CS281: Advanced ML **October 11, 2017**

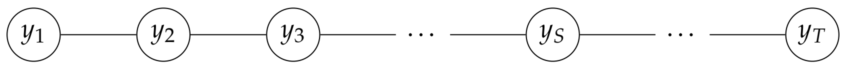
Lecture 11: Exact Inference: Belief Propagation

Lecturer: Sasha Rush *Scribes: Ismail Ben Atitallah, Hao Wu, Raphael Rovinov, Ziliang Che, Jiaoyang Huang*

11.1 Simple Marchov Chain

For a simple Markov Chain with time limit T and V classes per node:

$$\begin{aligned} p(y,t) &= \exp(\sum_t \theta_t^T(y_{t-1}, y_t) + \theta_t^o(y_t) - A(\theta)) \\ &\propto \prod_t \psi_t(y_{t-1}, y_t)\psi_t(y_t) \end{aligned}$$



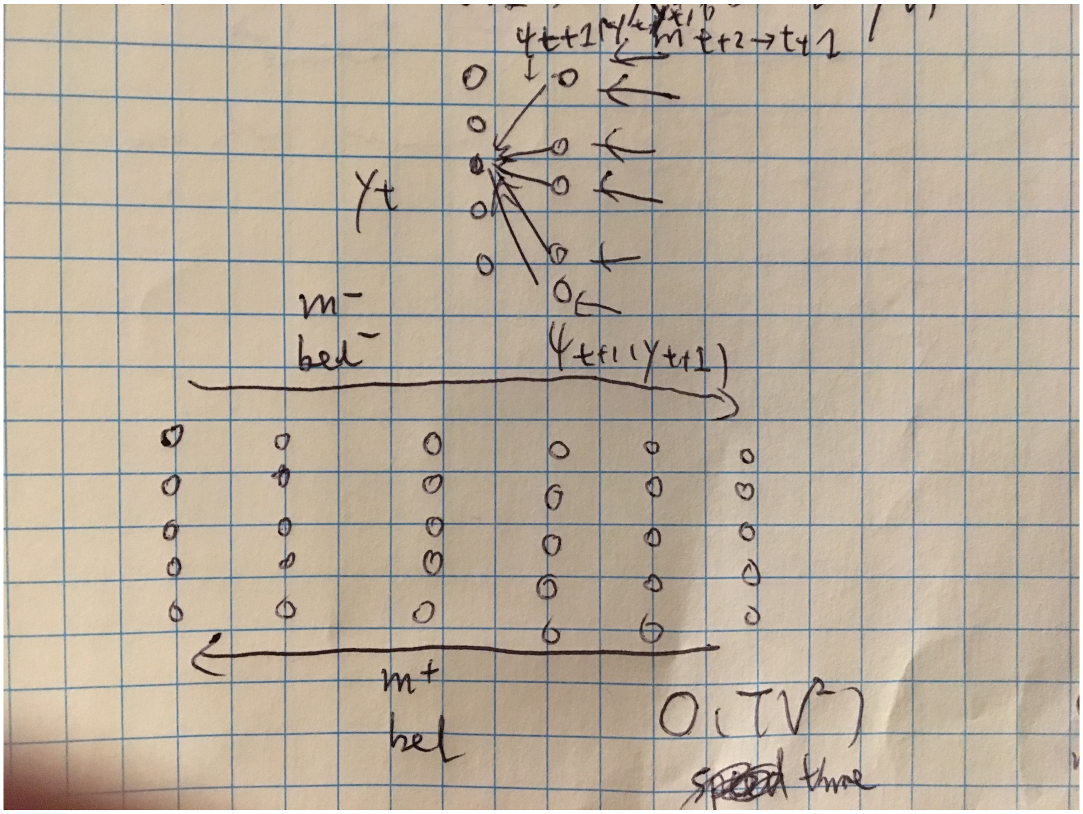
11.1.1 Distributive Property

Marginal:

$$\begin{aligned} p(y_s = v) &= \sum_{y_{1:T, y'_s = v}} \prod_t \psi_t(y'_{t-1}, y'_t)\psi(y'_t)/Z(\theta) \\ &= \sum_{y'_t} \psi_T(y'_T) \sum_{y_T - 1}^{\prime} \psi_{T-1}(y'_{t-1})\psi_T(y_{T-1}, y'_T) \sum \dots \sum_{y'_2} \psi_2(y'_2) \sum_{y'_1} \psi_1(y'_1, y'_2)\psi_1(y'_1) \end{aligned}$$

