

Практическое задание по ЕМ-алгоритму

Солоткий Михаил, 417 группа ВМК МГУ

17 ноября 2018 г.

1 Вывод формул

- Апостериорное распределение на координаты лица:

$$q_k(\mathbf{d}_k) = p(\mathbf{d}_k \mid \mathbf{X}_k, \boldsymbol{\theta}, \mathbf{A}) = \frac{p(\mathbf{X}_k \mid \mathbf{d}_k, \boldsymbol{\theta})p(\mathbf{d}_k \mid \mathbf{A})}{\int p(\mathbf{X}_k \mid \mathbf{d}_k, \boldsymbol{\theta})p(\mathbf{d}_k \mid \mathbf{A})d(\mathbf{d}_k)} = \frac{p(\mathbf{X}_k \mid \mathbf{d}_k, \boldsymbol{\theta})p(\mathbf{d}_k \mid \mathbf{A})}{\sum_{i=0}^{H-h} \sum_{j=0}^{W-w} p(\mathbf{X}_k \mid \mathbf{d}_k, \boldsymbol{\theta})p(\mathbf{d}_k \mid \mathbf{A})}$$
$$q(\mathbf{d}) = \prod_{k=1}^K q_k(\mathbf{d}_k)$$

- Точечные оценки параметров на М-шаге:

$$\begin{aligned} \mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A}) &= \mathbb{E}_{q(\mathbf{d})} \ln p(\mathbf{X}, \mathbf{d} \mid \boldsymbol{\theta}, \mathbf{A}) = \int q(\mathbf{d}) \ln p(\mathbf{X}, \mathbf{d} \mid \boldsymbol{\theta}, \mathbf{A})d(\mathbf{d}) = \\ &= \int q(\mathbf{d}) \sum_{k=1}^K \ln p(\mathbf{X}_k, \mathbf{d}_k \mid \boldsymbol{\theta}, \mathbf{A})d(\mathbf{d}) = \sum_{k=1}^K \int q(\mathbf{d}) \ln p(\mathbf{X}_k, \mathbf{d}_k \mid \boldsymbol{\theta}, \mathbf{A})d(\mathbf{d}) = \\ &= \sum_{k=1}^K \int q_k(\mathbf{d}_k) \ln p(\mathbf{X}_k, \mathbf{d}_k \mid \boldsymbol{\theta}, \mathbf{A})d(\mathbf{d}_k) = \sum_{k=1}^K \int q_k(\mathbf{d}_k) \ln p(\mathbf{X}_k \mid \mathbf{d}_k, \boldsymbol{\theta})d(\mathbf{d}_k) + \\ &+ \sum_{k=1}^K \int q_k(\mathbf{d}_k) \ln p(\mathbf{d}_k \mid \mathbf{A})d(\mathbf{d}_k) = \sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \left[\ln p(\mathbf{X}_k \mid \mathbf{d}_k = (i, j), \boldsymbol{\theta}) + \right. \\ &\left. + \ln \mathbf{A}(i, j) \right] \end{aligned}$$

Максимизируем по $\{\mathbf{A}, \mathbf{B}, \mathbf{F}, s^2\}$:

$$\begin{cases} \mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A}) \rightarrow \max_{\mathbf{A}} \\ \mathbf{A}(i, j) \geq 0 \\ \sum_{i,j} \mathbf{A}(i, j) = 1 \end{cases}$$
$$\begin{aligned} L(q, \boldsymbol{\theta}, \mathbf{A}, \boldsymbol{\lambda}, \mu) &= \sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \left[\ln p(\mathbf{X}_k \mid \mathbf{d}_k = (i, j), \boldsymbol{\theta}) + \ln \mathbf{A}(i, j) \right] + \sum_{i,j} \lambda_{ij} \mathbf{A}(i, j) - \\ &- \mu \left(\sum_{i,j} \mathbf{A}(i, j) - 1 \right) \end{aligned}$$

Необходимые условия максимума Каруша-Куна-Таккера:

$$\begin{cases} \nabla_{\mathbf{A}} L(q, \boldsymbol{\theta}, \mathbf{A}, \boldsymbol{\lambda}, \mu) = 0 \\ \mathbf{A}(i, j) \geq 0 \\ \sum_{i,j} \mathbf{A}(i, j) = 1 \\ \lambda_{ij} \geq 0 \\ \lambda_{ij} \mathbf{A}(i, j) = 0 \end{cases}$$

$$\nabla_{\mathbf{A}(i,j)} L(q, \boldsymbol{\theta}, \mathbf{A}, \boldsymbol{\lambda}, \mu) = \sum_{k=1}^K q_k(\mathbf{d}_k = (i, j)) \frac{1}{\mathbf{A}(i, j)} + \lambda_{ij} - \mu = 0$$

