隠れマルコフモデル

正田 備也

masada@rikkyo.ac.jp

Contents

マルコフモデル

- $\boldsymbol{z} = (z_1, \ldots, z_T)$ という確率変数の列がある
- z_t は、状態の集合 $\mathcal{S} = \{s_1, \ldots, s_K\}$ の要素を値としてとる
  - 例えば、$\mathcal{S}$ は天候の集合だとする
- 以下の（単純）マルコフ性の仮定をおく

$$p(z_i = s | z_1, \ldots, z_{i-1}) = p(z_i = s | z_{i-1}) \text{ for all } s \in \mathcal{S} \tag{1}$$

- また、s_k から s_l に遷移する確率 $p(z_t = s_l | z_{t-1} = s_k)$ は全ての t で等しいと仮定（斉時性）し、この確率を A_{s_k, s_l} と書く
- $\sum_{l=1}^{K} A_{s_k, s_l} = 1$ が、すべての k で成り立つ
- 便宜的に初期状態を確率変数 z_0 で表し、その値を s_0 とする
  - $A_{s_k, s_0} = 0$ が、すべての k について成り立つ

遷移行列

- 遷移確率をまとめて遷移行列 A として書くと

例. 天候の状態の集合 $S = \{s_0, s_{\text{sun}}, s_{\text{cloud}}, s_{\text{rain}}\}$

$$
A = \begin{array}{c} \\ s_0 \\ s_{\text{sun}} \\ s_{\text{cloud}} \\ s_{\text{rain}} \end{array}
\begin{array}{cccc} s_0 & s_{\text{sun}} & s_{\text{cloud}} & s_{\text{rain}} \\ \end{array}
\begin{bmatrix} 0 & 0.33 & 0.33 & 0.33 \\ 0 & 0.8 & 0.1 & 0.1 \\ 0 & 0.2 & 0.6 & 0.2 \\ 0 & 0.1 & 0.2 & 0.7 \end{bmatrix} \tag{2}
$$

cf. http://cs229.stanford.edu/section/cs229-hmm.pdf

状態列の尤度

- 状態の列 $\boldsymbol{z}_{1:T} \equiv (z_1, \ldots, z_T)$ の尤度 $p(\boldsymbol{z}_{1:T})$ は

$$
\begin{aligned}
p(\boldsymbol{z}_{1:T}; A) &= p(z_1, \ldots, z_T; A) \\
&= p(z_0, z_1, \ldots, z_T; A) \\
&= p(z_T | z_{T-1}, \ldots, z_1; A) \cdots p(z_2 | z_1; A) p(z_1 | z_0; A) \\
&= p(z_T | z_{T-1}; A) \cdots p(z_2 | z_1; A) p(z_1 | z_0; A) \\
&= \prod_{t=1}^{T} p(z_t | z_{t-1}; A) = \prod_{t=1}^{T} A_{z_{t-1}, z_t} \qquad (3)
\end{aligned}
$$

問. どこで単純マルコフ性の仮定を使っているか？

マルコフモデルの最尤推定

$\ln p(\boldsymbol{z}_{1:T}; A)$ を最大化することで A を推定する。

$$\ln p(\boldsymbol{z}_{1:T}; A) = \sum_{t=1}^{T} \ln A_{z_{t-1}, z_t} = \sum_{t=1}^{T} \sum_{l=1}^{K} \sum_{k=1}^{K} \mathbb{1}(z_{t-1} = s_l \wedge z_t = s_k) \ln A_{s_l, s_k} \tag{4}$$

$\mathbb{1}(\cdot)$ は、括弧内の命題が真のとき 1、偽のとき 0、という意味だとする。

$$\mathcal{L}(A, \{\lambda_k\}) = \sum_{t=1}^{T} \sum_{l=1}^{K} \sum_{k=1}^{K} \mathbb{1}(z_{t-1} = s_l \wedge z_t = s_k) \ln A_{s_l, s_k} + \sum_{l=1}^{K} \lambda_l \left(1 - \sum_{k=1}^{K} A_{s_l, s_k}\right) \tag{5}$$

$$\frac{\partial \mathcal{L}(A, \{\lambda_k\})}{\partial A_{s_l, s_k}} = \frac{\sum_{t=1}^{T} \mathbb{1}(z_{t-1} = s_l \wedge z_t = s_k)}{A_{s_l, s_k}} - \lambda_l \tag{6}$$