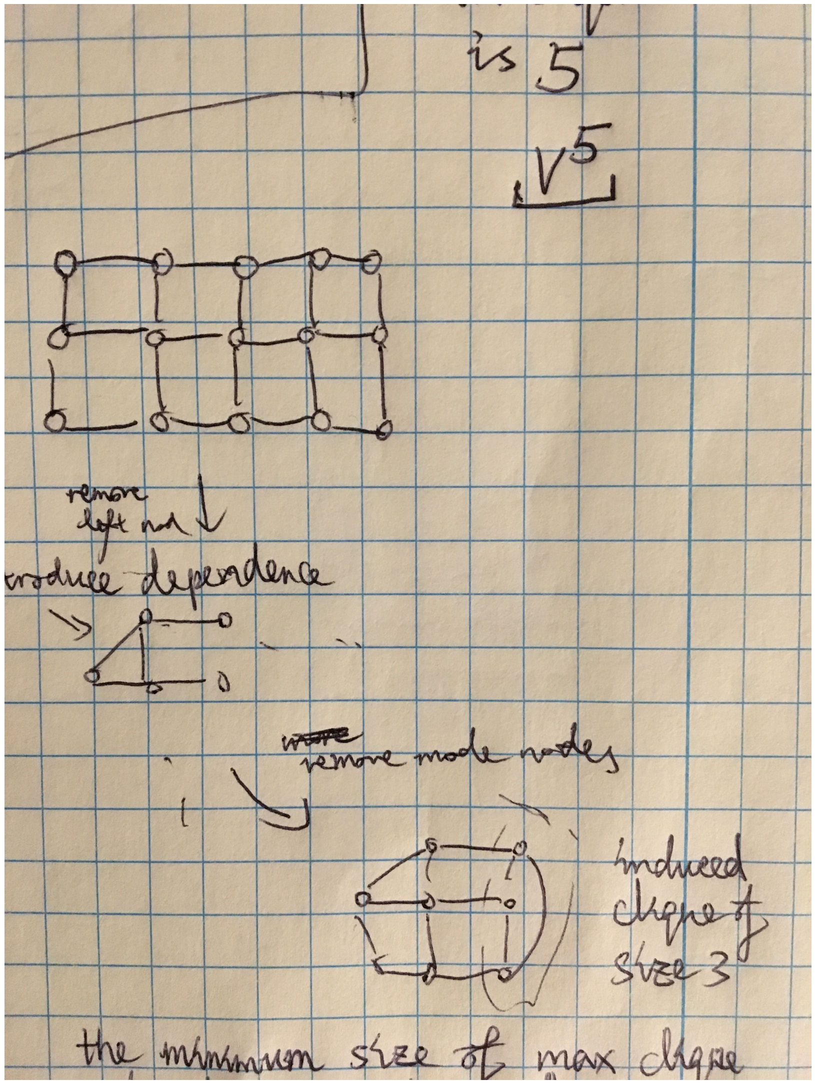
11.1.2 Compute All Marginals with Dynamic Programming

$$\begin{aligned} \mathrm{bel}_t^-(y_t) &\propto \psi_t(y_t)\mathrm{m}_{t-1\to t}^-(y_t) \\ \mathrm{m}_{t-1\to t}^- &= \sum_{y_{t-1}} \psi_{t-1}(y_{t-1}, y_t)\mathrm{bel}_{t-1}^-(y_{t-1}) \\ \mathrm{m}_{t+1\to t}^+ &= \sum_{y_{t+1}} \psi_{t+1}(y_{t+1}, y_t)\psi_{t+1}(y_{t+1})\mathrm{m}_{t+2\to t+1}^+(y_{t+1}) \\ p(y_t) = \mathrm{bel}_t(y_t) &\propto \mathrm{m}_{t+1\to t}(y_t)\mathrm{bel}_t^-(y_t) \end{aligned}$$



