多段決定問題

多変数関数 $f(\boldsymbol{x})$ の最大化

$$J = f(x_1, x_2, \cdots, x_n) \to \max$$

$n = 10, x_i \in \{0, 1\}$ の場合

	$\boldsymbol{x}$	J	
0	0000000000	$f(\boldsymbol{x}_0)$	
1	0000000001	$f(\boldsymbol{x}_1)$	
2	0000000010	$f(\boldsymbol{x}_2)$	
	$\vdots$		
k	0011101011	$f(\boldsymbol{x}_k)$	$\leftarrow$ 最大
	$\vdots$		
1023	1111111111	$f(\boldsymbol{x}_{2^n-1})$	

$$\max_i f(\boldsymbol{x}_i) = f(\boldsymbol{x}_k), \qquad \underset{i}{\operatorname{argmax}} f(\boldsymbol{x}_i) = k$$

多段決定問題

多変数関数 $f(\boldsymbol{x})$ の最大化

$$J = f(x_1, x_2, \cdots, x_n) \to \max$$

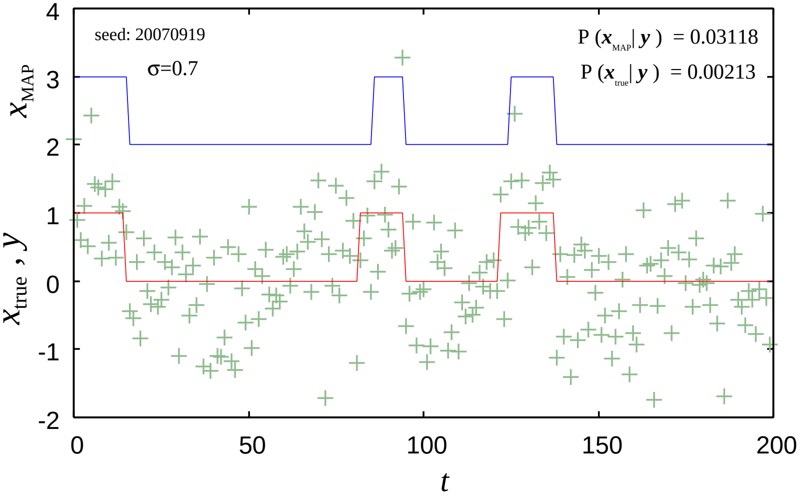
$n = 100, x_i \in \{0, 1\}$ の場合. $2^{100} = (2^{10})^{10} \approx (10^3)^{10} = 10^{30}$

	$\boldsymbol{x}$	J	
0	0000000000	$f(\boldsymbol{x}_0)$	
1	0000000001	$f(\boldsymbol{x}_1)$	
2	0000000010	$f(\boldsymbol{x}_2)$	
3	0000000011	$f(\boldsymbol{x}_3)$	
	$\vdots$		
k	0011101011	$f(\boldsymbol{x}_k)$	$\leftarrow$ 最大
	$\vdots$		
$\approx 10^{30}$	111 $\cdots$ 111	$f(\boldsymbol{x}_{2^n-1})$	

課題

- 例題： $n = 200,\ \boldsymbol{y} = (y_1, \cdots, y_n)$
- 目的：$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmax}}\ f(\boldsymbol{x}), x_i \in \{0, 1\}$ の計算
- まもとには計算できない.
- 動的計画法を使い，この問題を解決する.
- 仕組み理解し，コードを書き，問題を解く.
- 口頭試問をクリア
  （1 人 2 回まで挑戦可．最大 10 分/回）
  $\Rightarrow$ 成績に $10 \sim 20$ 点を加算

事後確率最大にする値 x_{MAP}



多段決定問題

多変数関数の最大化

$$J = p(\boldsymbol{x}|\boldsymbol{y}) \to \max$$

$n = 200, x_i \in \{0, 1\}$ の場合． $2^{200} = (2^{10})^{20} \approx (10^3)^{20} = 10^{60}$

	$\boldsymbol{x}$	J	
0	0000000000	$f(\boldsymbol{x}_0)$	
1	0000000001	$f(\boldsymbol{x}_1)$	
2	0000000010	$f(\boldsymbol{x}_2)$	
3	0000000011	$f(\boldsymbol{x}_3)$	
	$\vdots$		
k	0011101011	$f(\boldsymbol{x}_k)$	$\leftarrow$ 最大
	$\vdots$		
$\approx 10^{60}$	111 $\cdots$ 111	$f(\boldsymbol{x}_{2^n-1})$	