$\frac{\partial \mathcal{L}(A, \{\lambda_k\})}{\partial A_{s_l, s_k}} = 0$ より $A_{s_l, s_k} \propto \sum_{t=1}^{T} \mathbb{1}(z_{t-1} = s_l \wedge z_t = s_k)$ で、$\sum_{k=1}^{K} A_{s_l, s_k} = 1$ だから

$$\therefore A_{s_l, s_k} = \frac{\sum_{t=1}^{T} \mathbb{1}(z_{t-1} = s_l \wedge z_t = s_k)}{\sum_{k=1}^{K} \sum_{t=1}^{T} \mathbb{1}(z_{t-1} = s_l \wedge z_t = s_k)} = \frac{\sum_{t=1}^{T} \mathbb{1}(z_{t-1} = s_l \wedge z_t = s_k)}{\sum_{t=1}^{T} \mathbb{1}(z_{t-1} = s_l)} \tag{7}$$

Contents

隠れマルコフモデルHMM; hidden Markov model

- 状態 z_t は観測できず、各状態が生成する結果 x_t だけが観測できるとする
  - このとき、状態を表す確率変数 z_t は潜在変数 latent variable となる
- 隠れ状態の列 $\boldsymbol{z}_{1:T} = (z_1, \ldots, z_T)$ は、上述のとおり、単純マルコフ性と斉時性を持つとする
- 時点 t での観測結果を、確率変数 x_t で表す
- 観測データ x_t について、以下のような独立性の仮定をおく

$$p(x_t = v_w | \boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}) = p(x_t = v_w | z_t = s_k) \tag{8}$$

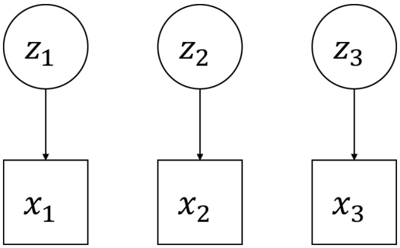
  - つまり、時点 t の観測値は、同じ時点の隠れ状態だけに依存する

観測データがカテゴリカルデータの場合

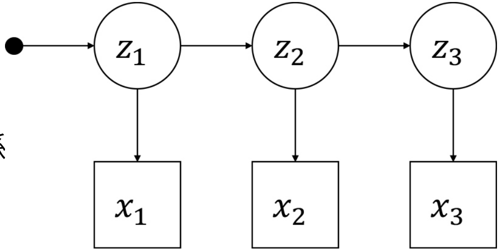
- ここでは、観測データはカテゴリカルデータだとする
  - 観測データが連続値で、$p(x_t|z_t)$ が、例えば正規分布の場合、以下の議論がどうなるか、考えてみよう
- 隠れ状態 s_k がであるときの、アイテム v_w の出現確率 $p(x_t = v_w|z_t = s_k)$ を、B_{s_k,v_w} と書くことにする
  - $(B_{s_k,v_1}, \ldots, B_{s_k,v_W})$ は、状態 s_k に対応するカテゴリカル分布の、パラメータである
  - $\sum_{w=1}^{W} B_{s_k,v_w} = 1$ が成り立つ
- 隠れマルコフモデルにおけるパラメータ推定では、A と B を推定することになる

