Lec 11. Markovian Random Fields & Hidden Markov Models.

$G(V, E)$. X_v for each $v \in V$, $X_v \in [q_v]$.

$(X_1, \dots X_n) \sim \mu(\cdot)$. μ_v – margial.

$p(X_v \mid X_{\neq v}) = p(X_v \mid X_{N(v)})$. Markov Property.

A distribution on $\Omega = \prod_{v \in V} [q_v]$ is called a

Gibbs distribution if $\forall\, \sigma \in \Omega$.

$$\mu(\sigma) = \prod_{\substack{A \subseteq V \\ A \text{ complete}}} V_A(\sigma) = \prod_{\substack{A \subseteq V \\ A \text{ clique}}} \hat{V}_A(\sigma).$$

Example

Hardcore.

Ising.

Thm. [Hammersley-Clifford] $p > 0$

p Satisfies Markov Property

$\Leftrightarrow$ μ is Gibbs.

Pf. " $\Leftarrow$ " "trivial".

" $\Rightarrow$ ". Assume μ is a distribution with Markov Property.

Will Prove $\forall D \subseteq V$, $p(x_D, 1_{D^c}) = \prod_{A \subseteq D} V_A(x)$.

Induction on D.

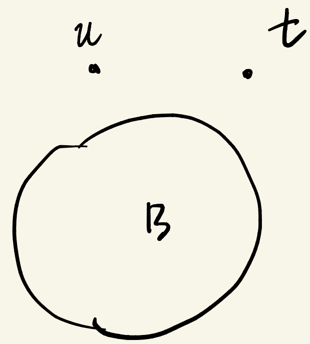
* . $V_\phi(x) = p(1)$

* . $V_D(x) = \dfrac{p(x_D, 1_{D^c})}{\prod_{A \subsetneq D} V_A(x)}$.

Verify that if $V_A(x) = 1$ if D is not complete.

induction on $|D|$.

$|D| = k+1.$



Goal.

$$p(x_D, 1_{D^c}) = \prod_{A \subsetneq D} V_D(x).$$

$$p(x_D, 1_{D^c}) = p(x_t, x_u, x_B, 1_{D^c})$$

$$= \frac{p(x_t, x_u, x_B, 1_{D^c})}{p(1_t, x_u, x_B, 1_{D^c})} \cdot p(1_t, x_u, x_B, 1_{D^c})$$

$$= \frac{p(x_t \mid x_u, x_B, 1_{D^c})}{p(1_t \mid x_u, x_B, 1_{D^c})} \cdot p(1_t, x_u, x_B, 1_{D^c})$$

$$= \frac{p(x_t \mid 1_u, x_B, 1_{D^c})}{p(1_t \mid 1_u, x_B, 1_{D^c})} \cdot p(1_t, x_u, x_B, 1_{D^c})$$

$$= \frac{p(x_t, 1_u, x_B, 1_{D^c})}{p(1_t, 1_u, x_B, 1_{D^c})} \cdot p(1_t, x_u, x_B, 1_{D^c})$$

Then $V_A \equiv 1$ if $\{t,u\} \subseteq A \subset D$.

$$= \frac{\prod_{A \subseteq B \cup \{t\}} V_A(x) \cdot \prod_{A \subseteq B \cup \{u\}} V_A(x)}{\prod_{A \subseteq B} V_A(x)} = \prod_{\substack{A \subsetneq D \\ \{t,u\} \not\subseteq A}} V_A(x)$$

How to sample from MRF ?

$\left(\frac{1}{2}, \frac{1}{2}\right)$. $\quad (0.1, 0.9)$.

HMM.

$$
\xi \quad
\begin{tikzcd}
X_0 \arrow[r, "A"] \arrow[d, "B"] & X_1 \arrow[r, "A"] \arrow[d, "B"] & X_2 \arrow[r] & \cdots \arrow[r] & X_n \arrow[d] \\
Y_0 & Y_1 & & & Y_n
\end{tikzcd}
$$

$$\theta = (\xi, A, B).$$

Likelihood function

$$L = L(\theta) = P_\theta(y_0, \dots, y_n) = \sum_{x_0, \dots x_n} P_\theta(x_0, \dots, x_n, y_0, \dots, y_n)$$

$$= \sum_{x} P_\theta(x, y).$$

Q. Given θ. How to compute $P_\theta(y_0, \dots, y_n)$. ?

DP.

