HMM アルゴリズムメモ

@fhiyo

2018 年 1 月 16 日

1 イントロダクション

このメモは，Bishop の pattern Recogniton of Machine Learning (PRML) [1] の Chapter13 に出てくる HMM (隠れマルコフモデル) についての説明の中で出てくる 3 つのアルゴリズム (Baum-Welch, Forward-Backward, Viterbi) の数式変形について書いたものである．

2 準備

2.1 マルコフモデル

まず，**マルコフモデル**について説明をする．マルコフモデルとは，時間に沿って順に観測されるデータである**時系列観測データ**の各データ (**観測変数**) が，そのデータが観測される時点 i よりも過去のデータに対してのみ従属関係にあるようなデータ列が従うモデルのことである．なお，ある時点 i での観測変数の個数は 1 つだけと仮定する．

ある時系列観測データを $X \stackrel{\text{def}}{=} \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N | \boldsymbol{x}_i \in \mathbb{R}^D\}$ (D : Dimension)

とおき，N 個のデータ点の同時分布を考える．これは確率の乗法定理を用いることで，

$$\begin{aligned}
&p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \\
&= p(\boldsymbol{x}_1) p(\boldsymbol{x}_2, \ldots, \boldsymbol{x}_N | \boldsymbol{x}_1) \\
&= p(\boldsymbol{x}_1) p(\boldsymbol{x}_2 | \boldsymbol{x}_1) p(\boldsymbol{x}_3, \ldots, \boldsymbol{x}_N | \boldsymbol{x}_1, \boldsymbol{x}_2) \\
&= \ldots \\
&= p(\boldsymbol{x}_1) \prod_{n=2}^{N} p(\boldsymbol{x}_n | \boldsymbol{x}_{n-1}, \ldots, \boldsymbol{x}_1)
\end{aligned} \tag{1}$$

と書ける．

2.1.1 1 次マルコフモデル

次に，**1 次マルコフモデル**について説明する．これはある時点 i での観測変数が，その 1 つ前の時点の観測変数にのみ従属であるような確率モデルである．式 (1) を参考にして 1 次マルコフモデルを定式化すると，

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = p(\boldsymbol{x}_1) \prod_{n=2}^{N} p(\boldsymbol{x}_n | \boldsymbol{x}_{n-1}) \tag{2}$$

のように書ける．ここで，式 (2) は

$$p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1},\ldots,\boldsymbol{x}_1) = p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1}) \qquad (n \in \{1,\ldots,N\})$$

という 1 次マルコフモデルの特性を定式化して導出した．

2.2 隠れマルコフモデル

隠れマルコフモデル (Hidden Markov Model, **HMM**) とは，データとして得られる状態 (観測変数) の他に，その観測変数がどの状態としてが出力されるかの確率に影響を与えるが，出力はされないために実際の状態がわからない離散的な値をとる**潜在変数** (**隠れ変数**) というものを導入したマルコフモデルのことである．以下，潜在変数を $Z \overset{\text{def}}{=} \{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N\}$ と表す．隠れマルコフモデルの具体的な条件を書くと，

- 観測変数 $\boldsymbol{x}_n$ は潜在変数 $\boldsymbol{z}_n$ の影響を受けて出力される
- 観測変数どうしは独立である
- 潜在変数どうしは 1 次マルコフモデルに従い生成される

となる．潜在変数の状態の個数である**状態数**を K，隠れマルコフモデルのパラメータをそれぞれ $\boldsymbol{\pi}, A, \boldsymbol{\phi}$ とおき，$\theta = \{\boldsymbol{\pi}, A, \boldsymbol{\phi}\}$ とおく．ただし，$\boldsymbol{\pi} = (\pi_1,\ldots,\pi_K)^{\mathrm{T}}$ は潜在変数の初期変数である $\boldsymbol{z}_1$ の状態を決定する確率である**初期確率** (←用語は適切か？) を表し，$A = [A_{jk}]_{K\times K}$ は, 任意の時点での潜在変数の状態が j のとき，次の時点で潜在変数の状態が k に遷移する確率である**遷移確率**を表す．また，$\boldsymbol{\phi} = (\phi_1,\ldots,\phi_K)^{\mathrm{T}}$ は任意の時点で潜在変数の状態が決定したとき，観測状態が出力される確率である**出力確率**を表す．ここで，$\pi_k \overset{\text{def}}{=} p(z_{1k}=1)$, $A_{jk} \overset{\text{def}}{=} p(z_{nk}|z_{n-1,j})$ とする．

このような条件のもとに隠れマルコフモデルを定式化すると以下のように表せる．

$$\begin{aligned}
p(X,Z|\theta) &= p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N|\theta) \\
&= p(\boldsymbol{z}_1|\theta)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N,\ldots,\boldsymbol{z}_N|\boldsymbol{z}_1,\theta) \\
&= p(\boldsymbol{z}_1|\boldsymbol{\pi})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_N|\boldsymbol{z}_1,\theta) \\
&= p(\boldsymbol{z}_1|\boldsymbol{\pi})\prod_{n=2}^{N} p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1},\theta)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N|\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N,\theta) \\
&= p(\boldsymbol{z}_1|\boldsymbol{\pi})\prod_{n=2}^{N} p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1},A)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N|\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N,\theta) \\
&= p(\boldsymbol{z}_1|\boldsymbol{\pi})\prod_{n=2}^{N} p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1},A)\prod_{n=1}^{N} p(\boldsymbol{x}_n|\boldsymbol{z}_n,\theta) \\
&= p(\boldsymbol{z}_1|\boldsymbol{\pi})\prod_{n=2}^{N} p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1},A)\prod_{n=1}^{N} p(\boldsymbol{x}_n|\boldsymbol{z}_n,\boldsymbol{\phi})
\end{aligned} \tag{3}$$