混合分布モデルと隠れマルコフモデル

- 混合分布モデル
  - 隠れ変数 z_i は独立



- 隠れマルコフモデル
  - 隠れ変数 z_t の間に依存関係
  - 線形で一方向的な依存関係



同時分布

- 隠れマルコフモデルでの観測変数と隠れ変数の同時分布は

$$
\begin{aligned}
p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}; A, B) &= p(\boldsymbol{x}_{1:T} | \boldsymbol{z}_{1:T}; B) p(\boldsymbol{z}_{1:T}; A) \\
&= \prod_{t=1}^{T} p(x_t | z_t; B) \times \prod_{t=1}^{T} p(z_t | z_{t-1}; A) \\
&= \prod_{t=1}^{T} B_{z_t, x_t} \times \prod_{t=1}^{T} A_{z_{t-1}, z_t}
\end{aligned} \tag{9}
$$

  - 状態列の確率 $p(\boldsymbol{z}_{1:T})$ の部分は、式 (3) と同じ
  - ただし、隠れマルコフモデルでは z_t は潜在変数

観測データの尤度

- 観測データの尤度 $p(\boldsymbol{x}; A, B)$ は

$$
\begin{aligned}
p(\boldsymbol{x}; A, B) &= \sum_{\boldsymbol{z}} p(\boldsymbol{x}, \boldsymbol{z}; A, B) \\
&= \sum_{z_1 \in \mathcal{S}} \cdots \sum_{z_T \in \mathcal{S}} \left(\prod_{t=1}^{T} p(x_t | z_t; B) \right) \left(\prod_{t=1}^{T} p(z_t | z_{t-1}; A) \right) \\
&= \sum_{z_1 \in \mathcal{S}} \cdots \sum_{z_T \in \mathcal{S}} \left(\prod_{t=1}^{T} B_{z_t, x_t} \right) \left(\prod_{t=1}^{T} A_{z_{t-1}, z_t} \right) \qquad (10)
\end{aligned}
$$

  - だが、この式をそのまま使って計算すると、計算量が $O(|\mathcal{S}|^T) = O(K^T)$ であるため、非現実的な計算時間になる

Contents

Forward procedure

- HMM のデータ尤度 $p(\boldsymbol{x}; A, B)$ は、動的計画法の一種である forward procedure によって、効率的に計算できる

  cf. https://web.stanford.edu/ jurafsky/slp3/A.pdf

- 時点 t までのデータ列 $(x_1, \ldots, x_t)$ の確率と、時点 t での隠れ状態 z_t が s_k である確率の同時確率を、$\alpha_t(k)$ とおく。つまり

$$\alpha_t(k) \equiv p(x_1, \ldots, x_t, z_t = s_k; A, B) \tag{11}$$

- すると、データ尤度を以下のように表すことができる

$$p(\boldsymbol{x}; A, B) = \sum_{k=1}^{K} p(\boldsymbol{x}_{1:T}, z_T = s_k; A, B) = \sum_{k=1}^{K} \alpha_T(k) \tag{12}$$

$$\alpha_1(k) = p(x_1, z_1 = s_k) = p(z_0 = s_0)p(z_1 = s_k|z_0 = s_0)p(x_1|z_1 = s_k) = A_{0,k}B_{k,x_1} \tag{13}$$

$$\begin{aligned}
\alpha_2(k) &= p(x_1, x_2, z_2 = s_k) = \sum_{z_1 \in \mathcal{S}} p(x_1, x_2, z_1, z_2 = s_k) \\
&= \sum_{z_1 \in \mathcal{S}} p(x_1, z_1)p(z_2 = s_k|x_1, z_1)p(x_2|z_2 = s_k, x_1, z_1) \qquad (14) \\
&= \sum_{z_1 \in \mathcal{S}} p(x_1, z_1)p(z_2 = s_k|z_1)p(x_2|z_2 = s_k) = \sum_{l=1}^{l} \alpha_1(l)A_{s_l,s_k}B_{s_k,x_2} \qquad (15)
\end{aligned}$$

$$\begin{aligned}
\alpha_3(k) &= p(x_1, x_2, x_3, z_3 = s_k) = \sum_{z_2 \in \mathcal{S}} p(x_1, x_2, x_3, z_2, z_3 = s_k) \\
&= \sum_{z_2 \in \mathcal{S}} p(x_1, x_2, z_2)p(z_3 = s_k|x_1, x_2, z_2)p(x_3|z_3 = s_k, x_1, x_2, z_2) \qquad (16) \\
&= \sum_{z_2 \in \mathcal{S}} p(x_1, x_2, z_2)p(z_3 = s_k|z_2)p(x_3|z_3 = s_k) = \sum_{l=1}^{K} \alpha_2(l)A_{s_l,s_k}B_{s_k,x_3} \qquad (17)
\end{aligned}$$

