Subject 20H_Q3 (April 13th) Date /

Attendants MRF and Sampling - conditional Ising model

$x \in \mathbb{R}^d$, $\underline{y \{0,1\}^d}$

↳ decision vector,
each y_i is the decision for pixel i

$p(y|x)$ conditional distribution
using Markov random field (MRF)

with possibly a grid-like undirected graph G

Conditional Ising Model

$$p(y|x) = \frac{1}{Z(x)} \exp\left(-\sum_i y_i \theta_i(x) - \sum_{i,j \in N_i} y_i y_j \theta_{ij}(x) \right)$$

Subject MRF　　Date 2

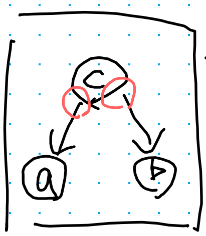
Attendants PRML #8

Bayesian Network: Directed Graphical Model

Markov Random Field: Undirected Graphical Model

8.1 Baysian Network

8.2 条件付き独立性



tail to tail

$$p(a,b) = p(a)p(b)$$

$$\Leftrightarrow a \perp\!\!\!\perp b \mid \emptyset$$

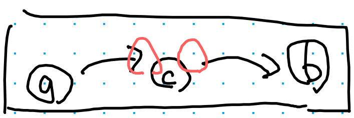
cで周辺化
↓
独立じゃない

$$p(a,b|c) = p(a|c)p(b|c) = \frac{p(a,b,c)}{p(c)}$$

$$\Leftrightarrow a \perp\!\!\!\perp b \mid c$$

cで条件付け
↓
独立
(Block)

head to tail

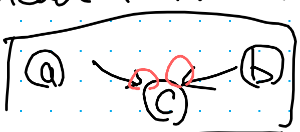


$$p(a,b,c) = p(a)p(c|a)p(b|c)$$

$$\Leftrightarrow a \perp\!\!\!\perp b \mid c$$

条件付き独立
(Block)

head to head



$$p(a,b,c) = p(a)p(b)p(c|a,b)$$

$$\Leftrightarrow a \perp\!\!\!\perp b \mid \emptyset$$

条件付き独立ではなくCで周辺化したら独立じゃない

8.3 マルコフ確率場(無向グラフィカルモデル)

Subject MRF PreDoc2 2014_Q3 Date 3

Attendants (PRML #8 DL Book #16)

$$p(y|x) = \frac{1}{Z(x)} \exp\left(-\sum_i y_i \theta_i(x) - \sum_i \sum_{j \in N_i} y_i y_j \theta_{ij}(x)\right)$$

$\theta_i(x)$ could be the output of NN (depending on i)

How to approximate inference

1 Why is not possible in general to exactly compute the marginal $p(y_i|x)$?

$$Z(x) = \sum_y p(x, y)$$

$$= \sum_{y_1} \sum_{y_2} \cdots \sum_{y_d} p(x, y)$$

Per sum of $\sum_{y_i}$, $y_i \in \{0, 1\}$ has two pattern then overall of pattern are $2^d = \underline{\underline{2^{8k}}}$

we need to compute sum of 2^{8k} term, it is not feasible

Subject 2014_Q3 MRF   Date 4

Attendants

[2] Let Y be random vector w/ values $y \in \{0,1\}^d$;

X w/ $x \in \mathbb{R}^d$. The MRF makes some conditional independence assumptions on Y.

In particular, consider the following conditional independence statement:

$$Y_i \perp\!\!\!\perp Y_j \mid Y_{S_i}, X$$

$i \neq j$ for several possible set S_i

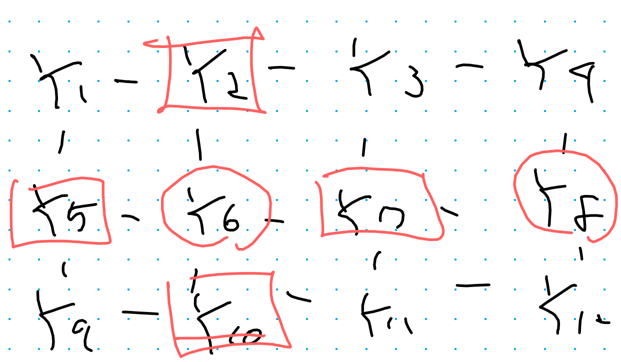
What is the smallest set S_i?

$$Y_i \perp\!\!\!\perp Y_j \mid Y_{S_i}$$

necessity and sufficient condition.