Т. к. $\mathbf{A}(i, j) > 0$ (для существования $\ln \mathbf{A}(i, j)$), $\lambda_{ij} = 0$

$$\mu = \sum_{k=1}^K q_k(\mathbf{d}_k = (i, j)) \frac{1}{\mathbf{A}(i, j)}$$

$$\mathbf{A}(i, j) = \sum_{k=1}^K q_k(\mathbf{d}_k = (i, j)) \frac{1}{\mu}$$

$$1 = \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} \sum_{k=1}^K q_k(\mathbf{d}_k = (i, j)) \frac{1}{\mu}$$

$$\mathbf{A}(i, j) = \frac{\sum_{k=1}^K q_k(\mathbf{d}_k = (i, j))}{\sum_{i=0}^{H-h} \sum_{j=0}^{W-w} \sum_{k=1}^K q_k(\mathbf{d}_k = (i, j))} = \frac{\sum_{k=1}^K q_k(\mathbf{d}_k = (i, j))}{K}$$

$$\begin{aligned} \mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A}) &\rightarrow \max_{\mathbf{B}} \\ &\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \left[\sum_{m=0}^{H-1} \sum_{n=0}^{W-1} \ln \mathcal{N}(\mathbf{X}_k(m, n) \mid \mathbf{B}(m, n), s^2) \left[(m < i) \vee (n < j) \vee \right. \right. \\ &\left. \vee (m \geq i + h) \vee (n \geq j + w) \right] \Big] \rightarrow \max_{\mathbf{B}} \\ &\left\{ \ln \mathcal{N}(x \mid \mu, \sigma^2) = -\frac{1}{2} \ln(2\pi) - \ln |\sigma| - \frac{(x-\mu)^2}{2\sigma^2} \right\} \\ &\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \left[\sum_{m=0}^{H-1} \sum_{n=0}^{W-1} (\mathbf{X}_k(m, n) - \mathbf{B}(m, n))^2 \left[(m < i) \vee (n < j) \vee (m \geq i + h) \vee \right. \right. \\ &\left. \left. \vee (n \geq j + w) \right] \right] \rightarrow \min_{\mathbf{B}} \\ \mathbf{B}(m, n) &= \frac{\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \mathbf{X}_k(m, n) \left[(m < i) \vee (n < j) \vee (m \geq i + h) \vee (n \geq j + w) \right]}{\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \left[(m < i) \vee (n < j) \vee (m \geq i + h) \vee (n \geq j + w) \right]} \end{aligned}$$

$$\begin{aligned} \mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A}) &\rightarrow \max_{\mathbf{F}} \\ &\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \sum_{t=0}^{h-1} \sum_{u=0}^{w-1} \ln p(\mathbf{X}_k(i+t, j+u) \mid \mathbf{F}(t, u), s^2) \rightarrow \max_{\mathbf{F}} \\ &\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \sum_{t=0}^{h-1} \sum_{u=0}^{w-1} (\mathbf{X}_k(i+t, j+u) - \mathbf{F}(t, u))^2 \rightarrow \min_{\mathbf{F}} \\ \mathbf{F}(t, u) &= \frac{\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \mathbf{X}_k(i+t, j+u)}{\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j))} = \frac{\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \mathbf{X}_k(i+t, j+u)}{K} \end{aligned}$$