初期確率 $\boldsymbol{\pi}$ について，$\pi_k = p(z_{1k}=1)$ より，

$$p(\boldsymbol{z}_1|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{1k}} \tag{4}$$

と表すことができる．さらに，$A_{jk} = p(z_{nk}|z_{n-1,j})$ より，

$$p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1},A) = \prod_{k=1}^{K}\prod_{j=1}^{K} A_{jk}^{z_{n-1,j}z_{nk}} \tag{5}$$

と表せる．

2.3 数学的準備

各アルゴリズムの説明のために必要な数学的準備を行う．ある時系列観測データを $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\}$，
その時系列観測データの裏にある潜在変数列を $Z = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_N\}$
と表す．ただし，$x_n \in \mathbb{R}^D$，$(n \in \{1, \dots, N\})$ とする．各 $\boldsymbol{z}_n$ は，潜在変数の状態数を K とすると，

$$
\begin{aligned}
&\boldsymbol{z}_n = \begin{bmatrix} z_{n1} \\ \vdots \\ z_{nK} \end{bmatrix} \\
&z_{nk} \in \{0, 1\}, (\forall n)\ n \in \{1, \dots, N\}, (\forall k)\ k \in \{0, \dots, K\} \\
&\sum_{k=1}^{K} z_{nk} = 1, n \in \{1, \dots, N\}
\end{aligned}
$$

を満たすものとする．つまり $\boldsymbol{z}_n$ はその K 個あるベクトルの成分のうちどれか 1 つだけが 1 で残りが 0 であるようないわゆる 1-of-K 表現のベクトルである．

これから EM アルゴリズムを用いて隠れマルコフモデルのパラメータの更新を行うにあたり，以下の記号を導入する．

$$
\begin{aligned}
Q(\theta, \theta^{old}) &\stackrel{\text{def}}{=} \mathbb{E}[\log p(X, Z|\theta)] = \sum_{Z} p(Z|\theta^{old}) \log p(X, Z|\theta) \\
\gamma(\boldsymbol{z}_n) &\stackrel{\text{def}}{=} p(\boldsymbol{z}_n|X, \theta^{old}) \\
\xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n) &\stackrel{\text{def}}{=} p(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n|X, \theta^{old}) \\
\gamma(z_{nk}) &\stackrel{\text{def}}{=} \mathbb{E}[z_{nk}] = \sum_{\boldsymbol{z}_n} \gamma(\boldsymbol{z}_n) z_{nk} \\
\xi(z_{n-1,j}, z_{nk}) &\stackrel{\text{def}}{=} \mathbb{E}[z_{n-1,j} z_{nk}] = \sum_{\boldsymbol{z}_{n-1}, \boldsymbol{z}_n} \xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n) z_{n-1,j} z_{nk}
\end{aligned}
$$

ここで，θ^{old} はパラメータの更新を行う前の古いパラメータを表している．上の各記号について簡単に説明すると，

- $Q(\theta, \theta^{old})$
  隠れマルコフモデルを表す確率分布 $p(X, Z|\theta)$ の対数尤度関数の期待値
- $\gamma(\boldsymbol{z}_n)$
  古いパラメータ θ^{old} 上での潜在変数 $\boldsymbol{z}_n$ の確率分布
- $\xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)$
  古いパラメータ θ^{old} 上での潜在変数の組 $(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)$ の同時確率分布
- $\gamma(z_{nk})$
  $z_{nk} = 1$ となる期待値
- $\xi(z_{n-1,j}, z_{nk})$
  $z_{n-1,j} = 1, z_{nk} = 1$ となる期待値

である．なお，$\gamma(z_{nk})$，$\xi(z_{n-1,j}, z_{nk})$ を負担率と呼ぶ．

3 Baum-Welch algorithm

3.1 概要

Baum-Welch algorithm とは，HMM において EM アルゴリズムを用いて HMM のパラメータを最適化するアルゴリズムである．

· 手順

1. E step
   現在推定されているパラメータに対する，モデルの尤度関数 $p(X, Z|\theta)$ の (対数の) 期待値を求める．
2. M step
   E step で求めた期待値を最大化するようなパラメータを求める．
3. 繰り返し
   1, 2 をパラメータの値が収束または値の更新時の差分が一定以下になるまで繰り返す．

3.1.1 入力

観測データ列 $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$，初期状態のパラメータ $\theta^{old} = \{\boldsymbol{\pi}, A, \boldsymbol{\phi}\}$
潜在変数が出力される期待値 $\gamma(z_{nk})$，$\xi(z_{n-1,j}, z_{nk})$ (EM の更新のたびに必要)