式 (14) から (15)、式 (16) から (17) は、自明でない。→ 条件付き独立性

$$
\begin{aligned}
\alpha_t(k) &= p(\boldsymbol{x}_{1:t}, z_t = s_k) \\
&= \sum_{z_{t-1} \in \mathcal{S}} p(\boldsymbol{x}_{1:t-1}, x_t, z_{t-1}, z_t = s_k) \\
&= \sum_{z_{t-1} \in \mathcal{S}} p(\boldsymbol{x}_{1:t-1}, z_{t-1}) p(z_t = s_k | \boldsymbol{x}_{1:t-1}, z_{t-1}) p(x_t | z_t = s_k, \boldsymbol{x}_{1:t-1}, z_{t-1}) \qquad (18) \\
&= \sum_{z_{t-1} \in \mathcal{S}} p(\boldsymbol{x}_{1:t}, z_{t-1}) p(z_t = s_k | z_{t-1}) p(x_t | z_t = s_k) = \sum_{l=1}^{K} \alpha_{t-1}(l) A_{s_l, s_k} B_{s_k, x_t} \qquad (19)
\end{aligned}
$$

式 (18) から (19) は、自明でない。下記の条件付き独立性を示す必要あり。

z_{t-1} が所与のもとで z_t と $\boldsymbol{x}_{1:t-1}$ は条件付き独立。

z_t が所与のもとで x_t と $\boldsymbol{x}_{1:t-1}, z_{t-1}$ は条件付き独立。

条件付き独立性 conditional independence

- C が与えられたときに A と B が独立である、つまり、$p(A, B|C) = p(A|C)p(B|C)$ が成り立つとき、A と B は C が所与のもとで条件付き独立である、と言う
- 条件付き確率の定義より、$p(A, B|C) = p(A|B, C)p(B|C)$ はいつでも成り立つ。よって

$$A \text{ と } B \text{ は } C \text{ が所与のもとで条件付き独立である}$$
$$\Leftrightarrow p(A|B, C) = p(A|C) \tag{20}$$

注. $p(A, B|C) = p(A|C)p(B|C)$ と $P(A, B) = P(A)P(B)$ は別物

cf. https://www.probabilitycourse.com/chapter1/1_4_4_conditional_independence.php

問 9.1

- $P(A, B, C|D) = P(A|D)P(B, C|D)$、つまり、
  D が所与のもとで A と B, C が条件付き独立ならば、
  $P(A, B|D) = P(A|D)P(B|D)$、つまり、
  D が所与のもとで A と B が条件付き独立になることを、
  示せ。

答え $P(A, B, C|D) = P(A|D)P(B, C|D)$ の両辺を C について周辺化すると、左辺は $\sum_C P(A, B, C|D) = P(A, B|D)$ となり、右辺は $\sum_C P(A|D)P(B, C|D) = P(A|D)P(B|D)$ となる。よって、$P(A, B|D) = P(A|D)P(B|D)$ が言える。