課題

- 例題： $n = 200,\ \boldsymbol{y} = (y_1, \cdots, y_n)$
- 目的：$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmax}}\ f(\boldsymbol{x}), x_i \in \{0, 1\}$ の計算
- まもとには計算できない.
- 動的計画法を使い，この問題を解決する.
- 仕組み理解し，コードを書き，問題を解く.
- 口頭試問をクリア
  （1 人 2 回まで挑戦可．最大 10 分/回）
  $\Rightarrow$ 成績に $10 \sim 20$ 点を加算
- 締切：定期試験の一週間前（7/23 木）まで

基本知識（確率・統計の復習）

- 平均 μ，分散 σ^2，標準偏差 σ
- 確率分布： 一様分布，正規分布
- 擬似乱数の生成
- 最尤推定
- 同時確率，条件付き確率
- マルコフ的情報源
- ベイズの公式，事前確率・事後確率
- 事後確率最大化

- 動的計画法（第８章）

確率，条件付き確率

	B_1 （風邪）	B_2 （風邪なし）	$p(A_i)$
A_1 （熱あり）	0.55	0.05	0.60
A_2 （熱なし）	0.10	0.30	0.40
$p(B_j)$	0.65	0.35	

例

同時確率 $p(A_1, B_1) = 0.55$

周辺確率 $p(A_1) = \sum_i p(A_1, B_i) = p(A_1) = 0.6$

条件付き確率

熱の有無を知る ⇒ 風邪であるかどうか検討がつく：

$$p(B_1|A_1) = \frac{p(B_1)p(A_1|B_1)}{p(A_1)} = \frac{p(A_1, B_1)}{p(A_1)} = \frac{0.55}{0.6} \approx 0.92$$

ベイズの公式

- ベイズの公式

  熱があったとしよう．　その時，風邪のあるなしの確率

$$p(B_1|A_1) = \frac{p(B_1)p(A_1|B_1)}{p(A_1)} = \frac{p(A_1, B_1)}{p(A_1)} = \frac{0.55}{0.6} \approx 0.92$$

$$p(B_2|A_1) = \frac{p(B_2)p(A_1|B_2)}{p(A_1)} = \frac{p(A_1, B_2)}{p(A_1)} = \frac{0.05}{0.6} \approx 0.08$$

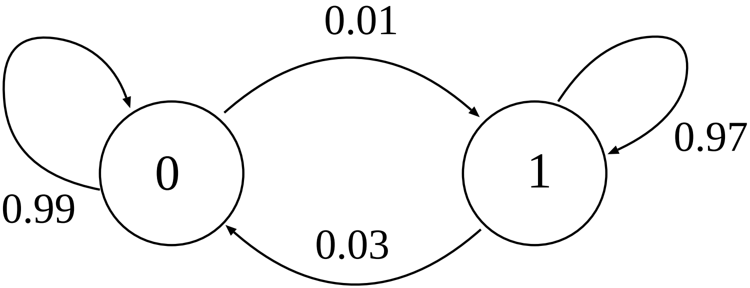
- 事後確率最大化（ベイズ推定）　$\underset{i}{\operatorname{argmax}}\ p(B_i|A_1) = 1$

  風邪であることの方が確率が大 $\Rightarrow$ 風邪であると推定

  入力（観測値）：熱のあるなし

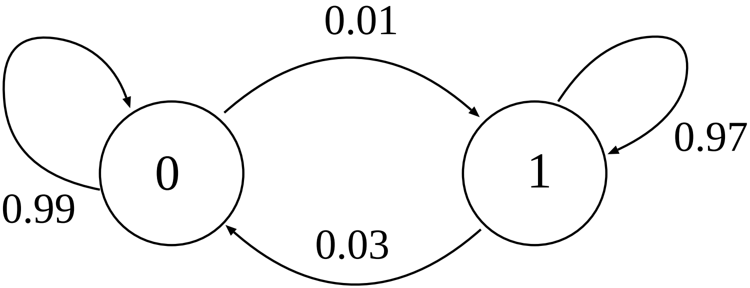
  $\Rightarrow$ 出力（推定値）　風邪かどうか

マルコフ的情報源



000000001111111111100000000

マルコフ的情報源



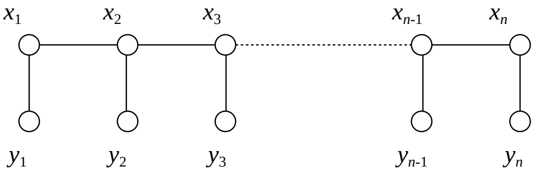
000000001111111111100000000

$$p(\boldsymbol{x}) = 0.5 \times 0.99 \times 0.99 \times 0.99 \times 0.99 \cdots$$