3.1.2 出力

最尤推定したパラメータ $\theta = \{\boldsymbol{\pi}, A, \boldsymbol{\phi}\}$

3.2 アルゴリズム

3.2.1 E step

尤度関数の期待値 $Q(\theta, \theta^{old})$ を $\gamma(z_{nk})$ と $\xi(z_{n-1,j}, z_{nk})$ の式に変形する．
そのように変形する理由は，section 4 で $\gamma(z_{nk})$ と $\xi(z_{n-1,j}, z_{nk})$ を陽に求めることができるからである．
式 (3) に注意すると，

$$
\begin{aligned}
Q(\theta, \theta^{old}) &= \sum_{Z} p(Z|\theta^{old}) \log p(X, Z|\theta) \\
&= \sum_{Z} p(Z|\theta^{old}) \log \left(p(\boldsymbol{z}_1|\boldsymbol{\pi}) \prod_{n=2}^{N} p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}, A) \prod_{n=1}^{N} p(\boldsymbol{x}_n|\boldsymbol{z}_n, \boldsymbol{\phi}) \right) \\
&= \sum_{Z} p(Z|\theta^{old}) \left(\log p(\boldsymbol{z}_1|\boldsymbol{\pi}) + \sum_{n=2}^{N} \log p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}, A) + \sum_{n=1}^{N} \log p(\boldsymbol{x}_n|\boldsymbol{z}_n, \boldsymbol{\phi}) \right) \\
&= \sum_{Z} p(Z|\theta^{old}) \log p(\boldsymbol{z}_{\boldsymbol{1}}|\boldsymbol{\pi}) + \sum_{Z} p(Z|\theta^{old}) \sum_{n=2}^{N} \log p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}, A) + \sum_{Z} p(Z|\theta^{old}) \sum_{n=1}^{N} \log p(\boldsymbol{x}_n|\boldsymbol{z}_n, \boldsymbol{\phi})
\end{aligned}
\tag{6}
$$

となる．さて，$p(Z|X,\theta^{old})$ について，Z が 1 次マルコフモデルであることに気をつけると，

$$\begin{aligned}
p(Z|X,\theta^{old}) &= p(\boldsymbol{z}_1,...,\boldsymbol{z}_N|X,\theta^{old}) \\
&= p(\boldsymbol{z}_1|X,\theta^{old})p(\boldsymbol{z}_2,....,\boldsymbol{z}_N|\boldsymbol{z}_1,X,\theta^{old}) \\
&\vdots \\
&= p(\boldsymbol{z}_1|X,\theta^{old})\prod_{n=2}^{N}p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1},X,\theta^{old})
\end{aligned} \tag{7}$$

と変形できる．$\sum_Z p(Z|X,\theta^{old})=1$ より，

$$\sum_{\boldsymbol{z}_1}p(\boldsymbol{z}_1|X,\theta^{old})\sum_{\boldsymbol{z}_2}p(\boldsymbol{z}_2|\boldsymbol{z}_1=\boldsymbol{z}'_1,X,\theta^{old})\cdots\sum_{\boldsymbol{z}_N}p(\boldsymbol{z}_N|\boldsymbol{z}_{N-1}=\boldsymbol{z}'_{N-1},X,\theta^{old})=1 \tag{8}$$

となる．ここで，式 (8) の各項は確率密度関数の定義から 1 である．

以上より，(7),(8) から，(6) の各項は，

$$\begin{aligned}
&\sum_Z p(Z|X,\theta^{old})\log p(\boldsymbol{z}_1,\pi) \\
&= \sum_{\boldsymbol{z}_1}p(\boldsymbol{z}_1|X,\theta^{old})\log p(\boldsymbol{z}_1,\pi)\sum_{\boldsymbol{z}_2,...\boldsymbol{z}_N}p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1},X,\theta) \\
&= \sum_{\boldsymbol{z}_1}p(\boldsymbol{z}_1|X,\theta^{old})\log p(\boldsymbol{z}_1,\pi)
\end{aligned} \tag{9}$$

$$\begin{aligned}
\sum_Z p(Z|\theta^{old})\sum_{n=2}^{N}\log p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1},A) &= \sum_Z p(Z|\theta^{old})\{\log p(\boldsymbol{z}_2|\boldsymbol{z}_1)+\cdots+\log p(\boldsymbol{z}_N|\boldsymbol{z}_{N-1})\} \\
&= \sum_Z p(Z|\theta^{old})\log p(\boldsymbol{z}_2|\boldsymbol{z}_1)+\cdots+\sum_Z p(Z|\theta^{old})\log p(\boldsymbol{z}_N|\boldsymbol{z}_{N-1}) \\
&= \sum_{n=2}^{N}\sum_Z p(Z|\theta^{old})\log p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}) \\
&= \sum_{n=2}^{N}\sum_{\boldsymbol{z}_n,\boldsymbol{z}_{n-1}}p(\boldsymbol{z}_n,\boldsymbol{z}_{n-1})\log p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}) \\
&\qquad \sum_{\boldsymbol{z}_1,...,\boldsymbol{z}_{n-2},\boldsymbol{z}_{n+1},...,\boldsymbol{z}_N}p(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{n-2},\boldsymbol{z}_{n+1},\ldots,\boldsymbol{z}_N|\boldsymbol{z}_n,\boldsymbol{z}_{n-1},\theta^{old}) \\
&= \sum_{n=2}^{N}\sum_{\boldsymbol{z}_n,\boldsymbol{z}_{n-1}}p(\boldsymbol{z}_n,\boldsymbol{z}_{n-1}|\theta^{old})\log p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1})
\end{aligned} \tag{10}$$

$$\begin{aligned}
&\sum_Z p(Z|X,\theta^{old})\sum_{n=1}^{N}\log p(\boldsymbol{x}_n|\boldsymbol{z}_n,\boldsymbol{\phi}) \\
&= \sum_{n=1}^{N}\sum_{\boldsymbol{z}_n}p(\boldsymbol{z}_n|X,\theta^{old})\log p(\boldsymbol{x}_n|\boldsymbol{z}_n,\boldsymbol{\phi})\sum_{\boldsymbol{z}_1,...,\boldsymbol{z}_{n-1},\boldsymbol{z}_{n+1},...,\boldsymbol{z}_N}p(\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{n-1},\boldsymbol{z}_{n+1},\ldots,\boldsymbol{z}_N|\boldsymbol{z}_n,X,\theta^{old}) \\
&= \sum_{n=1}^{N}\sum_{\boldsymbol{z}_n}p(\boldsymbol{z}_n|X,\theta^{old})\log p(\boldsymbol{x}_n|\boldsymbol{z}_n,\boldsymbol{\phi})
\end{aligned} \tag{11}$$

