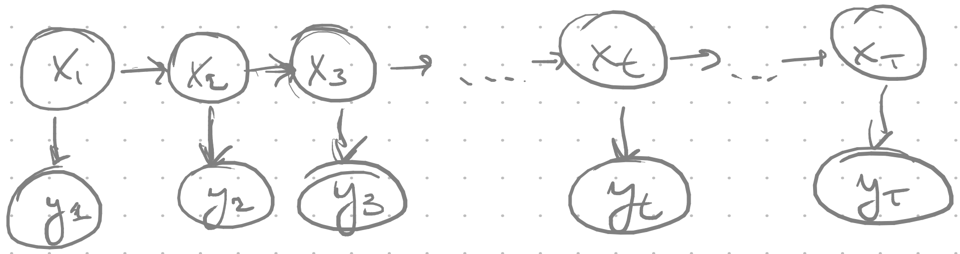
① Directed graphical models / Bayesian belief networks / J. Pearl causality

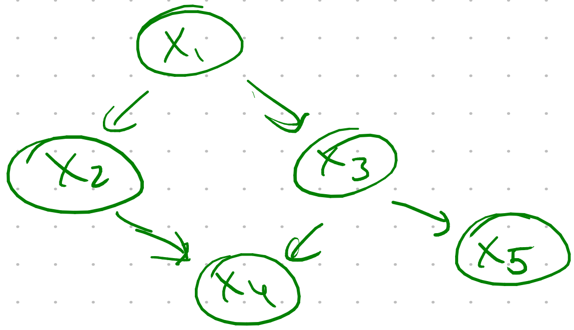


$$\left[\,p(X,Y) = p(x_1)\,p(y_1|x_1)\,p(x_2|x_1)\,p(y_2|x_2)\dots p(x_T|x_{T-1})\,p(y_T|x_T)\,\right]$$

$$p(x_1,x_2,\dots,x_n) = p(x_1)\,p(x_2|x_1)\,p(x_3|\cancel{x_1,x_2})\,p(x_4|\cancel{x_1,x_2,x_3})\dots p(x_n|\cancel{x_1,\dots,x_{n-1}})$$

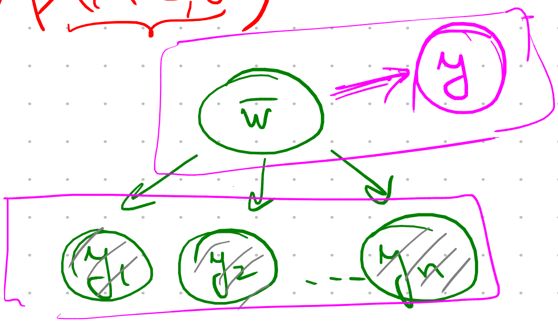
$$p(x_1,\dots,x_n) = \prod_{i=1}^{n} p(x_i\,|\,par(x_i))$$

$$\sum_{x_2} p(x_1 \dots x_n)$$



$$p(x_1 \dots x_5) = p(x_1)\,p(x_2|x_1)\,p(x_3|x_1)\,p(x_4|x_2,x_3)\,p(x_5|x_3)$$

$$p(\bar\theta, Z, X) = p(\bar\theta)\,\underbrace{p(Z|\bar\theta)}\,\underbrace{p(X|Z,\bar\theta)}$$



$$p(\bar w, Y) = p(\bar w)\cdot\prod_{n=1}^{N} p(y_n|\bar w)$$

$D = \{Y\}$ $\qquad p(\bar w|Y) = ?$

$$p(y|Y) = \int p(\bar w, y|Y)\,d\bar w$$

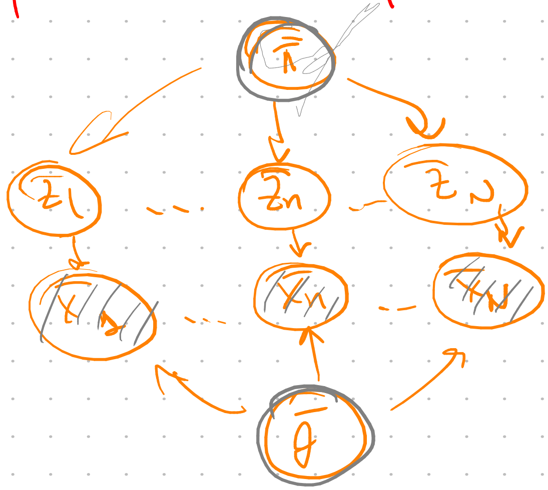
$p(\bar w, y, Y)$

$\downarrow$

$p(\bar w, y|Y)$

clustering $\quad p(\bar\pi,\ \bar\theta = \{\bar\mu_k, \Sigma_k\}_{k=1}^{K},\ Z,\ X)$

$$p(\bar\pi, \bar\theta, Z, X) = p(\bar\pi)\,p(\bar\theta)\prod_{n=1}^{N} p(\bar z_n|\bar\pi)\,p(\bar x_n|\bar z_n, \bar\theta)$$



