

第13章 系列データ

修士 1年
村下 昇平

*Pattern
Recognition
and
Machine
Learning*

* もくじ

- 「系列データ」とは？
- マルコフモデル
- 隠れマルコフモデル(HMM)
 - EMアルゴリズムによる最尤推定
 - フォワード・バックワードアルゴリズム
 - Viterbiアルゴリズム
 - HMMの応用
- 連続潜在変数モデル
 - 線形ガウスモデル
 - 粒子フィルタ

0. 系列データ:「系列データ」とは？

- これまでは独立同時分布に従うと仮定するデータ点の集合について考えてきたが…
- 本章では系列データを取り扱う。
 - 金融予測、音声認識などに用いる時系列データ
 - 文章における文字の系列
 - …など、連続した観測値の間に相関が見られるデータのこと。
- こうした系列データ(=過去の観測値)を用いて次の値を予測することがこの章の目的！

0. 系列データ: 用いられるモデル

- ある時点での値を予測するときに、それまでの全てのデータを考慮に入れることは現実的でない。
- 過去のデータのなかでも、直近の観測に最も強く依存していると考えるのが自然。



- …というわけで、この章で扱うマルコフモデルでは、未来の予測値が直近の観測値のみに依存し、それ以外の過去の観測値に対しては独立であるという仮定のもとに立っている。
- さらにこれを発展させたものとして、状態空間モデルである隠れマルコフモデルおよび線形動的システムについて考察する。これによってより複雑なモデルを構成することができる。
- 以上のモデルはいずれもグラフィカルモデルの枠組みを用いて特徴付けられ、積和アルゴリズムを用いて推論が行われる。

1. マルコフモデル: 一次マルコフ連鎖

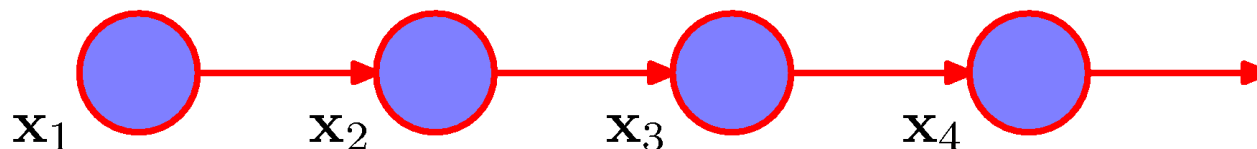
最も簡単なマルコフモデルとして、最も近い観測値のみに依存し、それ以外の過去の観測値から独立であると仮定したモデルが、以下の一次マルコフ連鎖である。

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N) = p(\mathbf{x}_1) \prod_{n=2}^N p(\mathbf{x}_n | \mathbf{x}_1, \dots, \mathbf{x}_{n-1}).$$

このような一次マルコフ連鎖においては遊向分離の性質から、時刻 n までの全ての観測値 \mathbf{x}_n の条件付き確率は

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N) = p(\mathbf{x}_1) \prod_{n=2}^N p(\mathbf{x}_n | \mathbf{x}_{n-1}).$$

とシンプルに表すことができる。こうしたマルコフ連鎖の応用では、ほとんどの場合 $p(\mathbf{x}_n | \mathbf{x}_{n-1})$ がみな同一であるという制約を課すが、そういったモデルは均一マルコフ連鎖として知られる。



1. マルコフモデル: 高次マルコフ連鎖

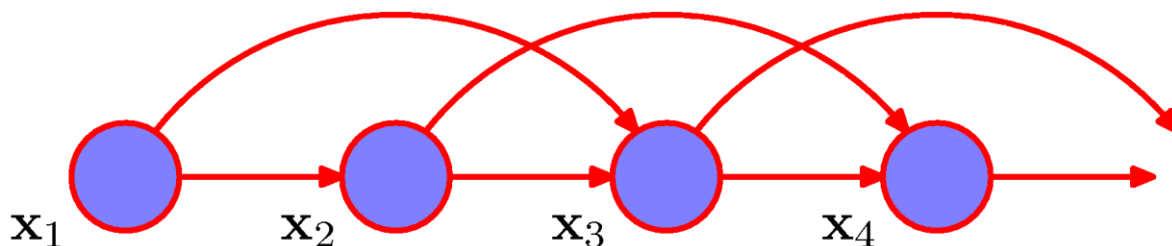
さらに一般性を求めるならば、次の値を予測するときに、いくつかの連続した観測データに見られる傾向を考慮したくなる…ここで出てくるのが高次マルコフ連鎖。

たとえば直前のさらに一つ前の値にも依存する、すなわち

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N) = p(\mathbf{x}_1)p(\mathbf{x}_2|\mathbf{x}_1) \prod_{n=3}^N p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{x}_{n-2}).$$

である場合、これを二次マルコフ連鎖と呼ぶ。

先ほどと同様有向分離を用いると結局、 \mathbf{x}_n は \mathbf{x}_1 から \mathbf{x}_{n-3} のすべての観測データから独立であることがわかる。



ただし、次数を上げると計算量が指数的に増大してしまう。

1. マルコフモデル: 状態空間モデル

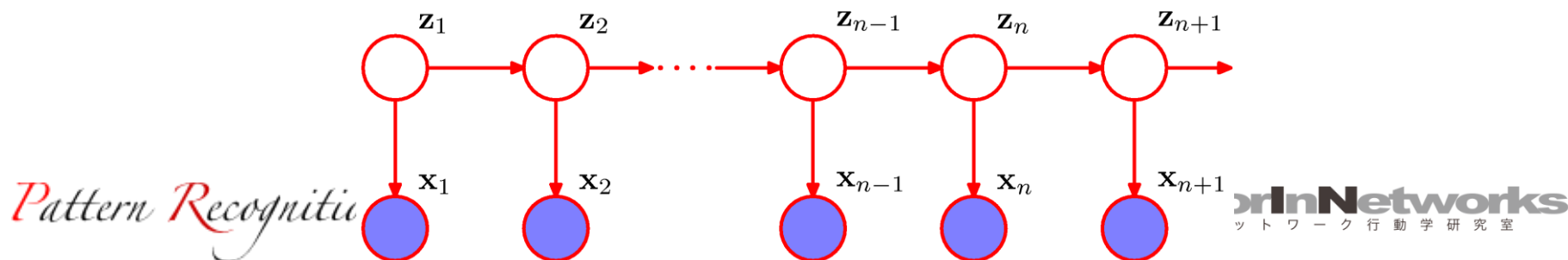
これまで出てきたようなマルコフモデルに対して、マルコフ連鎖を構成するような潜在変数を導入することで、比較的単純でより表現力の高いモデルを作ることができる。