のように変形できる．

(9)，(10)，(11) を (6) に代入すると，

$$
\begin{aligned}
Q(\theta, \theta^{old}) &= \sum_{Z} p(Z|\theta^{old}) \log p(X, Z|\theta) \\
&= \sum_{\boldsymbol{z}_1} p(\boldsymbol{z}_1|X, \theta^{old}) \log p(\boldsymbol{z}_1, \pi) + \sum_{n=2}^{N} \sum_{\boldsymbol{z}_n, \boldsymbol{z}_{n-1}} p(\boldsymbol{z}_n, \boldsymbol{z}_{n-1}|\theta^{old}) \log p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}, A) \\
&\qquad + \sum_{n=1}^{N} \sum_{\boldsymbol{z}_n} p(\boldsymbol{z}_n|X, \theta^{old}) \log p(\boldsymbol{x}_n|\boldsymbol{z}_n, \boldsymbol{\phi})
\end{aligned} \tag{12}
$$

式 (12) の各項について，式 (4), (5) に気をつけながら潜在変数の期待値についての表現に直すと，

$$
\begin{aligned}
&\sum_{\boldsymbol{z}_1} p(\boldsymbol{z}_1|X, \theta^{old}) \log p(\boldsymbol{z}_1|\boldsymbol{\pi}) \\
&= \sum_{\boldsymbol{z}_1} \gamma(\boldsymbol{z}_1) \log \sum_{k=1}^{K} \pi_k^{z_{1k}} \\
&= \gamma\left(\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) \log \sum_{k=1}^{K} \pi_k^{z_{1k}} + \cdots + \gamma\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}\right) \log \sum_{k=1}^{K} \pi_k^{z_{1k}} \\
&= \gamma(z_{11}) \log \pi_1 + \cdots + \gamma(z_{1K}) \log \pi_K \\
&= \sum_{k=1}^{K} \gamma(z_{1k}) \log \pi_k
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
&\sum_{n=2}^{N} \sum_{\boldsymbol{z}_n, \boldsymbol{z}_{n-1}} p(\boldsymbol{z}_n, \boldsymbol{z}_{n-1}|\theta^{old}) \log p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}) \\
&= \sum_{n=2}^{N} \sum_{\boldsymbol{z}_n, \boldsymbol{z}_{n-1}} \xi(\boldsymbol{z}_n, \boldsymbol{z}_{n-1}) \log \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{n-1,j} z_{nk}} \\
&= \sum_{n=2}^{N} \xi\left(\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) \log A_{11} + \cdots + \xi\left(\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}\right) \log A_{1K} + \cdots + \xi\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}\right) \log A_{KK} \\
&= \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \log A_{jk}
\end{aligned} \tag{14}
$$

$$
\begin{aligned}
&\sum_{n=1}^{N} \sum_{\boldsymbol{z}_n} p(\boldsymbol{z}_n|X, \theta^{old}) \log p(\boldsymbol{x}_n|\boldsymbol{z}_n, \boldsymbol{\phi}) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \log p(\boldsymbol{x}_n|\phi_k)
\end{aligned} \tag{15}
$$

(13),(14),(15) を (12) に代入すると，

$$Q(\theta, \theta^{old}) = \sum_{k=1}^{K} \gamma(z_{1k}) \log \pi_k + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \log A_{jk} + \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \log p(\boldsymbol{x}_n | \phi_k) \quad (16)$$

式 (16) より，隠れマルコフモデルを表す確率分布関数の現在のパラメータに関する期待値を潜在変数が出力される期待値である $\gamma(z_{nk})$，$\xi(z_{n-1,j}, z_{nk})$ によって表すことができた．

あとはこのモデルの期待値を最大化するパラメータを M step で求めればよい．

3.2.2 M step

ラグランジュの未定乗数法を用いて，式 (16) を最大化するパラメータを求める．パラメータの制約条件は，

$$\sum_{k=1}^{K} \pi_k = 1 \quad (17)$$

$$\sum_{k=1}^{K} A_{jk} = 1 (\forall j, j \in \{1, \ldots, K\}) \quad (18)$$

である．よって以下の関数 $F(\boldsymbol{\pi}, A, \boldsymbol{\phi})$ を考えると，各ラグランジュ未定乗数を $\lambda_i (i \in \{1, \ldots, K+1\})$ として，

$$F(\boldsymbol{\pi}, A, \boldsymbol{\phi}) = Q - \lambda_1 \left(\sum_{k=1}^{K} \pi_k - 1 \right) - \sum_{j=1}^{K} \left\{ \lambda_{j+1} \left(\sum_{k=1}^{K} A_{jk} - 1 \right) \right\}$$

という式で表すことができる. この関数を各パラメータに対して最大化する．

まず，$\boldsymbol{\pi}$ について考えると，

$$\begin{aligned} \frac{\partial F}{\partial \pi_k} &= \frac{\partial Q}{\partial \pi_k} - \lambda_1 \\ &= \frac{\gamma(z_{1k})}{\pi_k} - \lambda_1 = 0 \end{aligned}$$

より，

$$\pi_k = \frac{\gamma(z_{1k})}{\lambda_1}$$

となる．(17) から，

$$\begin{aligned} &\sum_{k=1}^{K} \pi_k = 1 \\ &\Leftrightarrow \sum_{k=1}^{K} \frac{\gamma(z_{1k})}{\lambda_1} = 1 \\ &\Leftrightarrow \lambda_1 = \sum_{k=1}^{K} \gamma(z_{1k}) \end{aligned}$$