$p(z_t|z_{t-1}, \boldsymbol{x}_{1:t-1}) = p(z_t|z_{t-1})$ を示す。

$$p(z_t|z_{t-1}, \boldsymbol{x}_{1:t-1}) = \frac{p(z_t, z_{t-1}, \boldsymbol{x}_{1:t-1})}{p(z_{t-1}, \boldsymbol{x}_{1:t-1})} \tag{21}$$

$$p(z_{t-1}, \boldsymbol{x}_{1:t-1}) = \sum_{\boldsymbol{z}_{1:t-2}} \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u) \tag{22}$$

$$\begin{aligned} p(z_t, z_{t-1}, \boldsymbol{x}_{1:t-1}) &= \sum_{\boldsymbol{z}_{1:t-2}} p(z_t|z_{t-1}) \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u) \\ &= p(z_t|z_{t-1}) \sum_{\boldsymbol{z}_{1:t-2}} \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u) \end{aligned} \tag{23}$$

$$\therefore p(z_t|z_{t-1}, \boldsymbol{x}_{1:t-1}) = \frac{p(z_t|z_{t-1}) \sum_{\boldsymbol{z}_{1:t-2}} \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u)}{\sum_{\boldsymbol{z}_{1:t-2}} \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u)} = p(z_t|z_{t-1}) \tag{24}$$

z_{t-1} が所与のもとで z_t と $\boldsymbol{x}_{1:t-1}$ は条件付き独立であることが言えた。

$p(x_t|z_t, z_{t-1}, \boldsymbol{x}_{1:t-1}) = p(x_t|z_t)$ を示す。

$$p(x_t|z_t, z_{t-1}, \boldsymbol{x}_{1:t-1}) = \frac{p(x_t, z_t, z_{t-1}, \boldsymbol{x}_{1:t-1})}{p(z_t, z_{t-1}, \boldsymbol{x}_{1:t-1})} \tag{25}$$

式 (23) より

$$p(z_t, z_{t-1}, \boldsymbol{x}_{1:t-1}) = p(z_t|z_{t-1}) \sum_{\boldsymbol{z}_{1:t-2}} \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u)$$

$$\begin{aligned} p(x_t, z_t, z_{t-1}, \boldsymbol{x}_{1:t-1}) &= \sum_{\boldsymbol{z}_{1:t-2}} p(x_t|z_t) p(z_t|z_{t-1}) \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u) \\ &= p(x_t|z_t) p(z_t|z_{t-1}) \sum_{\boldsymbol{z}_{1:t-2}} \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u) \end{aligned} \tag{26}$$

$$\therefore p(x_t|z_t, z_{t-1}, \boldsymbol{x}_{1:t-1}) = \frac{p(x_t|z_t) p(z_t|z_{t-1}) \sum_{\boldsymbol{z}_{1:t-2}} \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u)}{p(z_t|z_{t-1}) \sum_{\boldsymbol{z}_{1:t-2}} \prod_{u=1}^{t-1} p(z_u|z_{u-1}) p(x_u|z_u)} = p(x_t|z_t) \tag{27}$$

z_t が所与のもとで x_t と $\boldsymbol{x}_{1:t-1}, z_{t-1}$ は条件付き独立であることが言えた。

隠れマルコフモデルのEMアルゴリズム

- EMアルゴリズムの導出には、やはりJensenの不等式を使う

$$
\ln p(\boldsymbol{x}_{1:T}) = \ln \sum_{\boldsymbol{z}_{1:T}} p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T})
$$

$$
= \ln \sum_{\boldsymbol{z}_{1:T}} q(\boldsymbol{z}_{1:T}) \frac{p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T})}{q(\boldsymbol{z}_{1:T})} \geq \sum_{\boldsymbol{z}_{1:T}} q(\boldsymbol{z}_{1:T}) \ln \frac{p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T})}{q(\boldsymbol{z}_{1:T})} \tag{28}
$$

- EMアルゴリズムでは、この下界を最大化する

$$
\begin{aligned}
A, B &= \arg\max_{A,B} \sum_{\boldsymbol{z}_{1:T}} q(\boldsymbol{z}_{1:T}) \ln \frac{p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T})}{q(\boldsymbol{z}_{1:T})} \\
&= \arg\max_{A,B} \sum_{\boldsymbol{z}_{1:T}} q(\boldsymbol{z}_{1:T}) \ln p(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T})
\end{aligned} \tag{29}
$$