It requires blocks between Y_i and Y_j such as $S_i = N_i$

e.g. Y_6 Y_8



Y_2 Y_5 Y_7 Y_{10} is set of N_i

Subject 20H_Q3 MRF Date 5

Attendants Gibbs Sampling

Approximate the pair marginal $p(y_1, y_2 | x)$

suppose we initialize the Markov chain w/

$y^{(0)} = (0 \cdots 0)$

Describe how you will generate future samples

of Markov chain by using Gibbs Sampling

(w/ cyclic traversal through the graph)

Gives the simplest formula for the sampling

distribution. (as a function of the previous sample)

Finally, give a formula to approximate

$p(y_1, y_2 | x)$ by using the sample

from Markov chain.

Subject 20H Q3 MRF   Date 6

Attendants Gibbs Sampling は

1つづつ固定して Sampling

$\{z_i : i = 1, \dots M\}$ を考える

$t=0$ から T まで以下を行う

$$z_1' \sim p(z_1 \mid z_2, z_3 \cdots z_M)$$

$$z_2' \sim p(z_2 \mid z_1', z_3 \cdots z_M)$$

$$\vdots$$

$$z_M' \sim p(z_M \mid z_1', z_2' \cdots z_{M-1})$$

ある程度収束するまで

他の変数を固定したときの条件付き~~確~~分布

$$p(z_i \mid z_{-i})$$

$$(\text{where } z_{-i} = z_1 \cdots z_{i-1}, z_{i+1} \dots z_M)$$

Subject 20H_Q3 Gibbs Sampling Date 7

Attendants

$$p(y_i \mid y_{-i}, x) = \frac{p(y \mid x)}{p(y_{-i} \mid x)}$$

$$\propto \frac{\frac{1}{Z(x)} \exp\left(-\sum_i y_i \theta_i(x) - \sum_i \sum_{j \in N_i} y_i y_j \theta_{ij}(x)\right)}{\frac{1}{Z(x)} \exp\left(-\sum_{k \neq i} y_i \theta_i(x) - \sum_{k \neq i} \sum_{j \in N_i} y_i y_j \theta_{ij}(x)\right)}$$

↓ remain term depends on i

$$\propto \exp\left(y_i \theta_i(x) - \sum_{j \in N_i} y_i y_j \theta_{ij}(x)\right)$$

simplest formula for the sampling distribution

Subject 20H_Q3 Gibbs Sampling Date 8

Attendants

1. let $y^{(0)} = \underbrace{(0 \dots 0)}_{d}$

2. loop $t = 1$ to $t = T$

$y_1^{(t)} \sim P(y_1 \mid y_{-1}^{(t-1)}, x)$

where $y_{-1}^{(t-1)} = \underbrace{(0, \dots\dots 0)}_{d-1 \text{ elements}}$

$y_2^{(t)} \sim P(y_2 \mid y_{-2}^{(t-1)}, x)$

where $y_{-2}^{(t-1)} = (y_1^{(t)}, \underbrace{0, \dots, 0}_{d-2 \text{ elements}})$

$y_d^{(t)} \sim P(y_d \mid y_{-d}^{(t-1)}, x)$

where $y_{-2}^{(t-1)} = \underbrace{(y_1^{(t)}, y_2^{(t)} \dots y_{d-1}^{(t)})}_{d-1 \text{ elements.}}$

Subject 20H_Q3 Gibbs Sampling Date 9

Attendants Question 3.

3 $y^{(t)} = (y_1^{(t)}, y_2^{(t)} \dots y_d^{(t)})$

then loop 2

after $y^{(t)}$ is enough close

to stationarly distribution (after mixing time)

Then Sample $y_1^{(t)}, y_2^{(t)}$.

$p(y_1, y_2 | x) \approx$ num of samples which satisfy. $p(y_1^{(t)} = y_1, y_2^{(t)} = y_2 | x)$ / all num of sampling

Subject 2011_Q3 Gibbs Sampling

Date 9-(4)

Attendants

4. describe in your own words what is the mixing time of above Markov Chain?

<Bonus>

Give a precise mathematical characterization of mixing time of Markov Chain

$\boxed{\text{Answer}}$

mixing time $\tau(\varepsilon) = \max_{x \in \Omega} \left[\min \{ t \mid \forall s \geq t, \ d_{TV}(\pi, P_x^s) \leq \varepsilon \} \right]$

where x : a certain condition

Ω : condition set

t : step t (time)

$d_{TV}(\nu_1, \nu_2)$: total variation distance

$= \frac{1}{2} \sum_{x \in \Omega} | \nu_1(x) - \nu_2(x) |$

π : stational distribution

P_x^s : predicted distribution after s step

Subject 20H_Q3 DL Book Date 9-X

Attendants

平衡状態への収束

$$v^{(t)} = \left(V \operatorname{diag}(\lambda) V^{-1}\right)^t v^{(0)}$$

$$= V \operatorname{diag}(\lambda)^t V^{-1} v^{(0)}$$

Transition matrixは

固有値 < 1 は 0に収束

Subject 20H_Q3　　　　Date 10

Attendants MCMC : Markov chain Monte Carlo

1、所望の分布を定常分布にもつマルコフ連鎖を設計

2、十分な回数推移させて、定常分布からサンプリングします

↳ 何回？

- 近似サンプリング法

  → mixing time, (収束スピードの算定)