以上より，

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{k=1}^{K} \gamma(z_{1k})} \quad (19)$$

となる．

次に A について考える．

$$\frac{\partial F}{\partial A_{jk}} = \sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk}) \frac{1}{A_{jk}} - \lambda_{j+1} = 0$$

より，

$$A_{jk} = \frac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})}{\lambda_{j+1}}$$

となる．(18) から，

$$\begin{aligned}
&\sum_{k=1}^{K} A_{jk} = 1 \\
&\sum_{k=1}^{K} \frac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})}{\lambda_{j+1}} = 1 \\
&\Leftrightarrow \lambda_{j+1} = \sum_{k=1}^{K} \sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})
\end{aligned}$$

以上より，

$$A_{jk} = \frac{\sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})}{\sum_{k=1}^{K} \sum_{n=2}^{N} \xi(z_{n-1,j}, z_{nk})} \tag{20}$$

となる．

最後に，$\boldsymbol{\phi}$ に関して F を最大化する．ϕ に関しては出力の確率分布関数がどのような分布を仮定するかによるので，ここでは正規分布のときの例を考える．すなわち，

$$\begin{aligned}
p(\boldsymbol{x}_n|\phi_k) &= \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma_k) \\
&= \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma_k|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^{\mathrm{T}} \Sigma_k^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k)\} \\
\Leftrightarrow \log p(\boldsymbol{x}_n|\phi_k) &= C - \frac{1}{2} \log |\Sigma_k| - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^{\mathrm{T}} \Sigma_k^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k) \qquad (C : \text{定数})
\end{aligned}$$

よって，

$$\begin{aligned}
&\frac{\partial F}{\partial \mu_k} = \sum_{n=1}^{N} \gamma(z_{nk}) \left[(\boldsymbol{x}_n - \boldsymbol{\mu}_k) \{\Sigma_k^{-1} + (\Sigma_k^{-1})^{\mathrm{T}}\} \right] = 0 \\
&\Leftrightarrow \{\Sigma_k^{-1} + (\Sigma_k^{-1})^{\mathrm{T}}\} \sum_{n=1}^{N} \gamma(z_{nk})(\boldsymbol{x}_n - \boldsymbol{\mu}_k) = 0 \\
&\Leftrightarrow \boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{x}_n}{\sum_{n=1}^{N} \gamma(z_{nk})} \qquad (21)
\end{aligned}$$

また，

$$
\frac{\partial F}{\partial \Sigma_k} = \frac{\partial}{\partial \Sigma_k}\sum_{n=1}^{N}\sum_{k'=1}^{K}\gamma(z_{nk'})\left\{C - \frac{1}{2}\log|\Sigma_{k'}| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma_{k'}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} = 0
$$

$$
\Leftrightarrow \sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial \Sigma_k}\log|\Sigma_k| + \sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial \Sigma_k}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = 0 \tag{22}
$$

ここで，任意の行列 A とベクトル $\boldsymbol{x}$ について，

$$
\frac{\partial}{\partial A}\log|A| = \left(A^{-1}\right)^{\mathrm{T}}
$$

と，

$$
\begin{aligned}
\frac{\partial}{\partial A}\left(\boldsymbol{x}^{\mathrm{T}}A^{-1}\boldsymbol{x}\right) &= \frac{\partial}{\partial A}\mathrm{trace}\left(A^{-1}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right) \\
&= -A^{-1}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}A^{-1}
\end{aligned}
$$

である事実を利用すると[*1]，(22) は，

$$
\sum_{n=1}^{N}\gamma(z_{nk})\Sigma_k^{-1} - \sum_{n=1}^{N}\gamma(z_{nk})\Sigma_k^{-1}\left(\boldsymbol{x}_n-\boldsymbol{\mu}_k\right)\left(\boldsymbol{x}_n-\boldsymbol{\mu}_k\right)^{\mathrm{T}}\Sigma_k^{-1} = 0
$$

$$
\Leftrightarrow \Sigma_k = \frac{\sum_{n=1}^{N}\gamma(z_{nk})(\boldsymbol{x}_n-\boldsymbol{\mu}_k)(\boldsymbol{x}_n-\boldsymbol{\mu}_k)^{\mathrm{T}}}{\sum_{n=1}^{N}\gamma(z_{nk})} \tag{23}
$$

であることがわかる．なお，Σ_k^{-1} は精度行列と呼ばれるものであり，対称行列であることから $\left(\Sigma_k^{-1}\right)^{\mathrm{T}} = \Sigma_k^{-1}$ であることに注意する．

(19),(20),(21),(23) と，後述する Forward-Backward アルゴリズムを用いて $\gamma(\boldsymbol{z}_n)$ と $\xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)$ について計算すればモデルの期待値を陽に求めることができる．

4 Forward-Backward algorithm

4.1 概要

モデルの各潜在変数の事後確率である $\gamma(\boldsymbol{z}_n)$ と $\xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)$ を求めるためのアルゴリズムである．観測変数 X を取得したときの潜在変数の推定ができる．

なお section 4，及び section 5 では確率分布のパラメータ依存性を式に明記しないことにする．

4.1.1 入力

観測データ列 $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$，パラメータ $\theta^{old} = \{\boldsymbol{\pi}, A, \boldsymbol{\phi}\}$

4.1.2 出力

$\gamma(\boldsymbol{z}_n)$ と $\xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)$ の値