$$p(\boldsymbol{x}) = p(x_1) \times p(x_2|x_1) \times p(x_3|x_2) \times p(x_4|x_3) \cdots$$

事後確率 $p(\boldsymbol{x}|\boldsymbol{y})$ の最大化

横棒は条件付確率，縦棒は尤度に対応



$$\boldsymbol{x}^* = \underset{x_1, x_2, \cdots, x_n}{\operatorname{argmax}} \; p(x_1, x_2, \cdots, x_n, \; y_1, y_2, \cdots, y_n)$$

$$\boldsymbol{x}^* = \underset{x_1, x_2, \cdots, x_n}{\operatorname{argmax}} \; p(x_1) \prod_{i=2}^{n} p(x_i | x_{i-1}) \prod_{i=1}^{n} p(y_i | x_i)$$

使って良い知識

データ（観測値）： $\boldsymbol{y} = (y_1, \cdots, y_n)$

モデル： $p(\boldsymbol{x}), p(\boldsymbol{y}|\boldsymbol{x})$

$$
\begin{aligned}
p(x_1) &= \begin{cases} 0.5 & \text{if} \quad x_1 = 0 \\ 0.5 & \text{if} \quad x_1 = 1 \end{cases} \\
p(x_{i+1}|x_i) &= \begin{cases} 0.99 & \text{if} \quad x_i = 0,\ x_{i+1} = 0 \\ 0.01 & \text{if} \quad x_i = 0,\ x_{i+1} = 1 \\ 0.97 & \text{if} \quad x_i = 1,\ x_{i+1} = 1 \\ 0.03 & \text{if} \quad x_i = 1,\ x_{i+1} = 0 \end{cases} \\
p(y_i|x_i) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(y_i - x_i)^2}{2\sigma^2} \right\}
\end{aligned}
$$

以降，この課題では，特に指定しない限り $\sigma = 0.7$.

事後確率最大化

$$\boldsymbol{x}^* = \operatorname*{argmax}_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{y})$$

$p(\boldsymbol{x}|\boldsymbol{y})$ は知らない. $p(\boldsymbol{x}), p(\boldsymbol{y}|\boldsymbol{x})$ は与えられている.

事後確率最大化

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmax}} \; p(\boldsymbol{x}|\boldsymbol{y})$$

$p(\boldsymbol{x}|\boldsymbol{y})$ は知らない．$p(\boldsymbol{x}), p(\boldsymbol{y}|\boldsymbol{x})$ は与えられている．

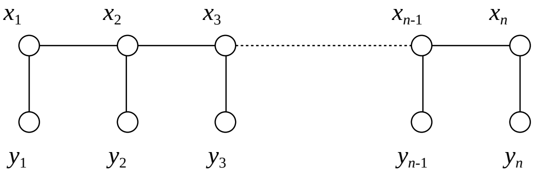
$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{y})} \quad \text{（ベイズの公式）}$$

$$= \frac{p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})}{p(\boldsymbol{y})}$$

$\boldsymbol{y}$ は観測値 $\Rightarrow p(\boldsymbol{y})$ は定数

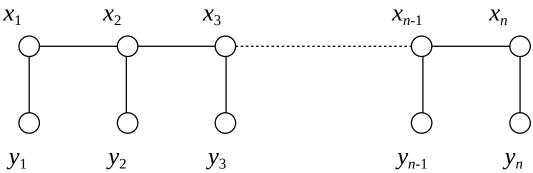
$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmax}} \; p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$$

事後確率最大化



$$
\begin{aligned}
\boldsymbol{x}^* &= \operatorname*{argmax}_{\boldsymbol{x}} \; p(\boldsymbol{x}) p(\boldsymbol{y}|\boldsymbol{x}) \\
&= \operatorname*{argmax}_{\boldsymbol{x}} \; p(x_1) \prod_{i=2}^{n} p(x_i|x_{i-1}) \prod_{i=1}^{n} p(y_i|x_i) \\
&= \operatorname*{argmax}_{\boldsymbol{x}} \; \log p(x_1) + \sum_{i=2}^{n} \log p(x_i|x_{i-1}) + \sum_{i=1}^{n} \log p(y_i|x_i)
\end{aligned}
$$