$$p(\bar\pi, \bar\theta|X) = \int p(\bar\pi, \bar\theta, Z|X)\,dZ$$

② d-разделимость

1) $n=1$ $\quad X_1 \quad p(x_1) = p(x_1)$

$X_1 \leftarrow X_2 \quad p(x_2)p(x_1|x_2)$

$=$

$X_1 \to X_2$

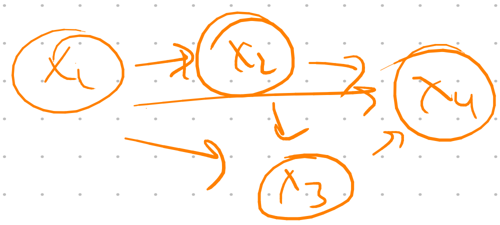
$p(x_1, x_2) = p(x_1)\,p(x_2|x_1)$

2) $n=2$ $\quad X_1 \quad X_2$

$p(x_1, x_2) = p(x_1)\,p(x_2)$

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

tournament



3) $n=3$ $\quad X_1 \quad X_2 \quad X_3$

$p(x_1)\,p(x_2)\,p(x_3)$

$X_1 \quad X_2 \to X_3$

$p(x_1)\,p(x_2)\,p(x_3|x_2)$

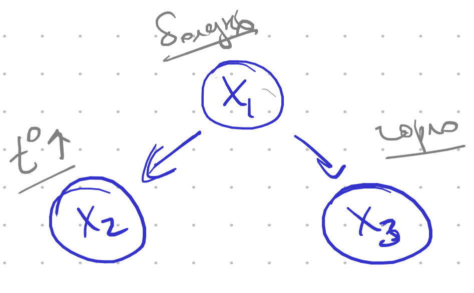
$= p(x_1)\,p(x_2|x_1)\,p(x_3|x_1, x_2)$

$X_1 \to X_2 \to X_3$ последовательная связь

$$\boxed{p(x_1, x_2, x_3)} = p(x_1)\,p(x_2|x_1)\,p(x_3|x_2) \quad \Downarrow ?$$

условная независ.: $\quad p(x_1, x_3 | x_2) \overset{?}{=} p(x_1|x_2)\,p(x_3|x_2)$

$$= \frac{p(x_1, x_2, x_3)}{p(x_2)} = \frac{p(x_1)\,p(x_2|x_1)\,p(x_3|x_2)}{p(x_2)}$$