これが状態空間モデルであり、このモデルの同時分布は以下で与えられる。

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{z}_1, \dots, \mathbf{z}_N) = p(\mathbf{z}_1) \left[\prod_{n=2}^N p(\mathbf{z}_n | \mathbf{z}_{n-1}) \right] \prod_{n=1}^N p(\mathbf{x}_n | \mathbf{z}_n).$$

ここで \mathbf{x}_n は観測値であり、 \mathbf{z}_n はそれに対応する潜在変数である。

このとき、 \mathbf{z}_n を与えたとき、 \mathbf{z}_{n-1} と \mathbf{z}_{n+1} が独立であるという重要な条件付き独立性を満たす。状態空間モデルでは、潜在変数を介して2つの観測変数 \mathbf{x}_n と \mathbf{x}_m をつなぐ経路は常に存在し、この経路は決して遮断されない(有向分離によって示される)。したがって、過去のすべての観測値を与えたときの、観測 \mathbf{x}_{n+1} の予測はすべての過去の観測値に依存する。しかしながら、観測変数はどの次数のマルコフ性も満たさない。つまり、表現力が高いモデルなのだ！



2. 隠れマルコフモデル

先の状態空間モデルのうち、潜在変数が離散変数であるとき
隠れマルコフモデル(HMM)を得る。
(ここで、観測変数は離散でも連続でもよい)

HMMは以下のように広く用いられている。

- 音声認識
- 自然言語モデル
- オンライン手書き文字認識
- タンパク質やDNAなどの生物学的配列の解析
- …などなど

2. HMM: 遷移確率行列

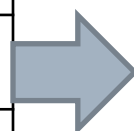
状態空間モデルであるので、潜在変数 z_n の確率分布は、条件付き分布 $p(z_n|z_{n-1})$ を通して直前の状態に依存する。

この潜在変数はK次元の二値変数なので、この条件付き分布は遷移確率を要素にもつ数表Aに対応する。ここで遷移確率 $A_{jk} \equiv p(z_{nk}=1 \mid z_{n-1,j}=1)$ すなわちAのjk要素は、 z_{n-1} での状態がjであったときに z_n での状態がkになる確率となる。

とうぜん、確率であるため $0 \leq A_{jk} \leq 1$ と $\sum_k A_{jk} = 1$ を満たす。

(例)今日の天気から明日の天気を予測

		明日		
		晴	雨	曇
今日	晴	$p[f f]=3/4$	$p[r f]=1/4$	$p[c f]=0$
	雨	$p[f r]=1/4$	$p[r r]=1/2$	$p[c r]=1/4$
	曇	$p[f c]=1/4$	$p[r c]=1/2$	$p[c c]=1/4$


$$\mathbf{A} = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

2. HMM: 一対K符号化法による条件付き分布の表現

条件付き分布は1対K符号化法によって以下の形で明示的に書ける。

$$p(z_n | z_{n-1}, \mathbf{A}) = \prod_{k=1}^K \prod_{j=1}^K A_{jk}^{z_{n-1,j} z_{nk}}$$

従って、例えば $z_{n-1}=(1,0,0)$ (昨日が晴)、 $z_n=(0,1,0)$ (今日が雨)である確率は

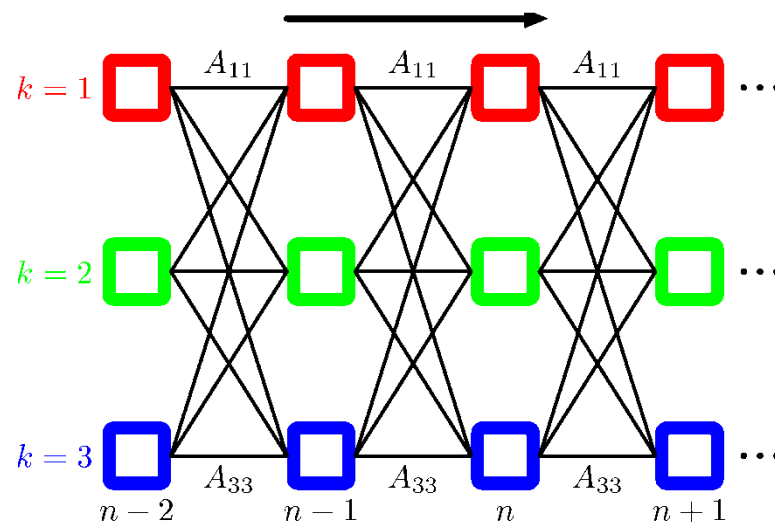
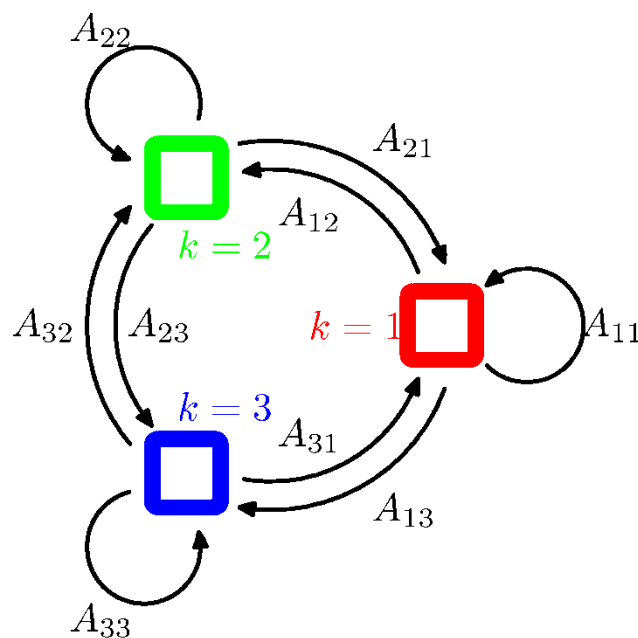
$$\begin{aligned} & p(z_n = (0,1,0) | z_{n-1} = (1,0,0)) \\ &= A_{11}^{z_{n-1,1} z_{n1}} A_{21}^{z_{n-1,2} z_{n1}} A_{31}^{z_{n-1,3} z_{n1}} A_{12}^{z_{n-1,1} z_{n2}} A_{22}^{z_{n-1,2} z_{n2}} A_{32}^{z_{n-1,3} z_{n2}} A_{13}^{z_{n-1,1} z_{n3}} A_{23}^{z_{n-1,2} z_{n3}} A_{33}^{z_{n-1,3} z_{n3}} \\ &= \left(\frac{3}{4}\right)^{1 \times 0} \left(\frac{1}{4}\right)^{0 \times 0} 0^{0 \times 1} \dots \end{aligned}$$