[*1] PRML の確率分布の章の練習問題の解答にもう少し細かい式展開があったはず．

4.2 アルゴリズム

Bayes' theorem を用いると，$\gamma(\boldsymbol{z}_n)$ は

$$\gamma(\boldsymbol{z}_n) = p(\boldsymbol{z}_n|X) = \frac{p(X|\boldsymbol{z}_n)p(\boldsymbol{z}_n)}{p(X)}$$

と表せる．ここで，任意の観測変数どうしは独立なので，

$$\begin{aligned}
\gamma(\boldsymbol{z}_n) &= \frac{p(X|\boldsymbol{z}_n)p(\boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N|\boldsymbol{z}_n)p(\boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n|\boldsymbol{z}_n)p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{z}_n)p(\boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{z}_n)p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{z}_n)p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{z}_n)}{p(X)} \\
&= \frac{\alpha(\boldsymbol{z}_n)\beta(\boldsymbol{z}_n)}{p(X)} \\
&= \frac{\alpha(\boldsymbol{z}_n)\beta(\boldsymbol{z}_n)}{\sum_{\boldsymbol{z}_n}\alpha(\boldsymbol{z}_n)\beta(\boldsymbol{z}_n)} \qquad (24)
\end{aligned}$$

と表すことが出来る．ただし，

$$\alpha(\boldsymbol{z}_n) \stackrel{\text{def}}{=} \begin{cases} \displaystyle\prod_{k=1}^{K} \{\pi_k p(\boldsymbol{x}_1|\phi_k)\}^{z_{nk}} & (n=1) \\ p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{z}_n) & (n \in \{2,\dots,N\}) \end{cases}$$

$$\beta(\boldsymbol{z}_n) \stackrel{\text{def}}{=} \begin{cases} 1 & (n=N) \\ p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{z}_n) & (n \in \{1,\dots,N-1\}) \end{cases}$$

とする．このように定義すると，$\alpha(\boldsymbol{z}_n)$,$\beta(\boldsymbol{z}_n)$ のそれぞれについて漸化式を立てることができ，

$$\begin{aligned}
\alpha(\boldsymbol{z}_n) &= p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_n) \\
&= p(\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n|\boldsymbol{z}_n) \\
&= p(\boldsymbol{z}_n)p(\boldsymbol{x}_n|\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-1}|\boldsymbol{x}_n,\boldsymbol{z}_n) \\
&= p(\boldsymbol{z}_n)p(\boldsymbol{x}_n|\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-1}|\boldsymbol{z}_n) \\
&= p(\boldsymbol{x}_n|\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-1},\boldsymbol{z}_n) \\
&= p(\boldsymbol{x}_n|\boldsymbol{z}_n)\sum_{\boldsymbol{z}_{n-1}} p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-1},\boldsymbol{z}_{n-1},\boldsymbol{z}_n) \\
&= p(\boldsymbol{x}_n|\boldsymbol{z}_n)\sum_{\boldsymbol{z}_{n-1}} p(\boldsymbol{z}_{n-1},\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-1}|\boldsymbol{z}_{n-1},\boldsymbol{z}_n) \\
&= p(\boldsymbol{x}_n|\boldsymbol{z}_n)\sum_{\boldsymbol{z}_{n-1}} p(\boldsymbol{z}_{n-1},\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-1}|\boldsymbol{z}_{n-1}) \\
&= p(\boldsymbol{x}_n|\boldsymbol{z}_n)\sum_{\boldsymbol{z}_{n-1}} p(\boldsymbol{z}_{n-1})p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-1}|\boldsymbol{z}_{n-1}) \\
&= p(\boldsymbol{x}_n|\boldsymbol{z}_n)\sum_{\boldsymbol{z}_{n-1}} p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n-1},\boldsymbol{z}_{n-1})p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}) \\
&= p(\boldsymbol{x}_n|\boldsymbol{z}_n)\sum_{\boldsymbol{z}_{n-1}} \alpha(\boldsymbol{z}_{n-1})p(\boldsymbol{z}_n|\boldsymbol{z}_{n-1}) \qquad (25)
\end{aligned}$$

β についても同様にして計算すると，

$$\begin{aligned}
\beta(\boldsymbol{z}_n) &= p(\boldsymbol{x}_{n+1},\ldots,\boldsymbol{x}_N|\boldsymbol{z}_n) \\
&= \sum_{\boldsymbol{z}_{n+1}} p(\boldsymbol{x}_{n+1},\ldots,\boldsymbol{x}_N,\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) \\
&= \sum_{\boldsymbol{z}_{n+1}} p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n)p(\boldsymbol{x}_{n+1},\ldots,\boldsymbol{x}_N|\boldsymbol{z}_{n+1},\boldsymbol{z}_n) \\
&= \sum_{\boldsymbol{z}_{n+1}} p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n)p(\boldsymbol{x}_{n+1},\ldots,\boldsymbol{x}_N|\boldsymbol{z}_{n+1}) \\
&= \sum_{\boldsymbol{z}_{n+1}} p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n)p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1})p(\boldsymbol{x}_{n+2},\ldots,\boldsymbol{x}_N|\boldsymbol{x}_{n+1},\boldsymbol{z}_{n+1}) \\
&= \sum_{\boldsymbol{z}_{n+1}} p(\boldsymbol{x}_{n+2},\ldots,\boldsymbol{x}_N|\boldsymbol{z}_{n+1})p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1})p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) \\
&= \sum_{\boldsymbol{z}_{n+1}} \beta(\boldsymbol{z}_{n+1})p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1})p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) \qquad (26)
\end{aligned}$$