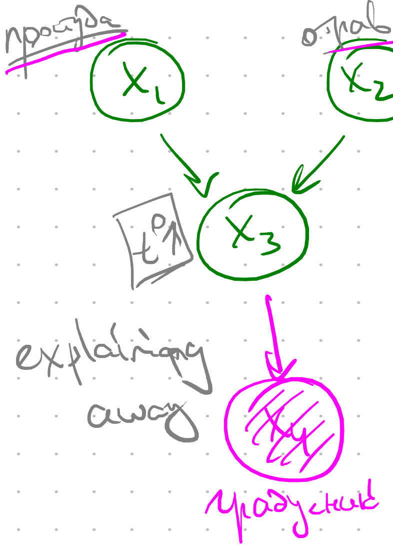
расходящаяся связь

$$p(x_1, x_2, x_3) = p(\cancel{x_1})\,p(x_2|x_1)\,p(x_3|x_1)$$

$$p(x_2, x_3 | x_1) \overset{?}{=} p(x_2|x_1)\,p(x_3|x_1)$$

$$= \frac{p(x_1, x_2, x_3)}{\cancel{p(x_1)}}$$

сходящаяся связь

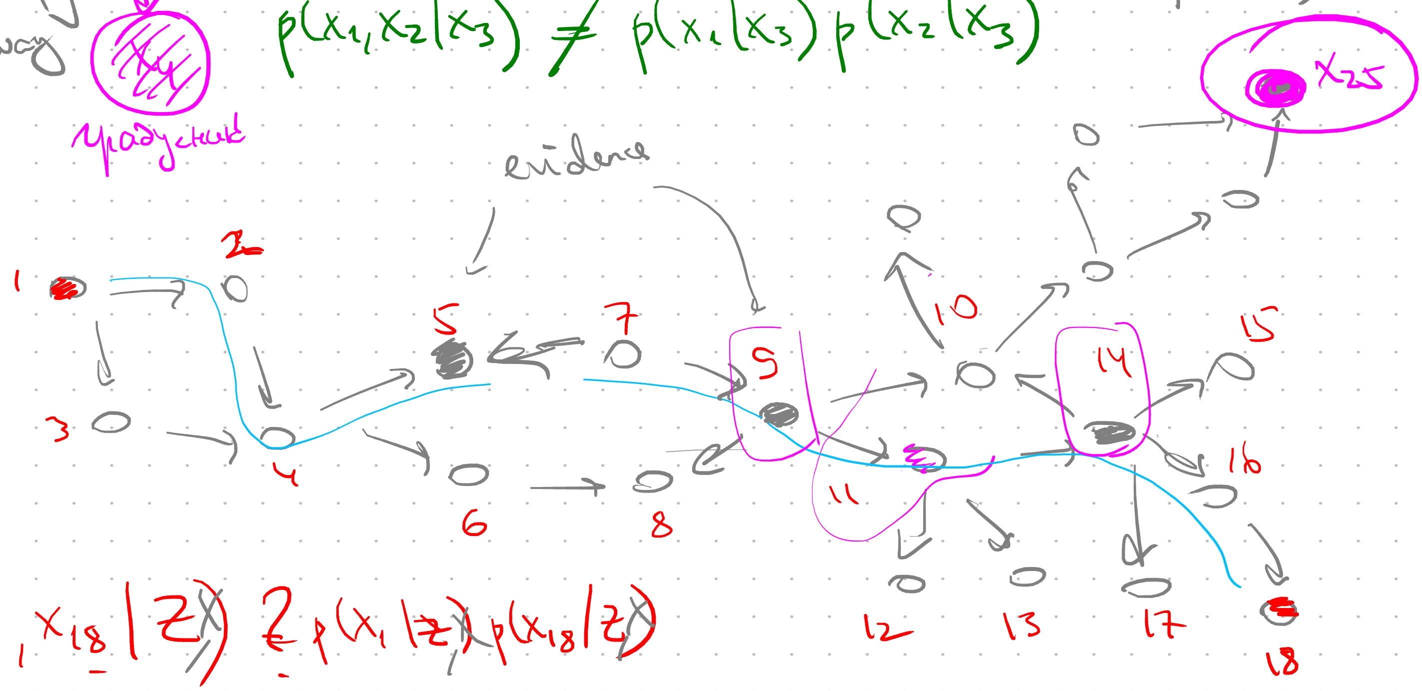


$$p(x_1, x_2, x_3) = p(x_1)\,p(x_2)\,p(x_3 | x_1, x_2)$$

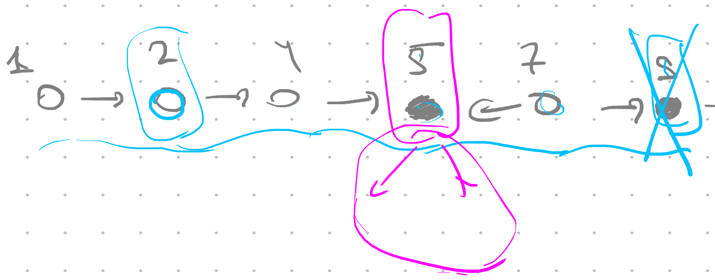
$$= \frac{p(x_1, x_2, x_3)}{p(x_1, x_2)}$$

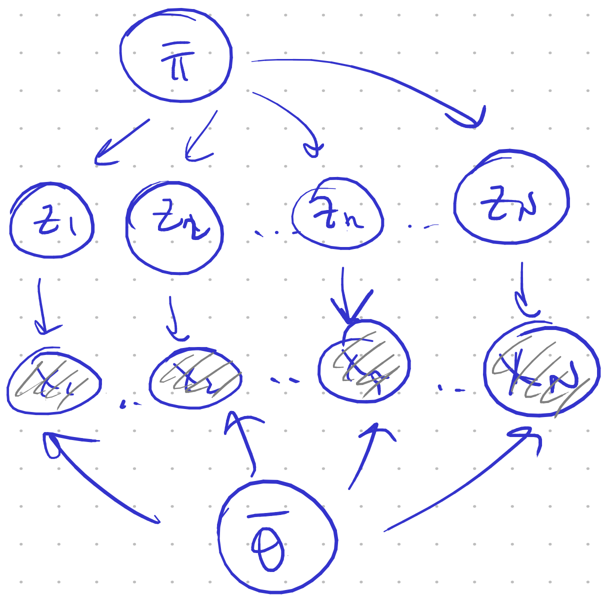
$$p(x_1, x_2) = p(x_1)\,p(x_2)$$

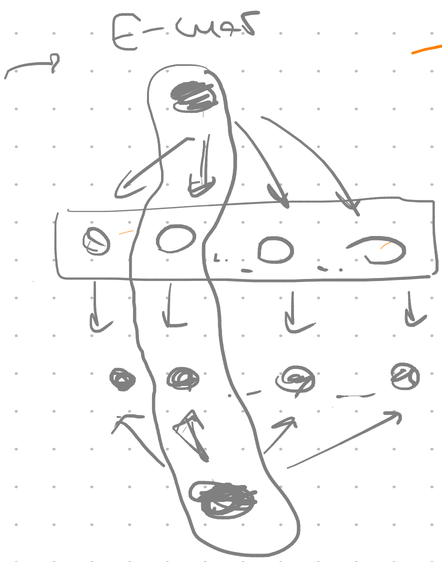
$$p(x_1, x_2 | x_3) \neq p(x_1 | x_3)\,p(x_2 | x_3)$$

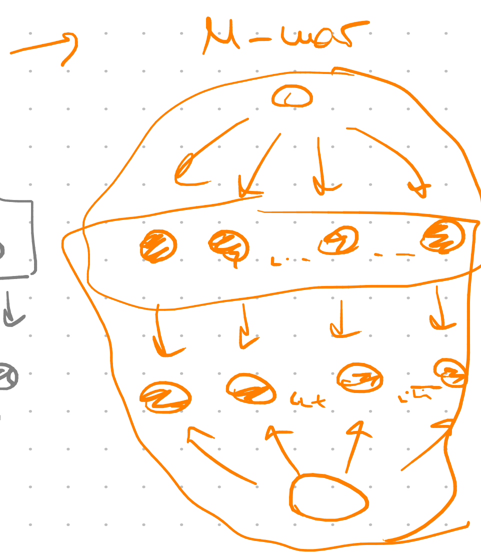


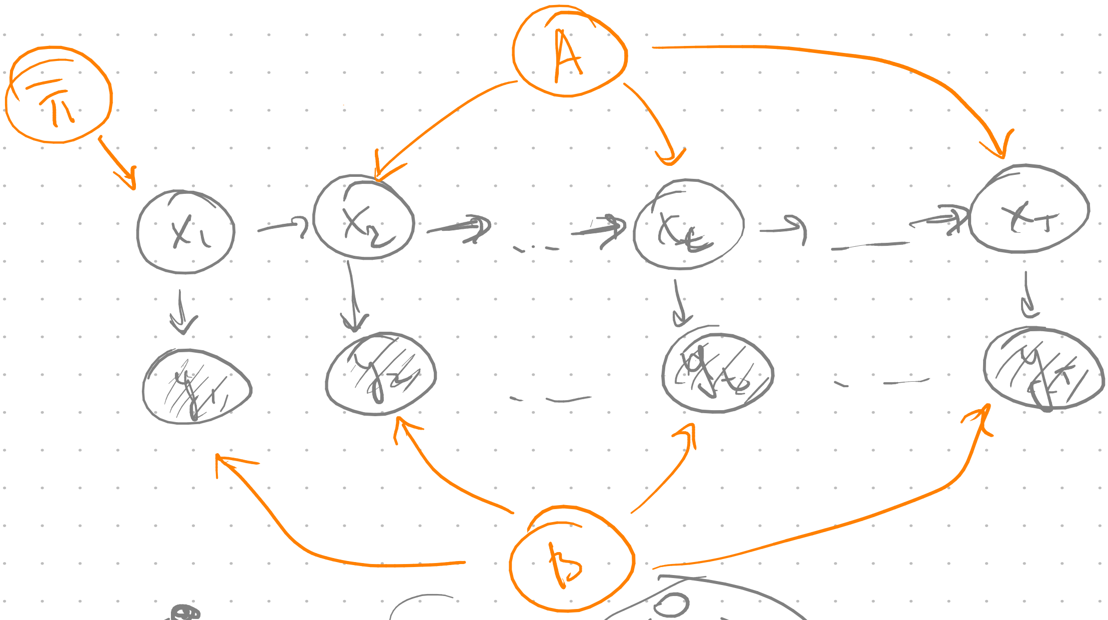
$$p(x_1, x_{18} | Z) \neq p(x_1 | Z)\,p(x_{18} | Z)$$