となり、結局求めたい $A_{12} = 1/4$ だけが残る。(指数の両方とも1であるものだけ！)

2. HMM: 状態遷移図と格子図

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

遷移行列は状態遷移図、あるいは格子図(遷移行列を時間的に展開したもの。トレリス図とも呼ばれる)によって図示することができる。



2. HMM: 出力確率

観測変数の条件付き確率分布 $p(x_n|z_n, \phi)$ を出力確率と呼ぶ。ここで ϕ はこの確率分布を支配するパラメータである。(以下の例での1/5やら2/3やらのこと！)

x_n は観測されるので、 ϕ の値が与えられたとき、分布 $p(x_n|z_n, \phi)$ は二値のベクトル z_n のK個の可能な状態に対応した、K個の要素をもつベクトルからなる。

(例)その日の行動から天気を予測

		観測変数		
		買い物	散歩	掃除
潜在変数	$z_n = \text{晴}$	$p(x=s z=f)=2/5$	$p(x=w z=f)=2/5$	$p(x=c z=f)=1/5$
	$z_n = \text{雨}$	$p(x=s z=r)=1/3$	$p(x=w z=r)=0$	$p(x=c z=r)=2/3$
	$z_n = \text{曇}$	$p(x=s z=c)=2/5$	$p(x=w z=c)=1/5$	$p(x=c z=c)=2/5$

「観測される」ということは、観測変数がいずれかに固定されるということなので、例えば「散歩」が観測された場合、 $p(x=\text{散歩})=(2/5, 0, 1/5)$ と、確かに3つの要素をもつベクトルになっている。

2. HMM: 一対K符号化法による出力分布の表現

先ほどの潜在変数についての条件付き分布の時と同様に、出力確率は以下の形でかける。

$$p(x_n | z_n, \phi) = \prod_{k=1}^K p(\mathbf{x}_n | \phi_k)^{z_{nk}}$$

従って先ほどの例、すなわち $z_n=(0,1,0)$ (雨), $x_n=(0,1,0)$ (散歩)の場合は

$$\begin{aligned} p(x_n = (0,1,0) | z_n = (0,1,0)) \\ &= p(\mathbf{x}_n | \phi_1)^{z_{n1}} p(\mathbf{x}_n | \phi_2)^{z_{n2}} p(\mathbf{x}_n | \phi_2)^{z_{n2}} \\ &= p(\mathbf{x}_n | \phi_1)^0 p(\mathbf{x}_n | \phi_2)^1 p(\mathbf{x}_n | \phi_2)^0 \end{aligned}$$

となり、結局求めたい $p=0$ だけが残る。

…1対K符号化法は便利だね！

2. HMM: 潜在/観測変数の同時確率分布

$$p(z_n | z_{n-1}, \mathbf{A}) = \prod_{k=1}^K \prod_{j=1}^K A_{jk}^{z_{n-1,j} z_{nk}} \quad p(x_n | z_n, \phi) = \prod_{k=1}^K p(\mathbf{x}_n | \phi_k)^{z_{nk}}$$

これらの式より、HMMにおける潜在変数と観測変数の同時確率分布は以下のようにかける。

$$p(\mathbf{X}, \mathbf{Z} | \theta) = p(\mathbf{z}_1 | \pi) \left[\prod_{n=2}^N p(\mathbf{z}_n | \mathbf{z}_{n-1}, \mathbf{A}) \right] \prod_{m=1}^N p(\mathbf{x}_m | \mathbf{z}_m, \phi).$$

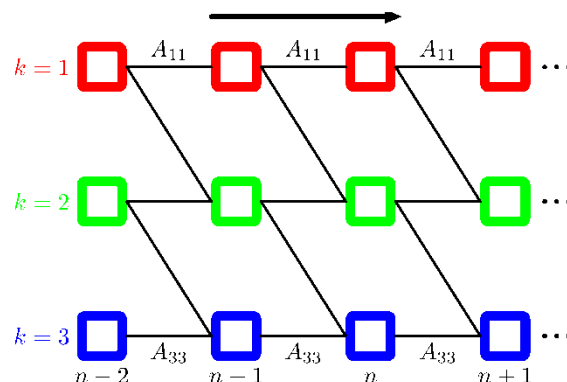
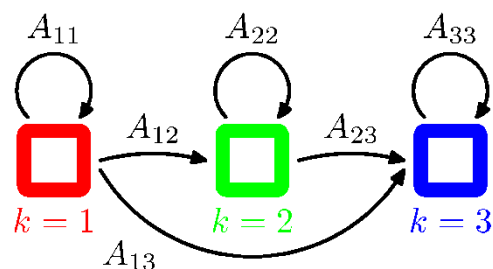
ここで、 \mathbf{X} は各時点における観測変数の、 \mathbf{Z} は各時点における潜在変数の集合であり、 θ は「最初の遷移確率(親ノードを持たないため \mathbf{A} と区別される) π 」、「状態遷移行列 \mathbf{A} 」、「出力確率のパラメータ ϕ 」といったパラメータ集合である。

