① а) Дивергенция Йенсена - Шеннона - симметризованная и сглаженная версия дивергенции Кульбака - Лейблера: $JSD(p\|q) = \frac{1}{2} KL(p\| \frac{p+q}{2}) + \frac{1}{2} KL(q \| \frac{p+q}{2})$

Свойства: $JSD(p\|q) \in [0, 1]$, $JSD(p\|q) = JSD(q\|p)$

$JSD(p\|q)$ показывает насколько близки распределения p и q

$JSD(p\|q)$ определена всегда

б) $KL(p\|q) = \int p(x) \ln \frac{p(x)}{q(x)} dx = - \int p(x) \ln \frac{q(x)}{p(x)} dx \geq$

$- \log \left(\int p(x) \cdot \frac{q(x)}{p(x)} dx \right) = - \log 1 = 0 \Rightarrow KL(p\|q) \geq 0 \Rightarrow$

$JSD(p\|q) \geq 0$

в) Модели плохо сравнивать по KL или JSD, т.к. вторая модель имеет имеет большую дисперсию. Используем S-score: $S\text{-}score = \exp\left(-\frac{1}{2}(\mu_1 - \mu_2)^T (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)\right) =$

$= \exp\left(-\frac{1}{2} \begin{pmatrix} 9 \\ 4 \end{pmatrix}^T \begin{pmatrix} 7000{,}1 & 0{,}1 \\ 0{,}1 & 5000{,}1 \end{pmatrix}^{-1} \begin{pmatrix} 9 \\ 4 \end{pmatrix}\right) \approx e^{-0{,}0071} \approx 1 \Rightarrow$ нельзя

считать разными

② $X_i \sim \sum_{j=1}^{k} \pi_j N(m, \sigma_j^2)$; $m \sim N(m_0, \sigma_0^2)$, $\frac{1}{\sigma_j^2} \sim \Gamma(\alpha_j, \beta_j)$

а) $p(X, m, \sigma^2 | \alpha, \beta, m_0, \sigma_0^2, \pi) = \prod_{i=1}^{n} \left(\sum_{j=1}^{k} \pi_j N(x_i | m, \sigma_j^2) \right) \cdot$

$\cdot \prod_{j=1}^{k} \Gamma\left(\frac{1}{\sigma_j^2} | \alpha_j, \beta_j\right) N(m | m_0, \sigma_0^2)$

б) $p(m, \sigma^2 | X, \alpha, \beta, m_0, \sigma_0^2, \pi) = \frac{p(X, m, \sigma^2 | \alpha, \beta, m_0, \sigma_0^2, \pi)}{p(X | \alpha, \beta, m_0, \sigma_0, \pi)} \propto p(X, m, \sigma^2 |$

$\alpha, \beta, m_0, \sigma_0^2, \pi) = \prod_{i=1}^{n} \left(\sum_{j=1}^{k} \pi_j \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_i - m)^2}{2\sigma^2}\right) \right) \cdot \prod_{j=1}^{k} \frac{1}{\Gamma(\alpha_j)} \beta_j^{\alpha_j} \cdot$

$\cdot \left(\frac{1}{\sigma_j^2}\right)^{\alpha_j - 1} \exp\left(-\frac{\beta_j}{\sigma_j^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(m - m_0)^2}{2\sigma_0^2}\right)$

в) $p(X, m, \sigma^2, Z | \alpha, \beta, m_0, \sigma_0, \pi) = \prod_{i=1}^{n} \left(\prod_{j=1}^{k} \pi_j N(m, \sigma_j^2 | \alpha, \beta, m_0, \sigma_0^2) \right)^{z_{ij}} \cdot$

$\prod_{j=1}^{k} \Gamma\left(\frac{1}{\sigma_j^2} | \alpha_j, \beta_j\right) N(m | m_0, \sigma_0^2)$