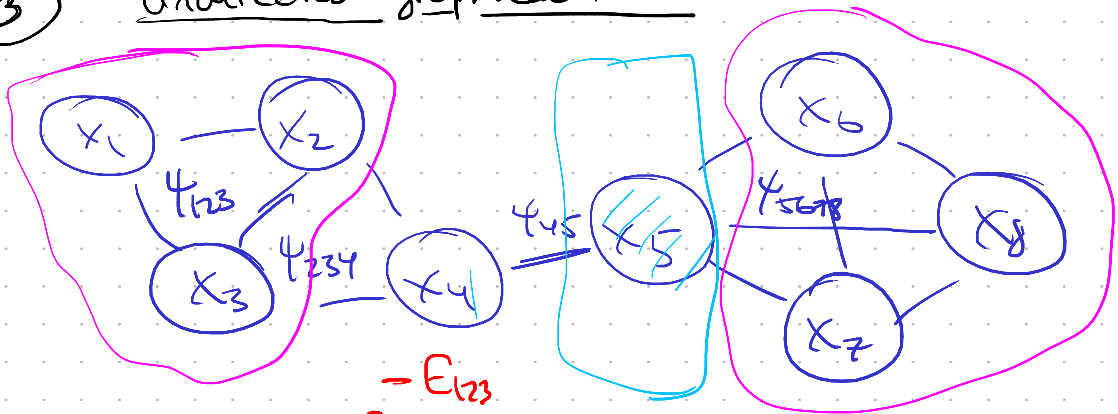






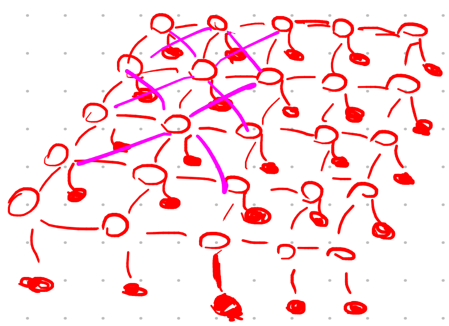
$$p(x, y, \pi, A, B) = p(\pi)\, p(A)\, p(B) \cdot$$
$$\times\, p(x_1 | \pi) \cdot p(y_1 | x_1, B)\, p(x_2 | x_1, A)\, p(y_2 | x_2, B) \ldots$$

③ Undirected graphical models



$E = E_{123} + \_\_\_\_$

$$f(x_1, \ldots, x_8) = \underbrace{\psi_{123}(x_1, x_2, x_3)}_{= e^{-E_{123}}}\, \psi_{234}(x_2, x_3, x_4)\, \psi_{45}(x_4, x_5)\, \psi_{5678}(x_5, x_6, x_7, x_8)$$



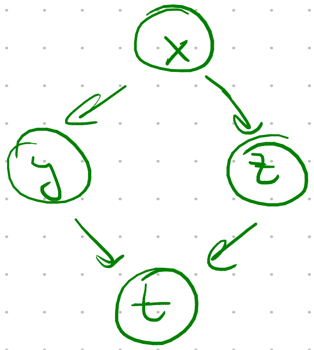
$x_i \quad y_i$

$$\underbrace{p(x, y)}_{\max} = \prod_i p(y_i | x_i) \cdot \prod_i \prod_{j \in nei(x_i)} p(x_i | x_j)$$

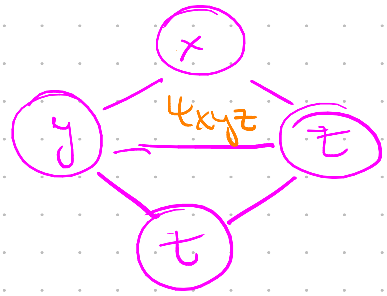
$p(y_i | x_i) \propto e^{-\alpha x_i y_i}$

$p(x_i | x_j) \propto e^{-\beta x_i x_j}$

$$\boxed{\min \sum \alpha x_i y_i + \sum_i \sum_{j \in nei(i)} \beta x_i x_j}$$



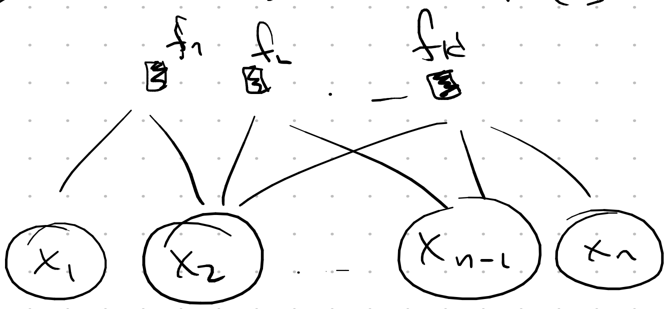
$$p(x,y,z,t) = \underbrace{p(x)p(y|x)}_{\psi_{xy}}\ \underbrace{p(z|x)}_{\psi_{xz}}\ \underbrace{p(t|y,z)}_{\psi_{yzt}}$$