- 完璧サンプリング法

  → coupling from the past

  (無限回の推移の結果を出力、定常分布に従う)

適切な Markov Chain の設計 (w/ エルゴード性)

↓

定常分布に asymptotic に近づく

↓

これからサンプリングできる

↳ 所望の dist になる

Subject 2014_Q3 Date / /

Attendants

condition space Ω

transition matrix $P \in \mathbb{R}^{n \times n}$

Markov chain M

s.t. irreducible (既約)

$$\forall \{x, y\} \in \binom{\Omega}{2}$$

↑
permutation

$$ {}_{\Omega}P_2 = \frac{\Omega!}{(\Omega-2)!} = \Omega \times (\Omega - 1)$$

みたいな順列の組

$\exists t > 0$

$$\Pr(X^t = y \mid X^0 = x) > 0$$

時刻tがある

どこから始まっても
どこにでも
行ける

Subject 20H_Q3　Date 12

Attendants

非周期的 (aperiodic)

整数

$\forall x \in \Omega$, $\gcd\{t \in \mathbb{Z}_{++} \mid \Pr(X^t = x \mid X^0 = x) > 0\} = 1$

conditional space

最大公約数

が成立

一度すで状態 $\{x, y\}$

になったらまたこれに至る

保証がない

既約で非周期的

(irreducible) (aperiodic)

な有限マルコフ連鎖をエルゴード的という

$\Leftrightarrow$ ergodic

エルゴード的なマルコフ連鎖は

唯一の定常分布を持ち、極限分布は stational dist に一致する。

Subject 20H-Q3 Date 13

Attendants

Markov Property (マルコフ性)

現在の X_t が X_{t-1} にしか依存しない

$$P(X_t | X_0, X_1, \dots X_{t-1}) = P(X_t | X_{t-1})$$

Irreducibility 既約性

任意の状態 $i \to j$ へ有限回 step でたどりつくことができる

Periodicity 周期性

$p^{(n)}(i,i) > 0$ となる全ての整数 $n \geq 1$ の最大公約数を状態 i の周期

HMC time-Homogeneous Markov Chain

推移確率(行列)が時刻 t に依存しない

$d(i) = 1$ のとき非周期的という

Stational Distribution (定常分布)

$\alpha P = \alpha$ を満たす確率ベクトル α, 既約で非周期的な場合は n を大きくしたときの推移確率は定常分布になる

観測が長いときの各状態の現れ回数が定常分布に近づく (エルゴード性 ergodic)

Subject 2017_Q3 Date 14

Attendants 詳細均衡式 (detailed balance equation)

$f : \Omega \to \mathbb{R}_{++}$

$$f(x)\,P(x,y) = f(y)\,P(y,x) \quad {}^\forall\{x,y\} \in \binom{\Omega}{2}$$

定常分布 π $\quad {}^\forall x \in \Omega$

$$\pi(x) = C \cdot f(x) \quad {}^\forall C : \text{constant}$$

Metropolis-Hasting Algo

状態 X

近傍 Y ← Xの近傍はk個、$\frac{1}{k}$で選ぶ

確率 $\dfrac{f(Y)}{f(X)+f(X)}$ で Yに推移、それ以外はXにとどまる

↳このマルコフ連鎖の定常分布はfに比例する

従って ① 効率的に近傍を見つけられる

② $x \in \Omega$ に対し $f(x)$ が容易に計算できれば

所望の分布が漸近的に得られる

Subject 20H_Q3　　　　Date 15

Attendants マルコフ連鎖の収束

Total variation distance (全変動距離)

of dist ν_1 ν_2

$$d_{TV}(\nu_1, \nu_2) = \max_{A \subseteq \Omega} \left\{ \sum_{x \in A} (\nu_1(x) - \nu_2(x)) \right|$$

$$= \frac{1}{2} \sum_{x \in \Omega} |\nu_1(x) - \nu_2(x)|$$

Mixing Time (混交時間)

状態空間 Ω を持つエルゴードマルコフ連鎖 M の混交時間は、

$$\tau(\varepsilon) = \max_{x \in \Omega} \{ \min \{ t \mid \forall s \ge t, \; d_{TV}(\pi, P_x^s) \le \varepsilon \} \}$$

π: マルコフ連鎖 M の定常分布とし、

P_x^s は初期状態 $x \in \Omega$ として時刻 $0 \to s \ge 0$ まで推移させた時のマルコフ連鎖 M の確率分布

$\Rightarrow$ $\tau(\varepsilon)$ 回以上推移させると定常分布との d_{TV} が ε 以下になる。

Subject 20H_Q3

Date 16

Attendants