$$\begin{aligned} \mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A}) &\rightarrow \max_{\mathbf{s}^2} \\ &\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} \left[\ln \mathcal{N}(\mathbf{X}_k(m, n) \mid \mathbf{B}(m, n), s^2) \left[(m < i) \vee (n < j) \vee (m \geq i + h) \vee \right. \right. \\ &\left. \left. \vee (n \geq j + w) \right] + \ln \mathcal{N}(\mathbf{X}_k(m, n) \mid \mathbf{F}(m-i, n-j), s^2) \left(1 - \left[(m < i) \vee (n < j) \vee (m \geq i + h) \vee \right. \right. \right. \\ &\left. \left. \left. \vee (n \geq j + w) \right] \right) \right] \rightarrow \max_{\mathbf{s}^2} \\ &\left\{ \ln(\mathcal{N}(x \mid \mu, \sigma^2))'_{\sigma} = \left(-\frac{1}{2} \ln(2\pi) - \ln(\sigma) - \frac{(x-\mu)^2}{2\sigma^2} \right)'_{\sigma} = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3} = 0; \sigma^2 = (x - \mu)^2 \right\} \\ s^2 &= \frac{\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} (\mathbf{X}_k(m, n) - \mathbf{B}(m, n))^2}{KHW} \left[(m < i) \vee (n < j) \vee (m \geq i + h) \vee \right. \\ &\left. \vee (n \geq j + w) \right] + \frac{\sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} (\mathbf{X}_k(m, n) - \mathbf{F}(m-i, n-j))^2}{KHW} \left(1 - \left[(m < i) \vee \right. \right. \\ &\left. \left. \vee (n < j) \vee (m \geq i + h) \vee (n \geq j + w) \right] \right) \end{aligned}$$

- М-шаг в случае EM-MAP:
 $\mathbb{E}_{q(\mathbf{d})} \ln p(\mathbf{X}, \mathbf{d} \mid \boldsymbol{\theta}, \mathbf{A}) \rightarrow \max_{\boldsymbol{\theta}, \mathbf{A}}$
 $(i_k^*, j_k^*) = \operatorname{argmax}(q_k(\mathbf{d}_k))$

$$\sum_{k=1}^K \ln p(\mathbf{X}_k, \mathbf{d}_k = (i_k^*, j_k^*) \mid \boldsymbol{\theta}, \mathbf{A}) + \sum_{k=1}^K \ln p(\mathbf{d}_k = (i_k^*, j_k^*) \mid \mathbf{A}) \rightarrow \max_{\boldsymbol{\theta}, \mathbf{A}}$$

- Точечные оценки параметров на М-шаге в EM-МАР (выведем их как частный случай уже выведенных выше оценок):

$$\mathbf{A}(i, j) = \frac{\sum_{k=1}^K \mathbb{I}[i = i_k^*, j = j_k^*]}{K}$$

На М-шаге могут появиться нули в матрице \mathbf{A} . Они не сломают вычисление максимальных значений остальных параметров на М-шаге, так как те не зависят от матрицы \mathbf{A} . Они также не повлияют на пересчёт вероятностей апостериорного распределения на Е-шаге на следующей итерации, так как те числа из матрицы \mathbf{A} , которые участвуют в формуле: $(i = i_k^*, j = j_k^*)$, как видно из формулы, больше 0.

$$\mathbf{B}(m, n) = \frac{\sum_{k=1}^K \mathbf{X}_k(m, n) \mathbb{I}[(m < i_k^*) \vee (n < j_k^*) \vee (m \geq i_k^* + h) \vee (n \geq j_k^* + w)]}{\sum_{k=1}^K \mathbb{I}[(m < i_k^*) \vee (n < j_k^*) \vee (m \geq i_k^* + h) \vee (n \geq j_k^* + w)]}$$

В результате вычисления выражения в некоторых случаях появляется неопределённость $\frac{0}{0}$, это случается, когда в исходном максимизируемом выражении не участвует $\mathbf{B}(m, n)$ и в таком случае параметр $\mathbf{B}(m, n)$ не обновляется.

$$\mathbf{F}(m, n) = \frac{\sum_{k=1}^K \mathbf{X}_k(i_k^* + m, j_k^* + n)}{K}$$

$$s^2 = \frac{\sum_{k=1}^K \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} (\mathbf{X}_k(m, n) - \mathbf{B}(m, n))^2}{KHW} \mathbb{I}[(m < i_k^*) \vee (n < j_k^*) \vee (m \geq i_k^* + h) \vee (n \geq j_k^* + w)] + \\ + \frac{\sum_{k=1}^K \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} (\mathbf{X}_k(m, n) - \mathbf{F}(m - i_k^*, n - j_k^*))^2}{KHW} \left(1 - \mathbb{I}[(m < i_k^*) \vee (n < j_k^*) \vee (m \geq i_k^* + h) \vee (n \geq j_k^* + w)]\right)$$