2. HMM: left-to-right HMM

遷移行列 A の $k < j$ となる A_{jk} 成分をゼロ、すなわち

$$A = \begin{pmatrix} A_{11} & \cdots & \cdots & A_{1k} \\ & \ddots & & \vdots \\ 0 & & \ddots & \vdots \\ 0 & 0 & & A_{kk} \end{pmatrix} \quad \text{とすることで、left-to-right HMM の遷移行列を得る。}$$

これは「次(あるいはそれ以降)の状態には遷移しうる」が「以前の状態には戻らない」ようなマルコフ連鎖を生成する。状態遷移図および格子図は以下のように描ける。(特に、格子図のほうは留まるか直近の次の状態に移るかの二択しかないものを表す)



このようなHMMは音声認識やオンライン文字認識など多くの分野で応用されている重要なモデルである。

2.1. HMMの最尤推定

というわけで、HMMについて最尤推定によってパラメータ $\theta = \{\pi, A, \phi\}$ を決定したい。

まず、HMMについての同時分布の式より、これを潜在変数について周辺化した以下の尤度関数が得られる。(なんで周辺化するのはよくわかんない)

$$p(\mathbf{X}|\theta) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\theta).$$

この尤度関数の最大化には様々な困難がつきまとう(詳細は下巻p.133を参照)。以下ではこれを効率的に扱う手法としてEMアルゴリズムを考えることにする。

2.1 HMMの最尤推定: EMアルゴリズム

EMアルゴリズムでは以下のステップで尤度関数を最大化する。第9章でやったのと基本的には同じだと考えてよい。

0. 最初にモデルパラメータをある初期集合に設定する。(これを θ^{old} とする)

1. Eステップ

まず潜在変数 $p(Z|X, \theta^{\text{old}})$ の事後分布を求め、この事後分布を用いて、パラメータ集合 θ の関数としての、完全データに対する尤度関数の対数の期待値(次式)を求める。

$$Q(\theta, \theta^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \theta^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\theta).$$

ここで潜在変数 z_n の周辺事後分布を γ 、2つの連続した潜在変数に対する同時事後分布を ξ とする、すなわち以下のように定めると…

$$\begin{aligned}\gamma(z_n) &= p(z_n|\mathbf{X}, \theta^{\text{old}}) \\ \xi(z_{n-1}, z_n) &= p(z_{n-1}, z_n|\mathbf{X}, \theta^{\text{old}}).\end{aligned}$$

どちらも確率であるから総和は1となり、また潜在変数の有り得る状態の個数は K であるから、 γ は K 個の要素を持つベクトル、 ξ は $K \times K$ 行列となる。

2.1 HMMの最尤推定: EMアルゴリズム

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n | \mathbf{X}, \theta^{\text{old}})$$

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n | \mathbf{X}, \theta^{\text{old}}).$$

これらは結局

◆ $\gamma(\mathbf{z}_n)$ のk番目の要素 z_{nk} は \mathbf{z}_n の状態がk(すなわち $z_{nk}=1$)となる条件付き確率

◆ $\xi(\mathbf{z}_{n-1}, \mathbf{z}_n)$ のjk要素 $\xi(z_{n-1,j}, z_{nk})$ は、 \mathbf{z}_{n-1} がj(すなわち $z_{n-1,j}=1$)であったときに \mathbf{z}_n がk(すなわち $z_{nk}=1$)となる確率であるといえる。

二値の確率変数である z_n の期待値を考えると、単にそれが値1をもつ確率がそのまま出てくるため、以下の式が得られる。

$$\gamma(z_{nk}) = \mathbb{E}[z_{nk}] = \sum_{\mathbf{z}_n} \gamma(\mathbf{z}_n) z_{nk}$$

$$\xi(z_{n-1,j}, z_{nk}) = \mathbb{E}[z_{n-1,j} z_{nk}] = \sum_{\mathbf{z}_{n-1}, \mathbf{z}_n} \xi(\mathbf{z}_{n-1}, \mathbf{z}_n) z_{n-1,j} z_{nk}.$$

z_n が1対K符号化されているんだから、確かにそうなるよな…

2.1 HMMの最尤推定: EMアルゴリズム

$$p(\mathbf{X}, \mathbf{Z}|\theta) = p(\mathbf{z}_1|\pi) \left[\prod_{n=2}^N p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) \right] \prod_{m=1}^N p(\mathbf{x}_m|\mathbf{z}_m, \phi). \quad Q(\theta, \theta^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \theta^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\theta).$$

ちょっと前に出てきた隠れマルコフモデルにおける \mathbf{X} と \mathbf{Z} の同時分布(上左式)を Q (上右式)に代入し、先ほどの γ , ξ を代入することで次式を得る。

$$\begin{aligned} Q(\theta, \theta^{\text{old}}) = & \sum_{k=1}^K \gamma(z_{1k}) \ln \pi_k + \sum_{n=2}^N \sum_{j=1}^K \sum_{k=1}^K \xi(z_{n-1,j}, z_{nk}) \ln A_{jk} \\ & + \sum_{n=1}^N \sum_{k=1}^K \gamma(z_{nk}) \ln p(\mathbf{x}_n|\phi_k). \end{aligned}$$

以上でEステップが終了する。

2.1 HMMの最尤推定: EMアルゴリズム

2. Mステップ

次のMステップでは先ほど求めた $\gamma(z_n)$ と $\xi(z_n)$ を定数とみなし、パラメータ $\theta = [\pi, A, \phi]$ に関して $Q(\theta, \theta^{\text{old}})$ を最大化する。