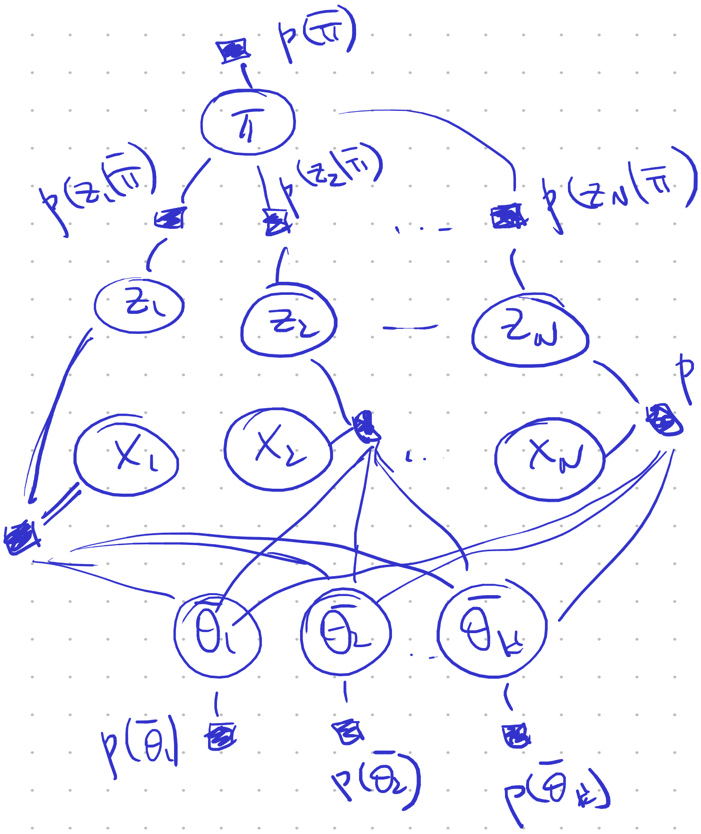


$$\psi_{yzt} \cdot \underbrace{\psi_{xyz}}_{=\ \psi_{xy} \cdot \psi_{xz}}$$

4 Factor graph

$$F(x_1 \dots x_n) = f_1(X_1) f_2(X_2) \dots f_k(X_k)$$





$$p(x_i | z_1, \bar\theta_1 \dots \bar\theta_k)$$

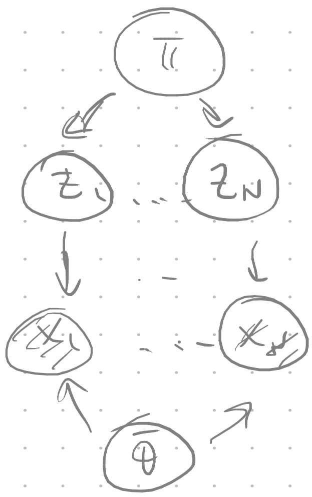
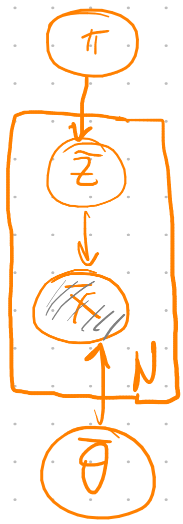
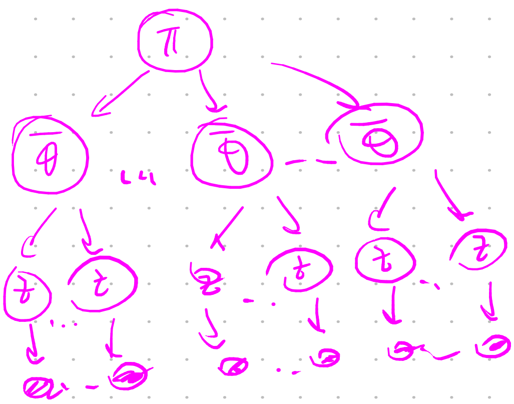
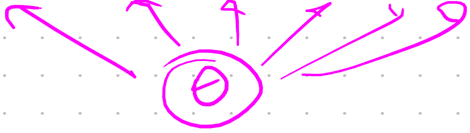
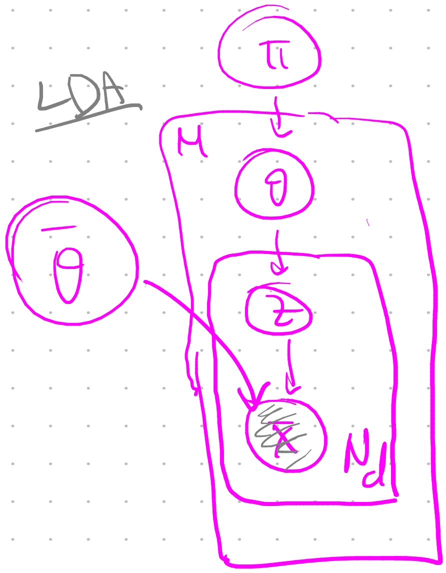


plate notation









5 Пример маргинализации

$$F(x_1 \ldots x_n) = f_1(x_1) \ldots f_k(x_k)$$

$$f(x_i) = \sum_{x_1 - \cancel{x_i} - x_n} F(x_1 \ldots x_n)$$

$$p(\bar{x} \mid D) \propto \int p(\bar{x}, \bar{\theta}, D)\, d\bar{\theta}$$

$$f(x_t) = \sum_{\cancel{x_t}} F(x_1 \ldots x_n)$$

$$F(x_1 \ldots x_n) = \psi_{12}(x_1, x_2)\, \psi_{23}(x_2, x_3) \ldots \psi_{T-1,T}(x_{T-1}, x_T)$$

$$f(x_t) = \sum_{x_1} \sum_{x_2} \ldots \sum_{x_{t-1}} \sum_{x_{t+1}} \ldots \sum_{x_n} \left(\underbrace{\psi_{12} \cdot \psi_{23} \ldots \psi_{t-1,t}}_{x_1 \ldots x_{t-1},\, x_t} \; \underbrace{\psi_{t,t+1} \ldots \psi_{T-1,T}}_{x_t,\, x_{t+1} \ldots x_T} \right) =$$