Maximum Likelihood

Given $y_0, \ldots y_n$. Find θ maximizing $P_\theta(y)$.

The Method of Expectation Maximization

Maximize $L(\theta) \iff$ maximize $\log L(\theta)$.

$$\theta_{t+1} = \underset{\theta}{\arg\max} \; \mathbb{E}_{\theta_t}\left[\log P_\theta(X, y) \mid Y = y \right].$$

Lem. If $\mathbb{E}_{\theta_0}\left[\log P_{\theta_1}(X, y) \mid Y = y \right] > \mathbb{E}_{\theta_0}\left[\log P_{\theta_0}(X, y) \mid Y = y \right]$

then $P_{\theta_1}(y) > P_{\theta_0}(y)$.

Pf. $$0 < \mathbb{E}_{\theta_0}\left[\log \frac{P_{\theta_1}(X, y)}{P_{\theta_0}(X, y)} \mid Y = y \right]$$

$$= \sum_x P_{\theta_0}(x \mid y) \cdot \log \frac{P_{\theta_1}(x, y)}{P_{\theta_0}(x, y)}$$

$$= \sum_x P_{\theta_0}(x \mid y) \log \frac{P_{\theta_1}(y)}{P_{\theta_0}(y)} - \sum_x P_{\theta_0}(x, y) \frac{P_{\theta_0}(x \mid y)}{P_{\theta_1}(x \mid y)}$$

$$= \log \frac{p_{\theta_1}(y)}{p_{\theta_0}(y)} - D_{KL}(p_{\theta_0}(\,), p_{\theta_1}(\,)) \leq \log \frac{p_{\theta_1}(y)}{p_{\theta_0}(y)}.$$

EM on HMM.

$$\log p_\theta(X, y) = \log \xi(X_0) + \sum_{t=0}^{n-1} \log A(X_t, X_{t+1})$$

$$+ \sum_{t=0}^{n} \log B(X_t, y_t).$$

$$\mathbb{E}_{\theta_0}[\log p_\theta(X,y) \mid Y=y]$$

$$= \underline{\sum_i \Pr_{\theta_0}[X_0 = i \mid Y=y] \cdot \log \xi(i)} \quad ①$$

$$+ \underline{\sum_{t=0}^{n-1} \sum_{i,j} \Pr_{\theta_0}[X_t = i, X_{t+1} = j \mid Y=y] \log A(i,j)} \quad ②$$

$$+ \underline{\sum_{t=0}^{n} \sum_i \Pr_{\theta_0}[X_t = i \mid Y=y] \log B(i, y_t).} \quad ③$$

① . $\xi(i) = \Pr_{\theta_0}[X_0 = i \mid Y=y]$

② $= \sum_i \left(\sum_j \left(\sum_t P_{\theta_0}[X_t = i, X_{t+1} = j \mid y] \right) \log A(i,j) \right)$

③ $= \sum_i \sum_j \left(\sum_{t: y_t = j} P_{\theta_0}[X_t = i \mid y] \right) \log B(i,j)$

<u>prop</u>. $p=(p_1,\dots,p_K)$, $q=(q_1,\dots,q_K)$.

"Kullback-Leibler distance"

$$D_{KL}(p\,\|\,q) \overset{\Delta}{=} \sum_i p_i \log p_i/q_i \geq 0.$$

$$\text{"} = \text{"} \iff p = q.$$

<u>pf</u>. W.L.O.G $p_i, q_i > 0$.

$\forall x > 0$, $\log x \leq x-1$

$$\sum_i p_i \log q_i/p_i \leq \sum_i p_i\left(\frac{q_i}{p_i}-1\right) = \sum (q_i - p_i)$$
$$= 0.$$