ここで、遷移確率についてのパラメータである π と A に関する最大化は適当なラグランジュ乗数を用いることにより簡単に求められ、以下のようになる。

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{j=1}^K \gamma(z_{1j})}$$
$$A_{jk} = \frac{\sum_{n=2}^N \xi(z_{n-1,j}, z_{nk})}{\sum_{l=1}^K \sum_{n=2}^N \xi(z_{n-1,j}, z_{nl})}.$$

2.1 HMMの最尤推定: EMアルゴリズム

出力分布についてのパラメータ ϕ を求める場合、例えば、出力分布がガウス密度分布のときには、 $p(x|\phi)=N(x|\mu_k, \Sigma_k)$ となり、 $Q(\theta, \theta^{\text{old}})$ の最大化により以下を得る。

$$\mu_k = \frac{\sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n}{\sum_{n=1}^N \gamma(z_{nk})}$$
$$\Sigma_k = \frac{\sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T}{\sum_{n=1}^N \gamma(z_{nk})}.$$

2.2. フォワード-バックワードアルゴリズム

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n | \mathbf{X}, \theta^{\text{old}})$$

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n | \mathbf{X}, \theta^{\text{old}}).$$

以下ではフォワード-バックワードアルゴリズムと呼ばれる、EMアルゴリズムのEステップにあたる上式を効率的に求める方法について議論する。

まず、有向分離を用いれば以下の条件付き独立が証明できる。

$$p(\mathbf{X} | \mathbf{z}_n) = p(\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{z}_n) \times p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N | \mathbf{z}_n) \quad (13.24)$$

$$p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1} | \mathbf{x}_n, \mathbf{z}_n) = p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1} | \mathbf{z}_n) \quad (13.25)$$

$$p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1} | \mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1} | \mathbf{z}_{n-1}) \quad (13.26)$$

$$p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N | \mathbf{z}_n, \mathbf{z}_{n+1}) = p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N | \mathbf{z}_{n+1}) \quad (13.27)$$

$$p(\mathbf{x}_{n+2}, \dots, \mathbf{x}_N | \mathbf{z}_{n+1}, \mathbf{x}_{n+1}) = p(\mathbf{x}_{n+2}, \dots, \mathbf{x}_N | \mathbf{z}_{n+1}) \quad (13.28)$$

$$p(\mathbf{X} | \mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1} | \mathbf{z}_{n-1}) \times p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N | \mathbf{z}_n) \quad (13.29)$$

$$p(\mathbf{x}_{N+1} | \mathbf{X}, \mathbf{z}_{N+1}) = p(\mathbf{x}_{N+1} | \mathbf{z}_{N+1}) \quad (13.30)$$

$$p(\mathbf{z}_{N+1} | \mathbf{z}_N, \mathbf{X}) = p(\mathbf{z}_{N+1} | \mathbf{z}_N). \quad (13.31)$$

2.2. フォワード-バックワードアルゴリズム

- 以下の変形では次の2つを頭にたたき込んでおきましょう。
- 1. 確率の乗法定理
 - $P\{A,B\} = P\{A\} \times P\{B|A\} = P\{B\} \times P\{A|B\}$
 - 特にAとBがCについて条件付き独立であるとき $P\{A,B|C\} = P\{A|C\} \times P\{B|C\}$
- 2. 同時確率のある変数についての総和をとると周辺確率となる。
 - たとえば $p(z_n) = \sum_{z_{n-1}} p\{z_{n-1}, z_n\}$ とか。

$$p(a | b, c) = p(a | c) \cdots (8.20)$$

2.2. フォワード-バックワードアルゴリズム

まずベイズの定理より次式を得る。

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

これを先ほどの式を用いて変形すると、 $p(\mathbf{X}|\mathbf{z}_n)=p(x_1\dots x_n|\mathbf{z}_n)p(x_{n+1}\dots x_N|\mathbf{z}_n)$ の前半部分と $p(\mathbf{z}_n)$ の積は確率の乗法定理より $p(x_1\dots x_n, \mathbf{z}_n)$ という同時確率となり…

$$\gamma(\mathbf{z}_n) = \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{z}_n)p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N|\mathbf{z}_n)}{p(\mathbf{X})} = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(\mathbf{X})}$$

を得る。ただしここで

$$\alpha(\mathbf{z}_n) \equiv p(\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{z}_n)$$

$$\beta(\mathbf{z}_n) \equiv p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N|\mathbf{z}_n)$$

と定義した。 α は時刻 n までの過去全ての観測データと \mathbf{z}_n の同時確率、 β は \mathbf{z}_n が与えられたときの時刻 $n+1$ から N までの全ての未来のデータの条件付き確率となっている。

2.2. フォワード-バックワードアルゴリズム: フォワード α 再帰

α について考える。目標は α を n に関する再帰式として計算量を N について線形にすることである

まず確率の乗法定理を用いて出力分布 $p(x_n|z_n)$ をくり出したい。以下のように変形すると...

$$\begin{aligned}\alpha(z_n) &= p(x_1, \dots, x_n, z_n) \\ &= p(x_1, \dots, x_n | z_n) p(z_n)\end{aligned}$$

$\{x_1 \dots x_{n-1}\}$ と x_n とは z_n について条件付き独立である(13.25)から、以下のようにくり出せる。

$$= p(x_n | z_n) p(x_1, \dots, x_{n-1} | z_n) p(z_n)$$

ここで、 $\{x_1 \dots x_n, z_n\}$ について周辺化させる。

$$= p(x_n | z_n) p(x_1, \dots, x_{n-1}, z_n)$$

こやることで次式を得

$$= p(x_n | z_n) \sum_{z_{n-1}} p(x_1, \dots, x_{n-1}, z_{n-1}, z_n)$$

2.2. フォワード-バックワードアルゴリズム: フォワード α 再帰

さらに確率の乗法定理を用いて遷移確率をくり出す。ここで $\{x_1 \dots x_n\}$ と z_n とが z_{n-1} について条件付き独立(13.26)であることに注意すると、以下のように変形できて…

$$\begin{aligned} &= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}, \mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}) \\ &= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1} | \mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}) \\ &= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}, \mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1}). \end{aligned}$$