事後確率最大化

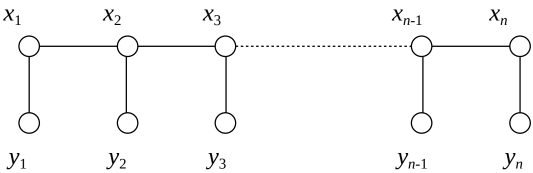


$$
\begin{aligned}
J &= \log p(x_1) + \sum_{i=2}^{n} \log p(x_i|x_{i-1}) + \sum_{i=1}^{n} \log p(y_i|x_i) \\
&= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1} \left\{ \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1}) \right\}
\end{aligned}
$$

y_i は定数 ⇒ 教科書 p.216 (8.2) 式と同じ形

$$
= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)
$$

事後確率最大化



$$
\begin{aligned}
J &= \log p(x_1) + \sum_{i=2}^{n} \log p(x_i|x_{i-1}) + \sum_{i=1}^{n} \log p(y_i|x_i) \\
&= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1} \left\{ \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1}) \right\}
\end{aligned}
$$

y_i は定数 ⇒ 教科書 p.216 (8.2) 式と同じ形

$$
= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)
$$

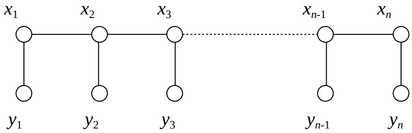
- $f_1(x_1) = \log p(x_1) + \log p(y_1|x_1)$
- $h_i(x_i, x_{i+1}) = \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1})$

基本知識（確率・統計の復習）

- 平均 μ，分散 σ^2，標準偏差 σ
- 確率分布： 一様分布，正規分布
- 擬似乱数の生成
- 最尤推定
- 同時確率，条件付き確率
- マルコフ的情報源
- ベイズの公式，事前確率・事後確率
- 事後確率最大化

- 動的計画法（第８章）

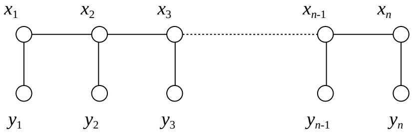
動的計画法



$$
\begin{aligned}
J &= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1}\Big\{\log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1})\Big\} \\
&= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)
\end{aligned}
$$

- x_1 に着目 $\Rightarrow f_1(x_1), h_1(x_1, x_2)$ にしか関係しない

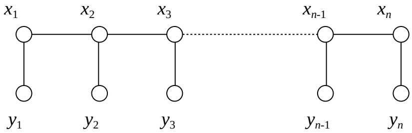
動的計画法



$$
\begin{aligned}
J &= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1} \Big\{ \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1}) \Big\} \\
&= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)
\end{aligned}
$$

- x_1 に着目 $\Rightarrow f_1(x_1), h_1(x_1, x_2)$ にしか関係しない
- $f_1(x_1) + h_1(x_1, x_2)$ を最大にするよう x_1 を選ぶ

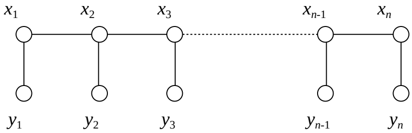
動的計画法



$$
\begin{aligned}
J &= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1} \Big\{ \log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1}) \Big\} \\
&= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)
\end{aligned}
$$

- x_1 に着目 $\Rightarrow f_1(x_1), h_1(x_1, x_2)$ にしか関係しない
- $f_1(x_1) + h_1(x_1, x_2)$ を最大にするよう x_1 を選ぶ
- ↑ x_2 の値がわかっていなければ選べない

動的計画法



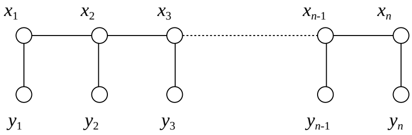
$$
\begin{aligned}
J &= \log p(x_1) + \log p(y_1|x_1) + \sum_{i=1}^{n-1}\Big\{\log p(x_{i+1}|x_i) + \log p(y_{i+1}|x_{i+1})\Big\} \\
&= f_1(x_1) + h_1(x_1, x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)
\end{aligned}
$$