- ELBO:

$$\begin{aligned} \mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A}) &= \sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \left[\ln p(\mathbf{X}_k \mid \mathbf{d}_k = (i, j), \boldsymbol{\theta}) + \ln \mathbf{A}(i, j) \right] - \\ &- \sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \ln q_k(\mathbf{d}_k = (i, j)) = \\ &= \sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \left[\sum_{m=0}^{H-1} \sum_{n=0}^{W-1} \left(\ln \mathcal{N}(\mathbf{X}_k(m, n) \mid \mathbf{B}(m, n), s^2) \mathbb{I}[(m < i) \vee (n < j) \vee \right. \right. \\ &\vee (m \geq i + h) \vee (n \geq j + w)] + \ln \mathcal{N}(\mathbf{X}_k(m, n) \mid \mathbf{F}(m - i, n - j), s^2) \left. \left(1 - \mathbb{I}[(m < i) \vee (n < j) \vee (m \geq i + h) \vee \right. \right. \\ &\vee (n \geq j + w)] \right) \right] + \ln \mathbf{A}(i, j) - \sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \ln q_k(\mathbf{d}_k = (i, j)) = \\ &= -\frac{1}{2} KHW \ln(2\pi s^2) - \sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \left[\frac{1}{2s^2} \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} \left((\mathbf{X}_k(m, n) - \mathbf{B}(m, n))^2 \mathbb{I}[(m < i) \vee \right. \right. \\ &\vee (n < j) \vee (m \geq i + h) \vee (n \geq j + w)] + (\mathbf{X}_k(m, n) - \mathbf{F}(m - i, n - j))^2 \left. \left(1 - \mathbb{I}[(m < i) \vee (n < j) \vee (m \geq i + h) \vee \right. \right. \\ &\vee (n \geq j + w)] \right) \right] + \ln \mathbf{A}(i, j) - \sum_{k=1}^K \sum_{i=0}^{H-h} \sum_{j=0}^{W-w} q_k(\mathbf{d}_k = (i, j)) \ln q_k(\mathbf{d}_k = (i, j)) \end{aligned}$$

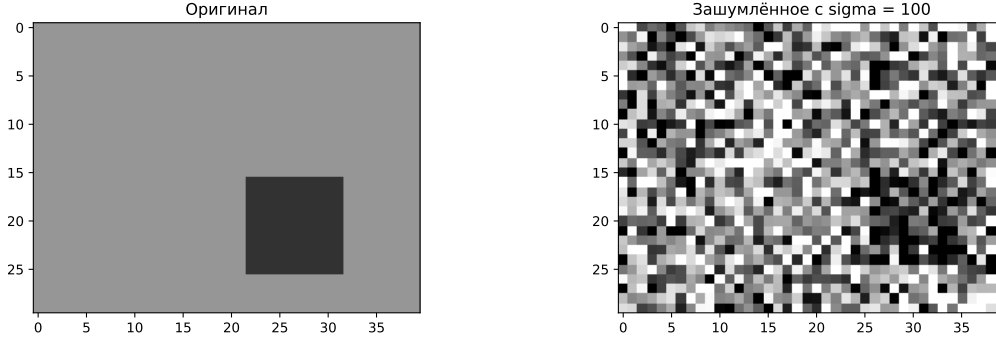
- ELBO для EM-МАР:

$$\begin{aligned} \mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A}) &= -\frac{1}{2} KHW \ln(2\pi s^2) - \sum_{k=1}^K \left[\frac{1}{2s^2} \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} \left((\mathbf{X}_k(m, n) - \mathbf{B}(m, n))^2 \mathbb{I}[(m < i_k^*) \vee (n < j_k^*) \vee \right. \right. \\ &\vee (m \geq i_k^* + h) \vee (n \geq j_k^* + w)] + (\mathbf{X}_k(m, n) - \mathbf{F}(m - i_k^*, n - j_k^*))^2 \left. \left(1 - \mathbb{I}[(m < i_k^*) \vee (n < j_k^*) \vee \right. \right. \\ &\vee (m \geq i_k^* + h) \vee (n \geq j_k^* + w)] \right) \right] + \ln \mathbf{A}(i_k^*, j_k^*) \end{aligned}$$

2 Эксперименты на синтетических данных

2.1 Генерация данных

В данном пункте описывается общая схема генерации выборки, здесь не фиксируются параметры, которые могут меняться в разных экспериментах, а просто обозначаются соответствующими буквами. Генерировалась выборка из K зашумлённых изображений размером 40×30 . Зашумлённые пиксели могут принимать дробные значения. Визуализация происходила с помощью функции `plt.imshow` с параметрами $v_min=0$, $v_max = 255$. Это означает, что все значения больше либо равные 255 изображались белыми, а значения меньше либо равные 0 – чёрными. Приведение к единому масштабу было сделано для удобной визуализации результата работы алгоритма. В качестве F – изображения с переменными координатами вершины, использовался квадрат 10×10 , заполненный значением 50. Фон был также монотонный и заполнен значением 150.