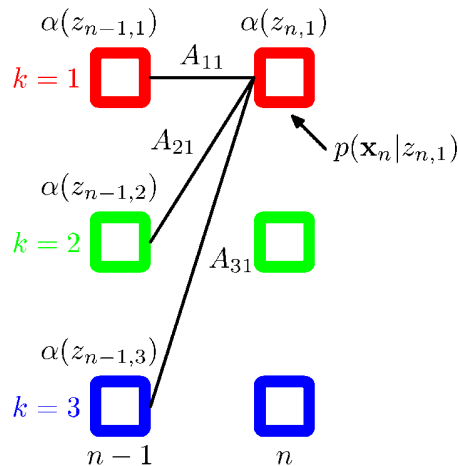
$p(\mathbf{x}_1 \dots \mathbf{x}_{n-1}, \mathbf{z}_{n-1})$ が $\alpha(\mathbf{z}_{n-1})$ であることから結局 α は以下の再帰式で表される。

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} \alpha(\mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1})$$

2.2. フォワード-バックワードアルゴリズム: フォワード α 再帰

$$\alpha(\mathbf{z}_n) = p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} \alpha(\mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1})$$

式変形はまあ、置いといて…この再帰式は以下の格子図のように解釈できる。



($n-1$)ステップでの $\alpha(\mathbf{z}_{n-1})$ の要素 $\alpha(\mathbf{z}_{n-1,j})$ について、 $p(\mathbf{z}_n | \mathbf{z}_{n-1})$ の値に相当する A_{j1} で与えられる重みを用いて、重み付け和をとり、さらにそれにデータの寄与 $p(\mathbf{x}_n | \mathbf{z}_{n+1})$ をかけることで、 $\alpha(\mathbf{z}_{n,1})$ を得る。

この回帰式に対して以下の初期条件を与え、順次計算していくことによってすべての潜在ノードの α を求めることができる。

$$\alpha(\mathbf{z}_1) = p(\mathbf{x}_1, \mathbf{z}_1) = p(\mathbf{z}_1) p(\mathbf{x}_1 | \mathbf{z}_1) = \prod_{k=1}^K \{\pi_k p(\mathbf{x}_1 | \phi_k)\}^{z_{1k}}.$$

2.2. フォワード-バックワードアルゴリズム: バックワードβ再帰

つぎに β について考える。

ここで、先ほどと同様に周辺化された z_{n+1} を明示的に総和の形で表してやると、

$$\begin{aligned}\beta(\mathbf{z}_n) &= p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N | \mathbf{z}_n) \\ &= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N, \mathbf{z}_{n+1} | \mathbf{z}_n)\end{aligned}$$

さらに確率の乗法定理より以下のように遷移確率をくり出せる(両方ともに z_n が条件として入っていることに注意！)

$$= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N | \mathbf{z}_n, \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n)$$

$\{\mathbf{x}_{n+1} \dots \mathbf{x}_N\}$ と \mathbf{z}_n とが \mathbf{z}_{n+1} について条件付き独立(13.27)であることより

$$= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N | \mathbf{z}_{n+1}) p(\mathbf{z}_{n+1} | \mathbf{z}_n)$$

2.2. フォワード-バックワードアルゴリズム: バックワードβ再帰

さらに $\{x_{n+2} \dots x_N\}$ と x_{n+1} とが z_{n+1} について条件付き独立(13.28)であることより出力確率をくり出すことができる

$$= \sum_{z_{n+1}} p(x_{n+2}, \dots, x_N | z_{n+1}) p(x_{n+1} | z_{n+1}) p(z_{n+1} | z_n).$$

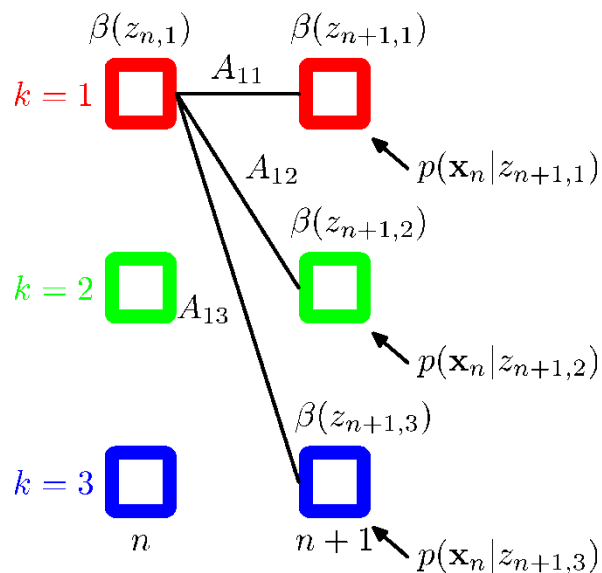
$p(x_{n+2} \dots x_N | z_{n+1})$ が $\beta(z_{n+1})$ であることから結局 β も以下の再帰式で表される。

$$\beta(z_n) = \sum_{z_{n+1}} \beta(z_{n+1}) p(x_{n+1} | z_{n+1}) p(z_{n+1} | z_n).$$

2.2. フォワード-バックワードアルゴリズム: バックワードβ再帰

$$\beta(z_n) = \sum_{z_{n+1}} \beta(z_{n+1}) p(x_{n+1}|z_{n+1}) p(z_{n+1}|z_n).$$

この再帰式は以下の格子図で表すことができ、 $\beta(z_{n+1})$ を用いて $\beta(z_n)$ を求める後ろ向きアルゴリズムとなっている。各ステップで x_{n+1} の観測の影響を出力確率 $p(x_{n+1}|z_{n+1})$ で吸収し、遷移行列 $p(z_{n+1}|z_n)$ を掛け、そして z_{n+1} について周辺化している。