$$\sum_{x_1 - x_{t-1}} \left[\sum_{x_{t+1} \ldots x_T} F_1(x_1 \ldots x_{t-1}, x_t) \cdot F_2(x_t, x_{t+1} \ldots x_T) \right]$$

$$= \left[\sum_{x_1} \sum_{x_2} \ldots \sum_{x_{t-1}} \psi_{12} \ldots \psi_{t-1,t} \right] \times \left[\sum_{x_{t+1}} \ldots \sum_{x_n} \psi_{t,t+1} \ldots \psi_{T-1,T} \right]$$

$$\sum_{x_{t-1}}\sum_{x_{t-2}}\cdots\sum_{x_1}\left(\psi_{12}\cdot\psi_{23}\cdot\ldots\cdot\psi_{t-1,t}\right) =$$

$$\sum_{x_{t-1}}\cdots\sum_{x_2}\left(\psi_{23}\cdots\psi_{t-1,t}\left(\sum_{x_1}\psi_{12}\right)\right) =$$

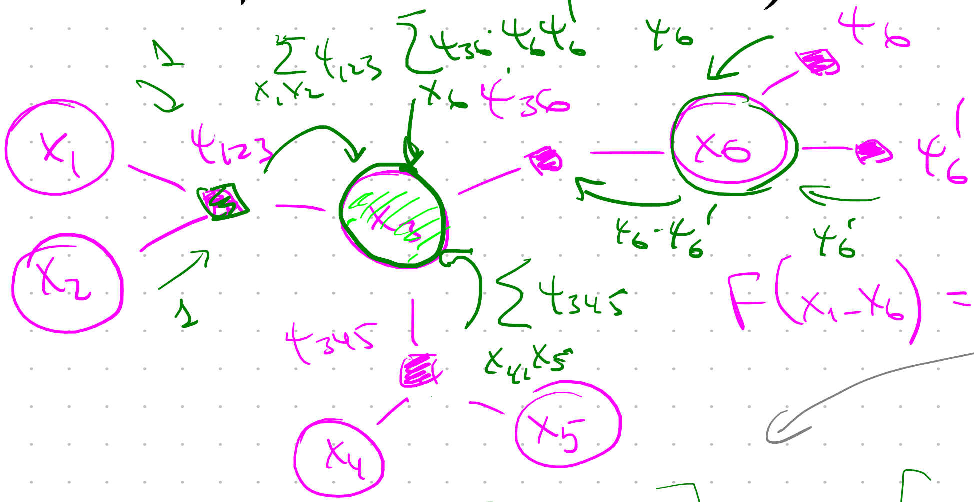
$$= \sum_{x_{t-1}}\cdots\sum_{x_3}\psi_{34}\cdots\psi_{t-1,t}\left[\sum_{x_2}\psi_{23}\left[\sum_{x_1}\psi_{12}\right]\right] = \ldots$$

$$= \sum_{x_{t-1}}\psi_{t-1,t}\left[\sum_{x_{t-2}}\psi_{t-2,t-1}\left[\cdots\sum_{x_3}\psi_{34}\left[\sum_{x_2}\psi_{23}\left[\sum_{x_1}\psi_{12}\right]\cdots\right]\right]\right]$$

$$\sum_{x_{t+1}}\cdots\sum_{x_n}\psi_{t,t+1}\psi_{t+1,t+2}\cdots\psi_{T-2,T-1}\psi_{T-1,T} = \ldots$$

$$= \sum_{x_{t+1}}\psi_{t,t+1}\left[\sum_{x_{t+2}}\psi_{t+1,t+2}\left[\cdots\sum_{x_{T-1}}\psi_{T-2,T-1}\left[\sum_{x_T}\psi_{T-1,T}\right]\cdots\right]\right]$$

$$f(x_t) = F_1(x_t)\cdot F_2(x_t)$$



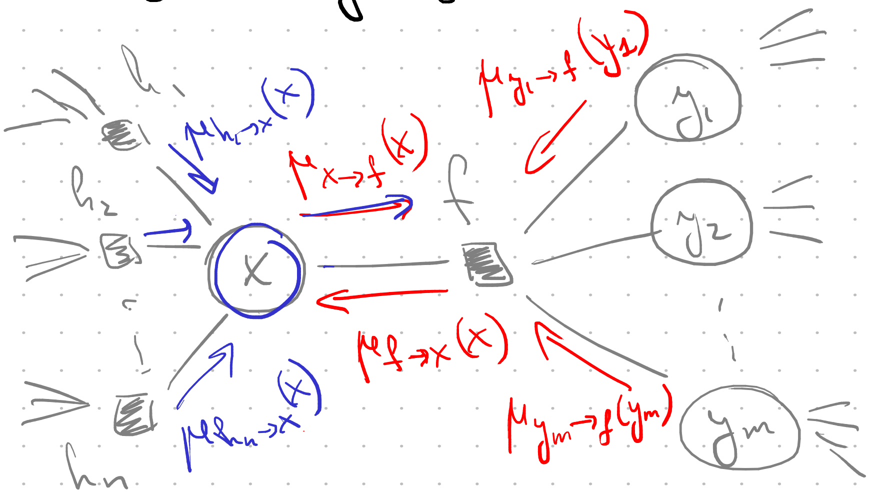
$$F(x_1 - x_6) = \psi_{123}\cdot\psi_{345}\cdot\psi_{36}\cdot\psi_6\cdot\psi_6'$$

$f(x_3)$

$$f(x_3) = \sum_{x_1,x_2,x_4,x_5,x_6}\psi_{123}\psi_{345}\left[\psi_{36}\psi_6\psi_6'\right] = \sum_{x_1-x_5}\left[\psi_{123}\cdot\psi_{345}\left[\sum_{x_6}\psi_{36}\psi_6\psi_6'\right]\right]$$

$$= \left[\sum_{x_6}\psi_{36}\psi_6\psi_6'\right]\cdot\left[\sum_{x_1,x_2}\psi_{123}\right]\cdot\left[\sum_{x_4,x_5}\psi_{345}\right]$$