2) $q(m, \sigma^2, Z) = q(m) q(Z) q(\sigma^2)$

$\log q(m) \propto \mathbb{E}_{q \setminus m} \log p(X, m, \sigma^2, Z | \alpha, \beta, m_0, \sigma_0^2, \pi) \propto \mathbb{E}_{q \setminus m} \left[\sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \cdot \left(-\frac{(x_i - m)^2}{2\sigma_j^2}\right) \right.$

$\left. - \frac{(m - m_0)^2}{2\sigma_0^2} \right] \propto m^2 \cdot \frac{1}{2} \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{k} \mathbb{E}_{q \setminus m} \frac{-z_{ij}}{\sigma_j^2}}_{A_m} - m \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{k} \mathbb{E}_{q / m} - \frac{z_{ij} x_i}{\sigma_j^2}}_{B_m} - m^2 \cdot \frac{1}{2\sigma_0^2} +$

$+ m \frac{m_0}{\sigma_0^2} = m^2 \frac{1}{2}\left(A_m - \frac{1}{\sigma_0^2}\right) + m\left(\frac{m_0}{\sigma_0^2} - B_m\right) = \frac{1}{2}\left(A_m - \frac{1}{\sigma_0^2}\right)\left(m^2 + 2 \cdot m\right.$

$\left. \left(\frac{m_0}{\sigma_0^2} - B_m\right)\left(A_m - \frac{1}{2\sigma_0^2}\right)^{-1}\right) \Rightarrow q(m) \sim N\left(\left(\frac{m_0}{\sigma_0^2} - B_m\right)\left(A_m - \frac{1}{2\sigma_0^2}\right)^{-1}, \left(A_m - \frac{1}{2\sigma_0^2}\right)^{-1}\right)$

$\log q(Z) \propto \mathbb{E}_{q \setminus Z} \log p(X, m, \sigma^2, Z | \alpha, \beta, m_0, \sigma_0^2, \pi) \propto \mathbb{E}_{q / Z} \left[\sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \right($

$\left. \log \pi_j - \frac{1}{2} \log 2\pi - \log \sigma_j - \frac{(x_i - m)^2}{2\sigma_j^2} \right) \Big] \propto \sum_{i=1}^{n} \sum_{j=1}^{k} z_{ij} \left(\log \pi_j - \frac{1}{2} \log(2\pi) - \right.$

$\left. - \frac{1}{2} \mathbb{E}_{q/Z} \log \sigma_j^2 - \frac{1}{2} \mathbb{E}_{q/Z} \frac{(x_i - m)^2}{2\sigma_j^2} \right)$

$\log q(\sigma^2) \propto \mathbb{E}_{q/\sigma^2} \log p(X, m, \sigma^2, Z | \alpha, \beta, m_0, \sigma_0^2, \pi) \propto \mathbb{E}_{q/\sigma^2} \left[\sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij} \cdot \right.$

$\left(\frac{\log \sigma_j^2}{2} - \frac{(x_i - m)^2}{2\sigma_j^2} \right) + \sum_{j=1}^{k} -(\alpha_j - 1) \log \sigma_j^2 - \frac{\beta_j}{\sigma_j^2} \propto \sum_{j=1}^{k} \frac{1}{\sigma_j^2} \left(-\sum_{i=1}^{n} \mathbb{E}_{q/\sigma^2} \frac{(x_i - m)^2}{2} z_{ij} - \right.$

$\left. - \beta_j \right) - \log \sigma_j^2 \left(\sum_{i=1}^{n} \mathbb{E}_{q/\sigma^2} \frac{z_{ij}}{2} + (\alpha_j - 1) \right)$

Использование обычного EM-алгоритма мешает сложность $q(Z, m, \sigma^2)$

g) $p(X \mid \alpha, \beta, m_0, \sigma_0, \bar{\pi}) \to \max\limits_{\alpha, \beta, m_0, \sigma_0, \bar{\pi}}$