($n+1$)ステップの $\beta(z_{n+1})$ の要素 $\beta(z_{n+1,k})$ についての重み付け和をとることで、 $\beta(z_n,1)$ を得る。ここで、各 $\beta(z_{n+1,k})$ に対応する重みは、 $p(z_{n+1}|z_n)$ の値に対応する出力密度 $p(x_n|z_{n+1,k})$ の値を乗じたものである。

また、 β に関する初期条件は次式で与えられる。

$$p(z_N|X) = \frac{p(X, z_N) \beta(z_N)}{p(X)}.$$

これで γ が求まる。

2.2. フォワード-バックワードアルゴリズム: $\xi(\mathbf{z}_{n-1}, \mathbf{z}_n)$ の求め方

ちなみに ξ についても、これまでの α 再帰と β 再帰で得られた結果をそのまま用いることができる。まずベイズの定理によって次のように変形する。

$$\begin{aligned}\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) &= p(\mathbf{z}_{n-1}, \mathbf{z}_n | \mathbf{X}) \\ &= \frac{p(\mathbf{X} | \mathbf{z}_{n-1}, \mathbf{z}_n) p(\mathbf{z}_{n-1}, \mathbf{z}_n)}{p(\mathbf{X})}\end{aligned}$$

(13.29)を用いると次式を得る。

$$= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_{n-1} | \mathbf{z}_{n-1}) p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{x}_{n+1}, \dots, \mathbf{x}_N | \mathbf{z}_n) p(\mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1})}{p(\mathbf{X})}$$

ここに現れた α および β の定義式の部分を置き換えると結局 ξ は次式で表される。

$$= \frac{\alpha(\mathbf{z}_{n-1}) p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{z}_n | \mathbf{z}_{n-1}) \beta(\mathbf{z}_n)}{p(\mathbf{X})}.$$

以上のようにして α 再帰および β 再帰の結果を用いて直接 ξ を求めることができ、Eステップが終了する。

2.2. フォワード-バックワードアルゴリズム: Mステップ

…まあ、Mステップに関しては、以前示した式の値を求めるだけの話。

2.2. HMMの最尤推定まとめ

- まずパラメータの初期値 θ^{old} の値を定める。
 - ここで $\theta = \{\pi, A, \phi\}$ である。
- Eステップ
 - フォワード α 再帰およびバックワード β 再帰を行い、その結果を用いて γ, ξ を求める。
 - この段階では γ および ξ が得られるため、尤度関数 Q も求めることができる。
- Mステップ
 - すでに説明したとおり、 γ と ξ を代入するだけで θ が求まる。
- ※ちなみに、先ほど条件つき独立から求められた α - β 再帰式は、積和アルゴリズムによってもっと簡単に求められるらしい。詳細は13.2.3を参照。
- 実際のフォワード・バックワードアルゴリズムの利用の際には、適切なスケーリング係数を適用する必要がある (α や β はむちゃくちゃ小さな値となるから)。詳細は13.2.4を参照。

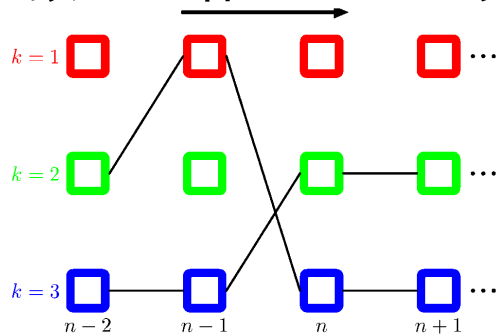
2.5 Viterbiアルゴリズム

- 隠れマルコフモデルにおける潜在変数は通常何らかの意味をもっているはず。
- …ってことは、その隠れ状態の最も確からしい系列を求めたくなりますよね。というかむしろそのためにやってたりすることも多い。
- ただし「個々の潜在変数に対して最も確からしい状態(= γ が最大)の集合を求める」と「潜在状態の最も確からしい系列を求める」こととはぜんぜんちがう。
- …そこで出てくるのがViterbiアルゴリズム。

2.5 Viterbiアルゴリズム

■ Viterbiアルゴリズムは…

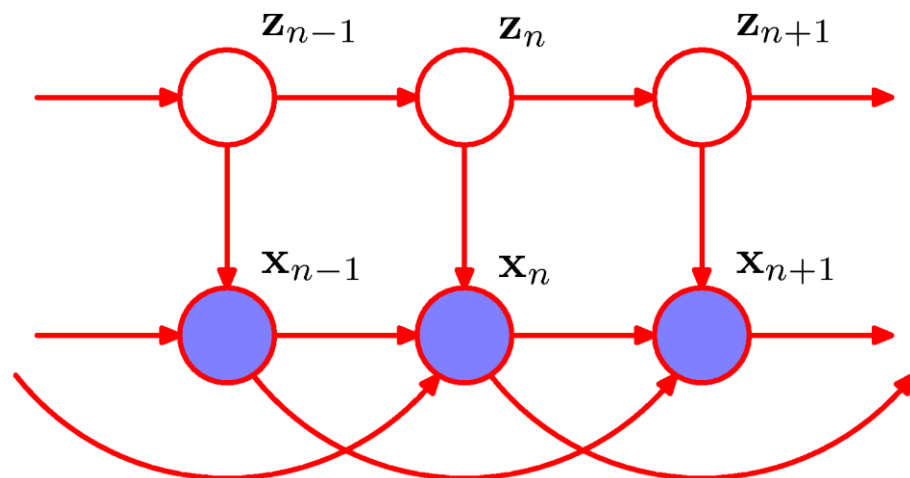
- 単純にやってしまうと N の大きさに従ってあり得る経路の数は指数的に増加してしまう。



- …が、Viterbiアルゴリズムを使うと、 N に対してただか線形に増加する計算量で最も確からしい系列を見つけることができる！
- イメージとしてはダイクストラ法に近いっぽい。
- max-sumアルゴリズムとして定式化されてるが、積和アルゴリズムの知識が必要そうなので省略。

