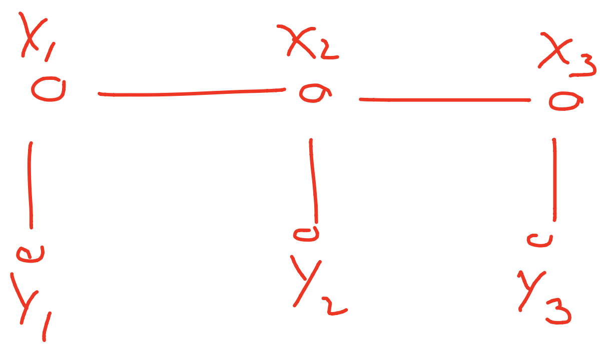
Last time we described a scheme to compute sums & maximas of Gibbs/MRFs.

An important example: HMM



$$P(x,y) = P(x_1)P(y_1|x_1)P(x_2|x_1)$$
$$P(y_2|x_2)P(x_3|x_2)P(y_3|x_3)$$

$$P(y) = \sum_{x} P(x,y)$$

$f_1(x_1)$ $\quad f_2 \quad f_3$

$$= \sum_{x_3}\sum_{x_2}\sum_{x_1} \Big\{ P(x_1)P(y_1|x_1)P(x_2|x_1)$$

$$\left(\left(\mathbb{P}(Y_2 \mid X_2)\right)\mathbb{P}(X_3 \mid X_2)\,\mathbb{P}(Y_3 \mid X_3)\right)$$

Could continue this process for $n > 3$

Thus we could define a recurrence.

$$f_1 = \mathbb{P}(X_1)\,\mathbb{P}(Y_1 \mid X_1)$$

$$f_i = \sum_{X_{i-1}} \left[f_{i-1}(X_{i-1})\,\mathbb{P}(X_i \mid X_{i-1}) \right] \mathbb{P}(Y_i \mid X_i)$$

"forward" probabilities.

<u>Rmk</u>: $f_1 = \mathbb{P}(X_1, Y_1)$ Suppose that

$$f_{i-1}(X_{i-1}) = \mathbb{P}(X_{i-1}, Y_{i-1}, \ldots, Y_1)$$

then

$$f_i(X_i) = \sum_{X \cdot \cdot} \mathbb{P}(X_{i-1}, Y_{i-1}, \ldots, Y_1)\,\mathbb{P}(X_i \mid X_{i-1})\,\mathbb{P}(Y_i \mid X_i)$$

$_{i-1}$

$$= \sum_{x_{i-1}} \mathbb{P}(x_i, y_{i-1}, \dots, y_1)\, \mathbb{P}(x_i \mid x_{i-1}, y_1, \dots, y_{i-1})\, \mathbb{P}(y_i \mid x_i)$$

$$= \sum_{x_{i-1}} \mathbb{P}(x_i, x_{i-1}, y_{1:i-1})\, \mathbb{P}(y_i \mid x_i, x_{i-1}, y_{1:i-1})$$

$$= \sum_{x_{i-1}} \mathbb{P}(x_{i-1}, x_i, y_{1:i-1}, y_i) = \mathbb{P}(x_i, y_{1:i})$$

In practice we work with logs

$$\tilde{f}_i(x_i) = \log f_i(x_i)$$

$$\tilde{f}_1(x_1) = \log \mathbb{P}(x_1) + \log \mathbb{P}(y_1 \mid x_1)$$

$$\tilde{f}_i(x_i) = \log f_i(x_i)$$

$$= \log \Big\{ \sum_{x_{i-1}} f_{i-1}(x_{i-1})\, \mathbb{P}(x_i \mid x_{i-1}) \Big\}$$

x_{i-1}

$$= \log \Big\{ \sum_{x_{i-1}} \exp\Big(\log f_{i-1}(x_{i-1}) + \log \mathbb{P}(X_i \mid X_{i-1}) \Big) \Big\}$$

$$+ \log \mathbb{P}(Y_i \mid X_i)$$

$$\boxed{= \log \sum_{x_{i-1}} \exp\Big\{ \tilde{f}_{i-1}(x_{i-1}) + \log \mathbb{P}(x_i \mid x_{i-1}) \Big\} + \log \mathbb{P}(y_i \mid x_i)}$$

Thus with

$$l_1 \oplus l_2 = \log\left(e^{l_1} + e^{l_2}\right)$$

$$\boxed{\tilde{f}_i(x_i) = \bigoplus \Big(\tilde{f}_{i-1}(x_{i-1}) + \log \mathbb{P}(x_i \mid x_{i-1}) \Big) + \log \mathbb{P}(y_i \mid x_i)}$$

Similarly if we want

$$\hat{X} = \arg\max_{X} \mathbb{P}(X|Y) = \arg\max_{X} \mathbb{P}(X,Y)$$

$$\max_{X} \mathbb{P}(X,Y) = \max_{x_3}\Big\{\max_{x_2}\Big\{\max_{x_1} \mathbb{P}(X_1)\mathbb{P}(Y_1|X_1)\mathbb{P}(X_2|X_1)\Big\}\mathbb{P}(Y_2|X_2)\mathbb{P}(X_3|X_2)\Big\}$$

$$\mathbb{P}(Y_3|X_3)\Big\}$$

Define

$$m_1(X_1) = \mathbb{P}(X_1)\mathbb{P}(Y_1|X_1)$$

$$m_i(X_i) = \max_{X_{i-1}} m_{i-1}(X_{i-1})\mathbb{P}(X_i|X_{i-1})\mathbb{P}(Y_i|X_i)$$

$$b_{i-1}(X_i) = \arg\max_{X_{i-1}} m_{i-1}(X_{i-1})\mathbb{P}(X_i|X_{i-1})$$

$_{i-1}$

Then use traceback to find the

x_i:

$$b_n = \underset{x_n}{\operatorname{argmax}}\ m_n(x_n)$$

$$\hat{x}_n = b_n$$

$$\hat{x}_{n-1} = b_{n-1}(\hat{x}_n)$$

$$\vdots$$

$$\hat{x}_1 = b_1(\hat{x}_2)$$

In practice

$$\tilde{m}_i(x_i) = \log m_i(x_i)$$

$$= \log \max m_{i-1}(x_{i-1})\, \mathbb{P}(X_i \mid X_{i-1})$$

$$+$$

$$\log \mathbb{P}(Y_i \mid X_i)$$

$$= \max_{x_{i-1}} \left\{ m_i(x_{i-1}) + \log \mathbb{P}(x_i \mid x_{i-1}) \right\} + \log \mathbb{P}(y_i \mid x_i)$$

Same as the softmax relation - just with hard max.

— "Viterbi algorithm"

<u>Rmk:</u> For optimization this is a dynamic program