となる．

(25), (26) 後は漸化式を解けば $\alpha(\boldsymbol{z}_n)$, $\beta(\boldsymbol{z}_n)$ は陽に求まる．次は $\gamma(\boldsymbol{z}_n)$, $\xi(\boldsymbol{z}_{n-1},\boldsymbol{z}_n)$ を α,β を使って表すことが出来れば任意の潜在変数の最尤値が求まり，目的が達成される．

ところで式 (24) の分母について，$p(X)=\sum_{\boldsymbol{z}_n}\alpha(\boldsymbol{z}_n)\beta(\boldsymbol{z}_n)$ は任意の n について成り立つので，

$$\begin{aligned}
p(X) &= \sum_{\boldsymbol{z}_n}\alpha(\boldsymbol{z}_n)\beta(\boldsymbol{z}_n) \\
&= \sum_{\boldsymbol{z}_N}\alpha(\boldsymbol{z}_N)\beta(\boldsymbol{z}_N) \\
&= \sum_{\boldsymbol{z}_N}\alpha(\boldsymbol{z}_N)
\end{aligned}$$

で計算することが出来る．つまり式 (24) は

$$\gamma(\boldsymbol{z}_n) = \frac{\alpha(\boldsymbol{z}_n)\beta(\boldsymbol{z}_n)}{\sum_{\boldsymbol{z}_N} \alpha(\boldsymbol{z}_N)} \tag{27}$$

と表せる．次に $\xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)$ について考える．

$$\begin{aligned}
\xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n) &= p(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n | X) \\
&= \frac{p(X|\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)p(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n | \boldsymbol{z}_{n-1}, \boldsymbol{z}_n)p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{z}_{n-1}, \boldsymbol{z}_n)p(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n | \boldsymbol{z}_{n-1}, \boldsymbol{z}_n)p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N | \boldsymbol{z}_{n-1}, \boldsymbol{z}_n)p(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n | \boldsymbol{z}_{n-1}, \boldsymbol{z}_n)p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N | \boldsymbol{z}_n)p(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1} | \boldsymbol{z}_{n-1}, \boldsymbol{z}_n)p(\boldsymbol{x}_n | \boldsymbol{z}_{n-1}, \boldsymbol{z}_n)p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N | \boldsymbol{z}_n)p(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1} | \boldsymbol{z}_{n-1})p(\boldsymbol{x}_n | \boldsymbol{z}_n)p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N | \boldsymbol{z}_n)p(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)}{p(X)} \\
&= \frac{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}, \boldsymbol{z}_{n-1})p(\boldsymbol{x}_n | \boldsymbol{z}_n)p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N | \boldsymbol{z}_n)p(\boldsymbol{z}_n | \boldsymbol{z}_{n-1})}{p(X)} \\
&= \frac{\alpha(\boldsymbol{z}_{n-1})p(\boldsymbol{x}_n | \boldsymbol{z}_n)p(\boldsymbol{z}_n | \boldsymbol{z}_{n-1})\beta(\boldsymbol{z}_n)}{p(X)} \\
&= \frac{\alpha(\boldsymbol{z}_{n-1})p(\boldsymbol{x}_n | \boldsymbol{z}_n)p(\boldsymbol{z}_n | \boldsymbol{z}_{n-1})\beta(\boldsymbol{z}_n)}{\sum_{\boldsymbol{z}_N} \alpha(\boldsymbol{z}_N)}
\end{aligned} \tag{28}$$

(27) と (28) より目標は達成された．

5 Viterbi algorithm

5.1 概要

先ほどの Forward-Backward アルゴリズムでは，各潜在変数についての事後確率を求めた．しかし，各々の潜在変数について，観測データ列が与えられたときの事後確率が最大となるような潜在変数の列 $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ を推定することはできない．そこで提案されるのが Viterbi アルゴリズムである．このアルゴリズムはある観測データ列 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ が出力されたときの潜在変数の列 $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ の事後確率を最大化するアルゴリズムである．

5.1.1 入力

パラメータ $\theta = \{\boldsymbol{\pi}, A, \boldsymbol{\phi}\}$

観測データ列 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$

5.1.2 出力

事後確率 $p(Z|X)$ を最大化する潜在変数の列 $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$

任意の時点 i における最大事後確率 $p(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_i | X, \theta)$

5.2 アルゴリズム

以下の式で与えられる $\omega(\boldsymbol{z}_n)$ を考える．

$$\omega(\boldsymbol{z}_n) = \begin{cases} \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{n-1}} \log p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n) & (n>1) \\ \quad \omega(\boldsymbol{z}_1) = \log(\boldsymbol{x}_1,\boldsymbol{z}_1) & (n=1) \end{cases} \tag{29}$$

以下，$\omega(\boldsymbol{z}_n)$ についての漸化式を解く．式 (29) より，

$$\begin{aligned}
\omega(\boldsymbol{z}_{n+1}) &= \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n} \log p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n+1},\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{n+1}) \\
&= \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n} \log p(\boldsymbol{z}_{n+1})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{n+1},\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n|\boldsymbol{z}_{n+1}) \\
&= \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n} \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1})p(\boldsymbol{z}_{n+1})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n|\boldsymbol{x}_{n+1},\boldsymbol{z}_{n+1}) \\
&= \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n} \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1})p(\boldsymbol{z}_{n+1})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n|\boldsymbol{z}_{n+1}) \\
&= \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1}) + \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n} \log p(\boldsymbol{z}_{n+1})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n|\boldsymbol{z}_{n+1}) \\
&= \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1}) + \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n} \log p(\boldsymbol{z}_{n+1})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n|\boldsymbol{z}_{n+1})
\end{aligned} \tag{30}$$