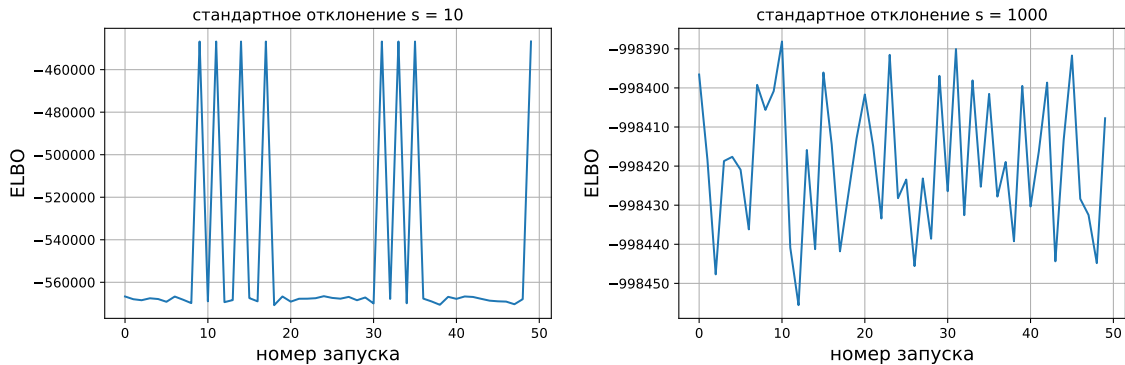


2.2 Инициализация параметров

Чтобы получались разные начальные приближения, значения пикселей параметров F , B генерировались независимо из равномерного распределения на отрезке $[0, 255]$ (предполагается, что пиксели изображений имеют именно такую область значений). Значения матрицы A генерировались из равномерного распределения на отрезке $[1, 2]$ (чтобы избежать нулевых значений), а затем перенормировались, чтобы сгенерированные значения имели смысл вероятностей. Истинные значения координат левых верхних вершин изображения генерировались из равномерного категориального распределения на множестве всех возможных положений этих вершин. Начальное приближение для стандартного отклонения не генерировалось, а просто было положено равным 1.

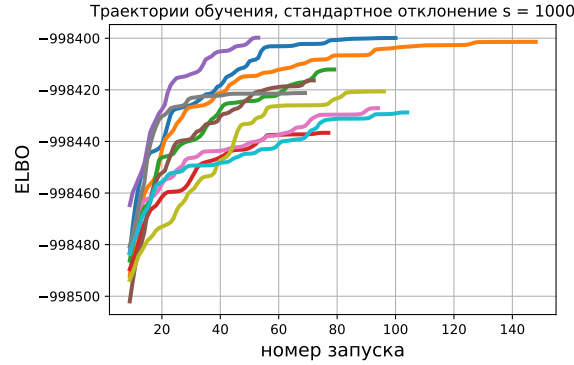
2.3 Зависимость от начального приближения

В данном эксперименте выборка имело размер $K = 100$. Использовались разные значения стандартного отклонения s : 10 и 1000, чтоб проверить зависимость также от уровня зашумления. Для каждого значения s ЕМ-алгоритм запускался по 50 раз из разных начальных приближений. Задавалось большое количество итераций: $2 \cdot 10^9$, для того, чтобы предотвратить ранний выход из алгоритма в случае тяжело оптимизируемого функционала. Выход происходил, когда значения ELBO на соседних итерациях отличались не больше, чем на $\text{tolerance} = 10^{-3}$. Запуск



Как можно видеть, в обоих случаях алгоритм значения функционала в результате отличаются сильно больше, чем на 10^{-3} , что свидетельствует о том, что локальные максимумы, к которым сошёлся алгоритм, разные. Для большего уровня зашумления наблюдается большее разнообразие значения $\log\text{-likelihood}$ в

точках локального максимума, то есть чем больше зашумление, тем больше запусков может понадобиться для отыскания лучшего локального максимума. Далее также представлен график ELBO по итерациям для разных запусков, на котором также видно, что за фиксированное число шагов метод из разных начальных приближений может достигать разных значений и в результате останавливаться в разных локальных максимумах.