2.6. HMMの拡張

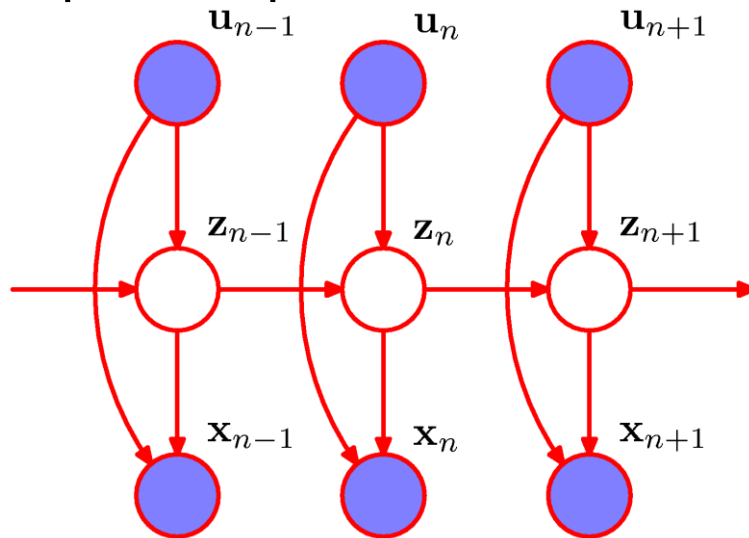
■ 自己回帰隠れマルコフモデル



- たくさんの時刻ステップだけ離れている観測変数間の相関をとりたいときに利用。
- 出力確率について、潜在変数だけでなく以前の観測変数(の部分集合)も条件として付け加える。

2.6. HMMの拡張

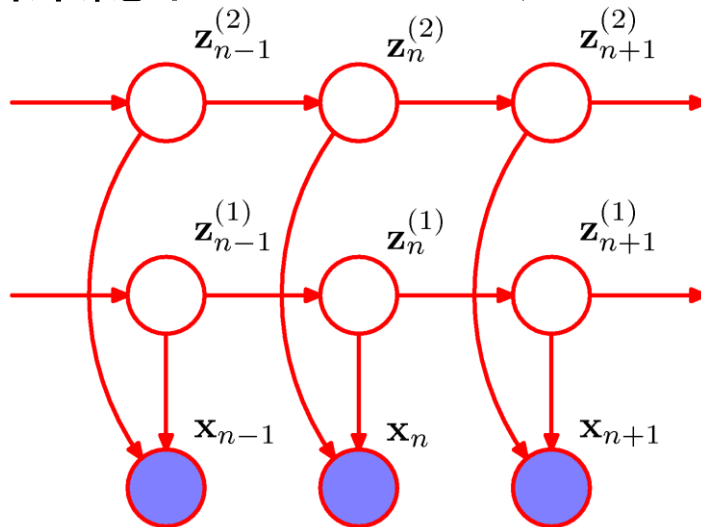
■ input-output隠れマルコフモデル



- HMMの枠組みを系列データに対する教師有り学習の領域まで拡張する。らしい。
- これまでの出力変数に加えて、潜在変数や出力変数に影響を与えるような新たな観測変数を考えるモデル。

2.6. HMMの拡張

■ 階乗隠れマルコフモデル



- お互いに独立な、複数の潜在変数のマルコフ連鎖があり、与えられた時刻のステップの観測変数の分布が同じ時刻ステップにおけるすべての潜在変数の状態に依存するようなモデル。
- 潜在状態をかなり少なく抑えることができるが、学習が複雑になるという欠点も持つ。

3. 線形動的システム

■ 線形動的システムとは？

- 状態空間モデルのうち、潜在変数と観測変数の両方が連続変数、特にガウス分布に従う、というモデル。12章の連続潜在変数モデルの一般化と見ることができる。
- 隠れマルコフモデルとの重要なちがいは、潜在変数が離散か連続(ガウス分布)か、ということ。

■ 直感的な説明。

- 時間軸上で変化している未知の量 $\{z_1 \dots z_N\}$ の値(位置とか)を、平均がゼロのガウスノイズを載せた観測値 $\{x_1 \dots x_N\}$ で返すセンサ(GPSとか)で計測したいとき、 z_n の値の推定のために、いちばん最近の数回の観測値を平均するとよいように思える…！
- もしセンサのノイズレベルが高い場合には、平均をとるときに比較的長い観測窓幅をとったほうがよいだろうし、ノイズレベルが低い場合には観測値を比較的そのまま使うほうがよいだろうし…というあたりを定式化したモデル。

3. 線形動的システム

遷移確率分布と出力確率分布はどちらもガウス分布であり、以下の一般的な形に書くことができる。

$$p(\mathbf{z}_n | \mathbf{z}_{n-1}) = \mathcal{N}(\mathbf{z}_n | \mathbf{A}\mathbf{z}_{n-1}, \mathbf{\Gamma})$$

$$p(\mathbf{x}_n | \mathbf{z}_n) = \mathcal{N}(\mathbf{x}_n | \mathbf{C}\mathbf{z}_n, \mathbf{\Sigma}).$$

このモデルパラメータ推定に対してはEMアルゴリズムを使うのが一般的である。

このEステップについては、線形動的システムにおけるフォワード-バックワード再帰のような効率的な推論アルゴリズムが積和アルゴリズムより導かれ、以下のように呼ばれる。

- フォワード再帰についてはカルマンフィルタ方程式
- バックワード再帰についてはカルマンスムーザ方程式あるいはRauch-Tung-Striebel (RTS) 方程式

くわしいメッセージの形については省略。

3.4. 粒子フィルタ

- もちろん線形ガウスでない動的システムも考えられる。これは、以下のような状況下で効果を発揮する。
 - 時系列に急激な構造変化が見られるとき
 - 異常値が存在するとき
 - 分布に非対称性や非正規性があるとき
 - 非線形システム
- このような非線形・非ガウスの状態空間モデルに対する扱いやすい推論アルゴリズムとしてサンプリング手法による定式化を用いたものが粒子フィルタ(ブートストラップフィルタ、モンテカルロフィルタ等とも呼ばれる)。