$F(\bar{\pi}, \alpha, \beta, m_0, \sigma_0) = \mathbb{E}_q \log p(X, m, \sigma^2, Z \mid \alpha, \beta, \sigma_0^2, \bar{\pi}) = \mathbb{E}_q \Big[\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{k} z_{ij} \cdot$

$(\log \pi_j - \frac{1}{2}\log(2\pi) - \log \sigma_j - \frac{(x_i - m)^2}{2\sigma_j^2}) + \sum\limits_{j=1}^{k} (-\log \Gamma(\alpha_j) + \alpha_j \log \beta_j -$

$- 2(\alpha_j - 1)\log \sigma_j - \frac{\beta_j}{\sigma_j^2}) - \frac{1}{2}\log(2\pi) - \log \sigma_0 - \frac{(m - m_0)^2}{2\sigma_0^2} \Big]$

$$\frac{\partial F}{\partial \alpha_j} = -\frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} + \log \beta_j - 2\mathbb{E}_q \log \sigma_j = 0 \Rightarrow \frac{\Gamma'(\alpha_j)}{\Gamma(\alpha_j)} = \log \beta_j - 2\mathbb{E}_q \log \sigma_j$$

$$\frac{\partial F}{\partial \beta_j} = \frac{\alpha_j}{\beta_j} - \mathbb{E}_q \frac{1}{\sigma_j^2} = 0 \Rightarrow \beta_j = \frac{\alpha_j}{\mathbb{E}_q \frac{1}{\sigma_j^2}}$$

$$\frac{\partial F}{\partial m_0} = -\frac{m_0}{\sigma_0^2} + \mathbb{E}_q \frac{m}{\sigma_0^2} = 0 \Rightarrow m_0 = \mathbb{E}_q m$$

$$\frac{\partial F}{\partial \sigma_0} = -\frac{1}{\sigma_0} + \mathbb{E}_q \frac{(m - m_0)^2}{\sigma_0^3} = 0 \Rightarrow \sigma_0^2 = \mathbb{E}_q (m - m_0)^2$$

$$\frac{\partial F}{\partial \pi_j} = \mathbb{E}_q \sum_{i=1}^{n} z_{ij} \frac{1}{\pi_j} - \lambda = 0 \Rightarrow \pi_j = \frac{\mathbb{E}_q \sum_{i=1}^{n} z_{ij}}{\sum_{j=1}^{k} \mathbb{E}_q \sum_{i=1}^{n} z_{ij}}$$

Для повышения разреженности надо ввести распределение на π

(3) Независимость: $p(x,y) = p(x)p(y)$

Условная независимость: $p(y,z|x) = p(y|x)p(z|x)$

Независимость $\not\Rightarrow$ условная независимость

$$\begin{tikzcd} X & Y \arrow[l] \\ Z \arrow[u] & \end{tikzcd}$$

$p(x,y,z) = p(z)p(y)p(x|z,y)$

$p(y,z) = \int p(y)p(z)p(x|y,z)\,dx = p(y)p(z) \Rightarrow$

y и z независимы

$p(x|y,z) \neq p(z|x)p(y|x)$, т.к. $p(x|y,z) \neq p(x|y)p(x|z) \Rightarrow$

y и z условно зависимы при x

Условная независимость $\not\Rightarrow$ независимость

$$\begin{tikzcd} X \arrow[r] \arrow[dr] & Y \\ & Z \end{tikzcd}$$

$p(x,y,z) = p(x)p(y|x)p(z|x)$

$p(y,z) = \int p(x)p(y|x)p(z|x)\,dx \neq p(y)p(z) \Rightarrow$

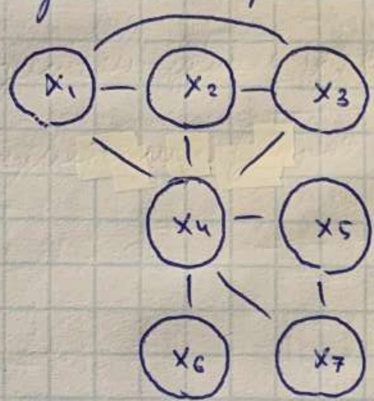
y и z зависимы

$p(y,z|x) = p(z|x)p(y|x,z) = p(z|x)p(y|x) \Rightarrow$ y и z условно независимы при x.