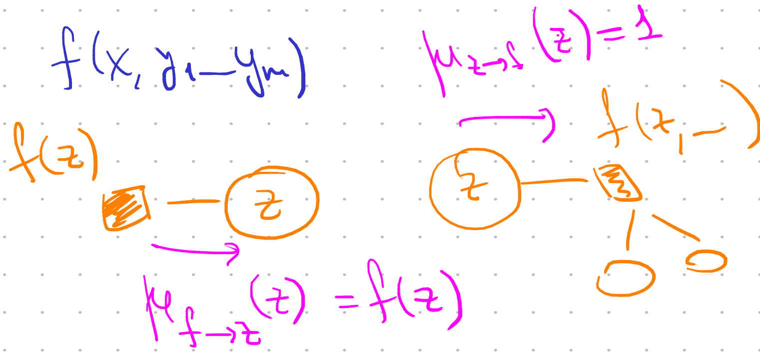
6 Message passing algorithm



$$F(x_1 \dots x_N) = \prod_{k=1}^{K} f_k(X_k)$$

$$f(x) = \sum_{x_1 \dots \cancel{x} \dots x_N} \prod_{k=1}^{K} f_k(X_k)$$

$f(x, y_1 \dots y_m)$

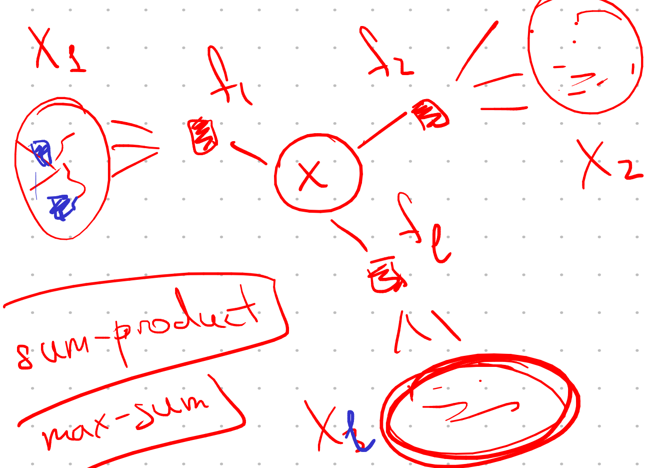


$$\mu_{x \to f}(x) = \prod_{i=1}^{n} \mu_{h_i \to x}(x)$$

[illegible]

$$\mu_{f \to x}(x) = \sum_{y_1 \dots y_m} f(x, y_1 \dots y_m) \cdot \mu_{y_1 \to f}(y_1) \dots \mu_{y_m \to f}(y_m)$$

$$f(x) = \sum_{x_1 - x_N} F(x, x_1, \dots, x_N) = \sum_{x_1 - x_N} \overbrace{f_1(x, X_1) \cdot F_1(X_1)}^{X_1} \dots \overbrace{f_\ell(x, X_\ell) F_\ell(X_\ell)}^{X_\ell}$$



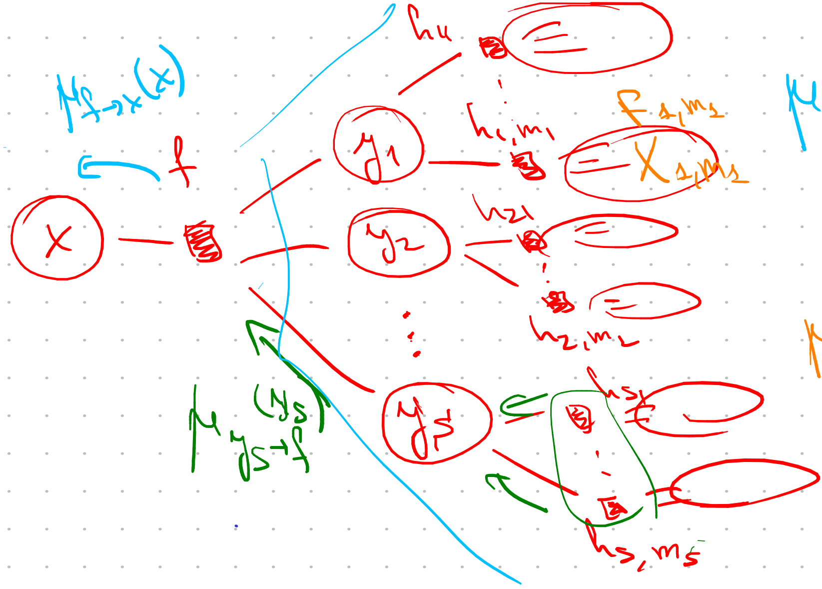
$$= \overbrace{\left(\sum_{X_1} f_1(x, X_1) F_1(X_1) \right)}^{\mu_{f_1 \to x}(x)} \times$$

$$\times \overbrace{\left(\sum_{X_2} f_2(x, X_2) F_2(X_2) \right)}^{\mu_{f_2 \to x}(x)} \times \dots$$

$$\times \underbrace{\left(\sum_{X_\ell} f_\ell(x, X_\ell) F_\ell(X_\ell) \right)}_{\mu_{f_\ell \to x}(x)}$$

sum-product

max-sum



$$\mu_{f\to x}(x) \stackrel{?}{=} \sum_X f(x,X)\cdot F'(X)$$

Дано:

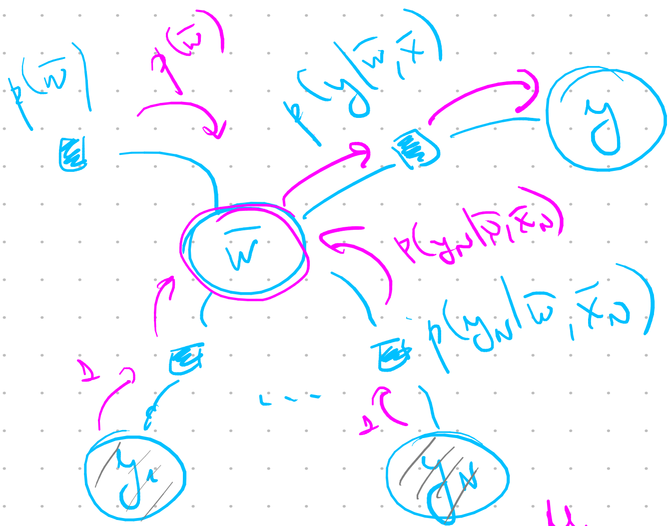
$$\mu_{h_{ij}\to y_i}(y_i) = \sum_{X_{ij}} h_{ij}(y_i, X_{ij})\cdot F_{ij}(X_{ij})$$

$$\mu_{f\to x}(x) = \sum_{y_1 \dots y_S} f(x, y_1 \dots y_S)\cdot \prod_{s=1}^{S} \underline{\mu_{y_s\to f}(y_s)} =$$

$$= \sum_{y_1 \dots y_S} f(x, y_1 \dots y_S)\cdot \prod_{i=1}^{S}\prod_{j=1}^{m_i} \mu_{h_{ij}\to y_i}(y_i) =$$

$$= \sum_{y_1 \dots y_S} f(x, y_1 \dots y_S)\cdot \prod_{i=1}^{S}\prod_{j=1}^{m_i} \underline{\left[\sum_{X_{ij}} h_{ij}(y_i, X_{ij})\cdot F_{ij}(X_{ij})\right]} =$$

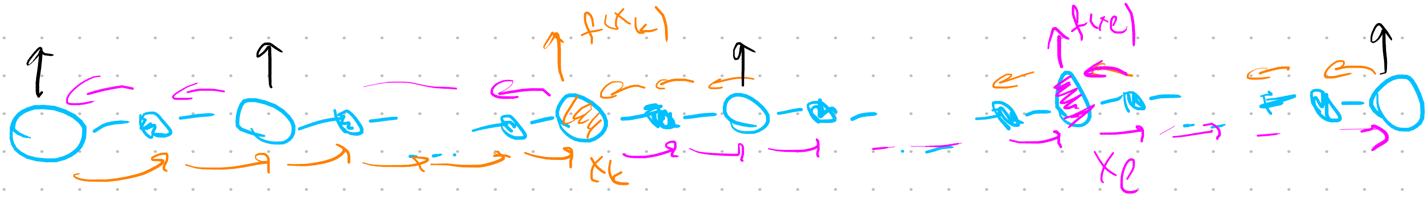
$$= \sum_{y_1 \dots y_S}\ \sum_{X_{11}\dots X_{S,m_S}} f(x, y_1 \dots y_S)\cdot \underbrace{\prod_{i=1}^{S}\prod_{j=1}^{m_i} h_{ij}(y_i, X_{ij})\, F_{ij}(X_{ij})}_{=\ F'(X)}$$

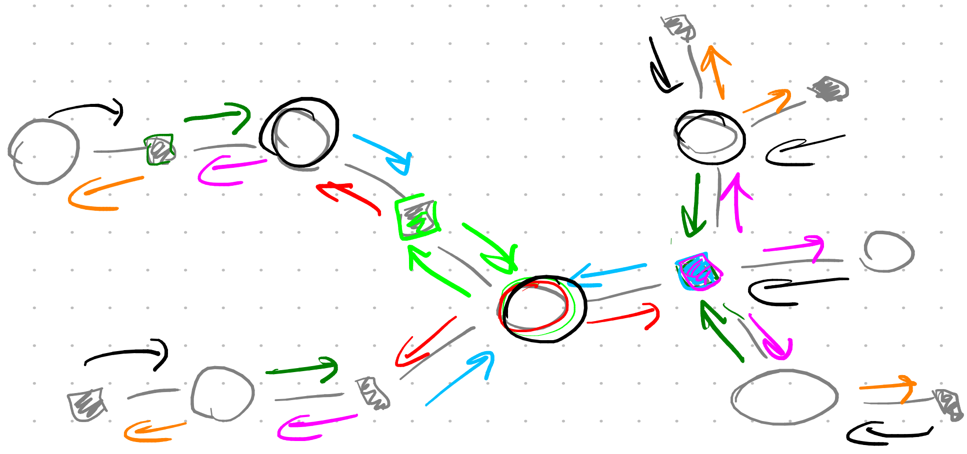


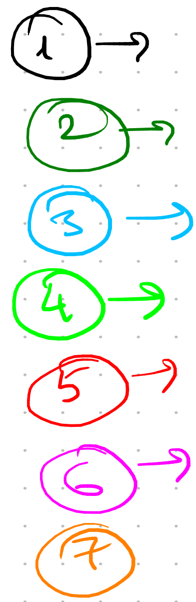
$$\underline{p(y|y_1 \dots y_N) = \int p(y, \bar{w}|y_1 \dots y_N)\, d\bar{w}}$$

$$\mu_{\bar{w}\to p(y|\bar{w},\bar{x})}(\bar{w}) = \underline{p(\bar{w})\cdot \prod_n p(y_n|\bar{w},\bar{x}_n)}$$

$$\mu_{p\to y}(y) = \underline{\int p(\bar{w})\cdot \prod_n p(y_n|\bar{w},\bar{x}_n)\cdot p(y|\bar{w},\bar{x})\, d\bar{w}}$$







– фактор-граф – дерево
– факторы „простые“

← message passing

не дерево → можно сделать деревом

дерево, НО факторы сложные → Expectation Propagation

Approximate Inference

- Loopy Belief Propagation
- Sampling
- Variational approx.