- x_1 に着目 $\Rightarrow f_1(x_1), h_1(x_1, x_2)$ にしか関係しない
- $f_1(x_1) + h_1(x_1, x_2)$ を最大にするよう x_1 を選ぶ
- ↑ x_2 の値がわかっていないければ選べない
- そこで，x_2 の可能なすべての値（$0, 1$）に対して，以下を計算
- $f_2(x_2) = \max_{x_1}\{f_1(x_1) + h_1(x_1, x_2)\}$

  $\hat{x}_1(x_2) = \operatorname*{argmax}_{x_1}\{f_1(x_1) + h_1(x_1, x_2)\}$

  $J = f_2(x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)$

動的計画法

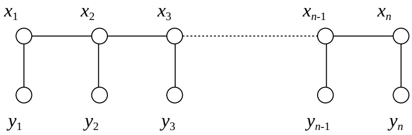


$$J = f_2(x_2) + h_2(x_2, x_3) + \cdots + h_{n-1}(x_{n-1}, x_n)$$

- 変数の数が 1 つ減った．これを続けていけばよい．
- x_2 に着目 $\Rightarrow f_2(x_2), h_2(x_2, x_3)$ にしか関係しない
- $f_2(x_2) + h_2(x_2, x_3)$ を最大にするよう x_2 を選ぶ
- $\uparrow$ x_3 の値がわかっていないければ選べない
- そこで，x_3 の可能なすべての値（$0, 1$）に対して，以下を計算
- $f_3(x_3) = \max_{x_2}\{f_2(x_2) + h_2(x_2, x_3)\}$

  $\hat{x}_2(x_3) = \operatorname*{argmax}_{x_2}\{f_2(x_2) + h_2(x_2, x_3)\}$

動的計画法

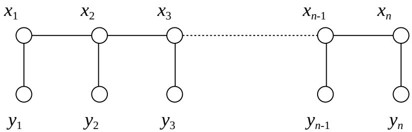


$$J \quad = \quad f_{n-1}(x_{n-1}) + h_{n-1}(x_{n-1}, x_n)$$

- x_n の可能なすべての値（$0, 1$）に対して，以下を計算
- $f_n(x_n) = \max_{x_{n-1}} \{f_{n-1}(x_{n-1}) + h_{n-1}(x_{n-1}, x_n)\}$

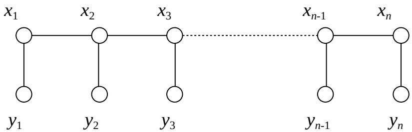
  $\hat{x}_{n-1}(x_n) = \operatorname*{argmax}_{x_{n-1}} \{f_{n-1}(x_{n-1}) + h_{n-1}(x_{n-1}, x_n)\}$
- $x_n^* = \operatorname*{argmax}_{x_n} f_n(x_n)$

動的計画法



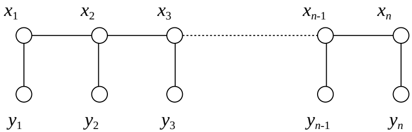
- $x_n^* = \underset{x_n}{\operatorname{argmax}}\, f_n(x_n)$

動的計画法



- $x_n^* = \operatorname*{argmax}_{x_n} f_n(x_n)$
- $x_{n-1}^* = \hat{x}_{n-1}(x_n^*)$

動的計画法



- $x_n^* = \underset{x_n}{\operatorname{argmax}}\, f_n(x_n)$
- $x_{n-1}^* = \hat{x}_{n-1}(x_n^*)$
- $x_{n-2}^* = \hat{x}_{n-2}(x_{n-1}^*)$
- $\cdots \to x_1^* = \hat{x}_1(x_2^*)$

データ構造

- $i = 1, \cdots, 200, \quad a = 0, 1$
- $x_i \cdots$ `int x[i]` 真の値
- $y_i \cdots$ `double y[i]` 観測データ
- $\hat{x}_i \cdots$ `int xhat[i]` 推定値
- $f_i(x_i) \cdots$ `double f[i][a]`
  i 番目の変数の値が $x_i = a$ のとき，そこまでに至る最適経路の尤もらしさ．
- $\hat{x}_i(x_{i+1}) \cdots$ `int xhat[i][a]`
  $x_{i+1} = a$ のとき，x_i が取るべき値

終