ここで，$P = p(\boldsymbol{z}_{n+1})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n|\boldsymbol{z}_{n+1})$ とおくと，

$$\begin{aligned}
P &= p(\boldsymbol{z}_{n+1})p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n|\boldsymbol{z}_{n+1}) \\
&= p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n,\boldsymbol{z}_{n+1}) \\
&= p(\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) \\
&= p(\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots|\boldsymbol{z}_{n+1},\boldsymbol{z}_n)p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) \\
&= p(\boldsymbol{z}_n)p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots|\boldsymbol{z}_n)p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) \\
&= p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n)p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n)
\end{aligned} \tag{31}$$

と変形できる．式 (31) を式 (30) に代入すると，

$$\begin{aligned}
\omega(\boldsymbol{z}_{n+1}) &= \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1}) + \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n} \log P \\
&= \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1}) + \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n} \log p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n)p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) \\
&= \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1}) + \max_{\boldsymbol{z}_n} \left\{ \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{n-1}} \log p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n)p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) \right\} \\
&= \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1}) + \max_{\boldsymbol{z}_n} \left\{ \log p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) + \max_{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_{n-1}} \log p(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n) \right\} \\
&= \log p(\boldsymbol{x}_{n+1}|\boldsymbol{z}_{n+1}) + \max_{\boldsymbol{z}_n} \{\log p(\boldsymbol{z}_{n+1}|\boldsymbol{z}_n) + \omega(\boldsymbol{z}_n)\}
\end{aligned} \tag{32}$$

となる．

式 (32) を $\omega(\boldsymbol{z}_1)$ から順に繰り返し計算することで $\omega(\boldsymbol{z}_N)$ を求めたら，$\boldsymbol{z}'_N = \arg\max_{\boldsymbol{z}_N} \omega(\boldsymbol{z}_N)$ を求め，このときの $\boldsymbol{z}'_N$ を出力する．つぎに，

$$\omega(\boldsymbol{z}'_N) = \log p(\boldsymbol{x}_N|\boldsymbol{z}'_N) + \max_{\boldsymbol{z}_{N-1}} \{\log p(\boldsymbol{z}'_N|\boldsymbol{z}_{N-1}) + \omega(\boldsymbol{z}_{N-1})\}$$

の式から，

$$\boldsymbol{z}'_{N-1} = \arg\max_{\boldsymbol{z}_{N-1}} \{\log p(\boldsymbol{z}'_N|\boldsymbol{z}_{N-1}) + \omega(\boldsymbol{z}_{N-1})\}$$

となるような $\boldsymbol{z}'_{N-1}$ を求めることができる．以下これを $\boldsymbol{z}'_1$ まで繰り返し求めることで，観測データ列が与えられたときの事後確率を最大化するような潜在変数の列 $\boldsymbol{z}'_1, \ldots, \boldsymbol{z}'_N$ を求めることができる．

また，計算の過程で任意の時点 i における最大事後確率 $p(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_i|X, \theta)$ を求めることが出来る．

6 その他

6.1 未解決

- 各アルゴリズムの計算量の見積もりの証明
- Baum-Welch の正規分布の場合の平均と分散の最大化の計算← 18/01/16 証明完了．
- $p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N|\boldsymbol{z}_n) = p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N)$ にはなぜならない？←おそらく解決．section 6.2.1 を参照．

- section 3.1 Baum-Welch
  期待値は全てのパラメータに対して上に凸な関数なのか？そうでなければ微分して 0 になるところで最大値をとることはいえない．

6.2 気づいたことメモ

6.2.1 独立変数，従属変数

$p(\boldsymbol{x}_n|\boldsymbol{z}_{n-1}, \boldsymbol{z}_n) = p(\boldsymbol{x}_n|\boldsymbol{z}_n)$ だが，$p(\boldsymbol{x}_n|\boldsymbol{z}_{n-1}) \neq p(\boldsymbol{x}_n)$

↑なぜ？

仮説．時間のあるときにきちんと計算したい．おそらく条件付き確率の定義から注意深く計算していけば証明出来る気がする

$\boldsymbol{z}_n$ が与えられたときは $\boldsymbol{x}_n$ は $\boldsymbol{z}_{n-1}$ に対して独立．$\boldsymbol{x}_n$ は $\boldsymbol{z}_n$ によってのみ決定されるから．ただし，$\boldsymbol{z}_n$ がないとき，つまり $p(\boldsymbol{x}_n|\boldsymbol{z}_{n-1})$ のときは $\boldsymbol{x}_n$ は $p(\boldsymbol{z}_{n-1})$ に対して従属．なぜならば $\boldsymbol{z}_n$ が $\boldsymbol{z}_{n-1}$ に対して従属であり， $\boldsymbol{x}_n$ が $\boldsymbol{z}_n$ に対して従属だから．

6.2.2 Forward-Backward

Forward-Backward は

- Viterbi と組み合わせて使われる場合と，
- Baum-Welch と組み合わせて使われる場合

の 2 種類がある．

前者ではまず Viterbi をやって，モデルの最大事後確率を求める．つぎにその $p(Z|X, \theta)$ を使って，各 $p(\boldsymbol{z}_n|X, \theta)$ を求める．この場合は $\xi(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n)$ は求めていない．

後者では Baum-Welch で EM する際に γ, ξ の具体的な値が必要になるから，そのときに Forward-Backward で計算する．

参考文献

[1] Pattern Recognition and Machine Learning, Christopher M. Bishop, Springer

[2] Biological sequence analysis, R.Durbin, CAMBRIDGE