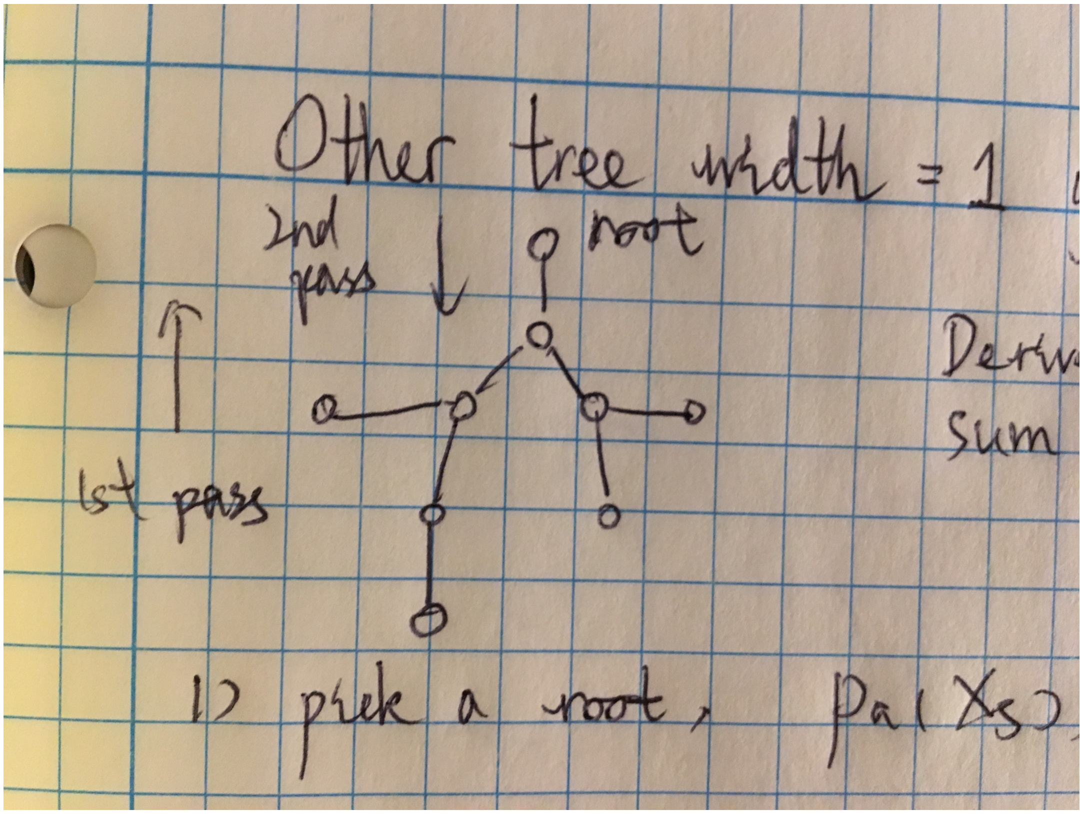
11.2 Other Graphs

$$p(y_s = v) = \sum_{y', y'_s = v} \prod_c \psi_c(y'_c)$$



Computing this sum product might be hard. Here we define the minimum size of maximum clique indcued -1 to be the *treewidth* of the graph.

11.2.1 Compute Marginals for treewidh = 1 Graph



Here we derive the generalization of the forward-backward sum-product algorithm

1. Pick a root s, $\text{pa}(x_s)$,$\text{ch}(x_s)$
2. Upward pass:

$$\text{m}_{s \to t}(x_t) = \sum_{x_s} \psi_{s-t}(x_s, x_t)\text{bel}_s^-(x_s)$$
$$\text{bel}_t^-(x_t) \propto \psi_t(x_t) \prod_{s \in \text{ch}(t)} m_{s-t}^-(x_t)$$

3. Downward pass:

$$\text{bel}_s(x_s) \propto \text{bel}_s^-(x_s) \prod_{t \in \text{pa}(s)} \text{m}_{t \to s}^+(x_s)$$
$$\text{m}_{t \to s}^+(x_s) = \sum_{x_t} \psi s - t(x_s, x_t)\psi_t(x_t) \prod_{c \in \text{ch}(t), c \neq s} = \text{m}_t^-(x_t)$$

11.3 Parallel Protocol for Sum Product

$$\text{bel}_s(x_s) \propto \psi_s(x_s) \prod_{t \in \text{nbr}(s)} \text{m}_{t \to s}(x_s)$$
$$\text{m}_(s \to t) = \sum_{x_s} \psi_s(x_s)\psi_{s-t} \prod_{u \in \text{nbr}(s), u \neq t} (x_s)$$

11.4 Final Notes

For this lecture we have been ultilizing + and $\times$ and distributive property

11.4.1 Commutative Semi-ring

$$\begin{array}{ccc}
+ & \text{max} & \cap \\
\times & \times & \cup \\
\vee & \vee & \vee \\
\text{marginal} & \text{argmax} & \text{satisfying assignment}
\end{array}$$