1 Without Mean Field approximation

Задана выборка:

$$\mathcal{D} = \{\mathbf{x}_i\}_{i=1\dots m}$$

Задана модель:

$$p(\mathbf{x}, \mathbf{z}|\theta) = p(\mathbf{z}|\mathbf{x}, \theta)\, p(\mathbf{x}|\theta) \tag{1}$$

Параметр θ ищется из условия:

$$p(\mathbf{x}|\theta) \to \max_{\theta} \tag{2}$$

Рассмотрим следующий функционал:

$$\mathcal{L}(q, \theta) = \log p(\mathbf{x}|\theta) - \mathrm{D}_{KL}\left(q(\mathbf{z})\, ||p(\mathbf{z}|\mathbf{x}, \theta)\right) \tag{3}$$

Заменим оптимизируемый функционал 2 на функционал 3:

$$\mathcal{L}(q, \theta) = \log p(\mathbf{x}|\theta) - \mathrm{D}_{KL}\left(q(\mathbf{z})\, ||p(\mathbf{z}|\mathbf{x}, \theta)\right) \to \max_{q \in Q, \theta} \tag{4}$$

Функционал 4 оптимизируется при помощи EM-алгоритма.

EM-algorithm Итерируемся до сходимости:

1. E-step:

$$q^t(\mathbf{z}) = \arg\max_{q \in Q} \mathcal{L}\left(q, \theta^{t-1}\right)$$

2. M-step:

$$\theta^t = \arg\max_{\theta} \mathcal{L}\left(q^t, \theta\right)$$

2 With Mean Field approximation

Пусть для вариационного распределения $q(\mathbf{z})$ выполняется следующее:

$$q(\mathbf{z}) = \prod_{i=1}^{d} q_i(z_i) \tag{5}$$

Найдем $\mathcal{L}(q, \theta)$ используя 5:

$$\begin{aligned}\mathcal{L}(q, \theta) &= \log p(\mathbf{x}|\theta) - \mathrm{D}_{KL}\left(q(\mathbf{z})\, ||p(\mathbf{z}|\mathbf{x}, \theta)\right) = \\ &= \log p(\mathbf{x}|\theta) - \mathrm{D}_{KL}\left(\prod_{i=1}^{d} q_i(z_i)\, ||p(\mathbf{z}|\mathbf{x}, \theta)\right) = \\ &= \log p(\mathbf{x}|\theta) - \int \prod q_i(z_i) \log \frac{\prod q_i(z_i)}{p(\mathbf{z}|\mathbf{x}, \theta)} d\mathbf{z}\end{aligned} \tag{6}$$

Минимизируем $\mathcal{L}(q, \theta)$ по $q_k(z_k)$:

$$\begin{aligned}\mathcal{L}(q, \theta) &=^{q_k} \log p(\mathbf{x}|\theta) - \int \prod q_i(z_i) \log \frac{\prod q_i(z_i)}{p(\mathbf{z}|\mathbf{x}, \theta)} d\mathbf{z} = \\ &= -\int \prod q_i(z_i) \log \frac{\prod q_i(z_i)}{p(\mathbf{x}, \mathbf{z}|\theta)} d\mathbf{z} \int \to \max_{q_k},\end{aligned} \tag{7}$$

где $=^{q_k}$ обозначает, что в текущий момент выполняется максимизация по k-й переменной и следовательно все остальные параметры являются константой относительно q_k поэтому мы эту константу занулям, так-как она нам не интересна.

$$
\begin{aligned}
\mathcal{L}(q,\theta) &=^{q_k} -\int \prod q_i(z_i)\log\frac{\prod q_i(z_i)}{p(\mathbf{x},\mathbf{z}|\theta)}d\mathbf{z} =^{q_k}\\
&=^{q_k} -\sum_{i=1}^{d}\int\prod_{i=1}^{d} q_i(z_i)\log q_i(z_i)\,d\mathbf{z} + \int\prod_{i=1}^{d} q_i(z_i)\log p(\mathbf{x},\mathbf{z}|\theta)\,d\mathbf{z} =\\
&=^{q_k} -\int\prod_{i=1}^{d} q_i(z_i)\log q_k(z_k)\,d\mathbf{z} + \int\prod_{i=1}^{d} q_i(z_i)\log p(\mathbf{z}|\mathbf{x})\,d\mathbf{z} + \cancel{\text{const}} =\\
&= -\int q_k(\mathbf{z}_k)\left[\log q_k(z_k) - \int\prod_{i\neq k} q_k(z_k)\log p(\mathbf{x},\mathbf{z}|\theta)\,d\mathbf{z}_{/k}\right]dz_k =\\
&=^{q_k} -\int q_k(z_k)\log\frac{q_k(z_k)}{\exp\left(\mathsf{E}_{/k}\log p(\mathbf{x},\mathbf{z}|\theta)\right)}
\end{aligned}
\tag{8}
$$

Получаем решение E-шага в случае Mean-Field апроксимации:

$$
\log q_k^t(z_k) \propto \mathsf{E}_{q/q_k}\log p\left(\mathbf{x},\mathbf{z}|\theta^{t-1}\right) \tag{9}
$$

Теперь минимизируем $\mathcal{L}(q,\theta)$ по θ:

$$
\begin{aligned}
\mathcal{L}(q,\theta) &=^{\theta} \log p(\mathbf{x}|\theta) - \int\prod q_i(z_i)\log\frac{\prod q_i(z_i)}{p(\mathbf{z}|\mathbf{x},\theta)}d\mathbf{z} =\\
&= -\int\prod q_i(z_i)\log\frac{\prod q_i(z_i)}{p(\mathbf{x},\mathbf{z}|\theta)}d\mathbf{z}\int \to \max_{\theta},
\end{aligned}
\tag{10}
$$

где $=^{\theta}$ обозначает, что в текущий момент выполняется максимизация по переменной θ.

$$
\begin{aligned}
\mathcal{L}(q,\theta) &=^{\theta} -\int\prod q_i(z_i)\log\frac{\prod q_i(z_i)}{p(\mathbf{x},\mathbf{z}|\theta)}d\mathbf{z} =\\
&=^{\theta}\int q(\mathbf{z})\log p(\mathbf{x},\mathbf{z}|\theta)\,d\mathbf{z} = \mathsf{E}_q\log p(\mathbf{x},\mathbf{z}|\theta)
\end{aligned}
\tag{11}
$$

Получаем решение M-шага в случае Mean-Field апроксимации:

$$
\mathsf{E}_{q^t}\log p(\mathbf{x},\mathbf{z}|\theta) \to \max_{\theta} \tag{12}
$$

EM-algorithm Итерируемся до сходимости:

1. E-step:
$$
\log q_k^t(z_k) \propto \mathsf{E}_{q/q_k}\log p\left(\mathbf{x},\mathbf{z}|\theta^{t-1}\right)
$$

2. M-step:
$$
\mathsf{E}_{q^t}\log p(\mathbf{x},\mathbf{z}|\theta) \to \max_{\theta}
$$