④ Ориентированные — с ориентированными рёбрами в графе зависимостей.

Неориентированные — с неориентированными рёбрами в графе зависимостей

1. Соединяем родителей каждой вершины
2. Удаляем ориентацию



⑤ Марковское одеяло - узлы в графической модели, которые изолируют данный узел от остальной части графа. Если известны значения узлов в марковском одеяле, то узел условно независим относительно других узлов

Ориентированная модель: родители узла, дети узла, родители детей узла

Неориентированная модель: соседи узла

6

(7) $p(a,b,c) = \frac{1}{Z} \psi_{ab}(a,b)\, \psi_{bc}(b,c); \quad a,b,c \in \{0,1\}$

a	b	c	Z
0	0	0	100
0	0	1	100
0	1	0	0,2
0	1	1	1,4
1	0	0	5
1	0	1	5
1	1	0	0,4
1	1	1	2,8

1. $p(a,c|b) = p(a|b)\,p(c|b)$, т.к. a и c не соединены ребром, а все остальные переменные это b =>

$a,c \mid b$ – условно независимы

$a,b \mid c$ – условно зависимы, т.к. есть ребро

$a - b$

$b,c \mid a$ – условно зависимы, т.к. есть ребро $b - c$

2. $p(a=1 \mid b=0, c=1) = \frac{5}{5+100} = 0{,}048$

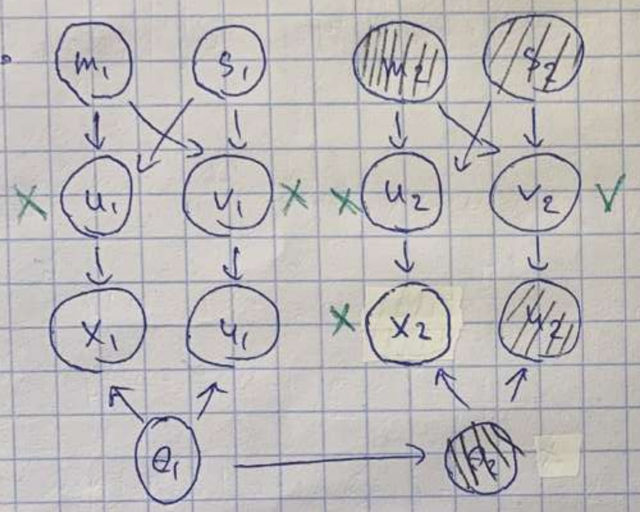
$p(c=1 \mid b=1, a=0) = \frac{1{,}4}{1{,}4+0{,}2} = 0{,}875$

$p(c=0 \mid a=1, b=1) = \frac{0{,}4}{0{,}4+2{,}8} = 0{,}125$

$p(c=0, a=1, b=1) = \frac{0{,}4}{214{,}8} = 0{,}002$

(7) 1. $X_1 \perp X_2$ — нет, т.к. $C = \emptyset$, то в первый пункт теоремы мы никогда не попадём, и во второй пункт теоремы мы не попадём, т.к. на пути $X_1 \to \theta_1 \to \theta_2 \to X_2$ нет перед-перед

2. —

3. 

в u_2 и x_2 три пути блокируют s_2, m_2, θ_2 (хвост-хвост)

в u_1, и V_2 все пути блокируют x_1 или или y_1 (перед-перед)

4. —

(9) а) Да, задают. При преобразовании ориентированного графа в неориентированный новых ребер не появляется => задают одно и то же

б) Нет, не задают. При преобразовании ориентированного графа в неориентированный появляются новые ребра => не задают одно и то же