2.4 Качество восстановления при разных уровнях зашумления и размере выборки

Стоит ожидать, что при одном и том же уровне зашумления, качество восстановления изображений будет не хуже, если добавить в выборку данных. Возможно начиная с некоторого K данных уже достаточно и проблема будет в большом зашумлении, и при увеличении выборки качество будет расти незначительно. Как было выяснено в предыдущем пункте, для сильно зашумлённых изображений может понадобиться много рестартов, чтобы сойтись к хорошему локальному максимуму, поэтому если ограничивать количество рестартов, может так оказаться, что на большей выборке качество хуже из-за неподходящего случайного начального приближения. При одном и том же объёме выборки и увеличении параметра s стоит ожидать ухудшение качества.

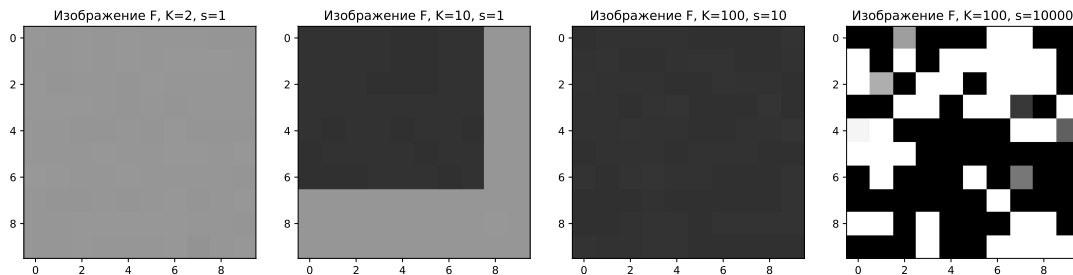
Введём двумерную сетку параметров s и K : $K \in \{1, 10, 100, 1000\}$, $s \in \{1, 10, 100, 1000, 10000\}$. Ниже в таблице представлено $\frac{\mathcal{L}(q, \theta, \mathcal{A})}{K}$ в результате запуска ЕМ-алгоритма с 20 рестартами и максимально 100 итерациями на каждом рестарте, tolerance = 10^{-3} , как и прежде. Результаты приведены в таблице ниже, жирным выделены максимальные с своих столбцов значения.

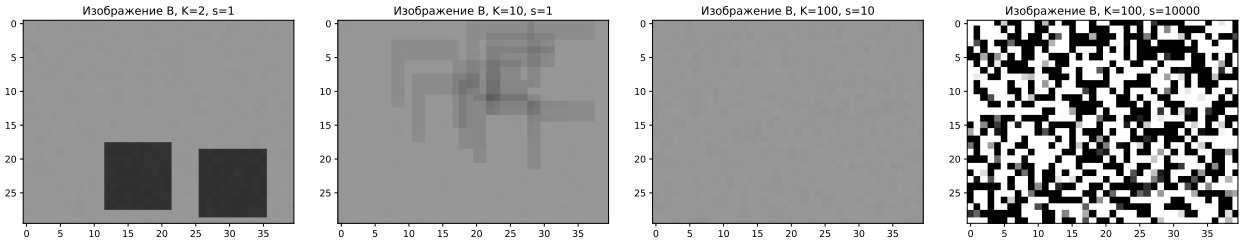
	$s = 1$	$s = 10$	$s = 100$	$s = 1000$	$s = 10000$
$K = 2$	-1280	-4702	-6817	-9516	-12292
$K = 10$	-1636	-5529	-7198	-9907	-12691
$K = 100$	-1699	-4464	-7225	-9988	-12748
$K = 1000$	-1708	-4469	-7234	-9991	-12754

Таблица 1: Значения $\frac{\mathcal{L}(q, \theta, \mathcal{A})}{K}$

Судя по результатам, при увеличении зашумлённости качество действительно падает, переход с $K = 100$ к $K = 1000$ не даёт значительного прироста log-likelihood, но в случае одного и того же значения s и при различных малых K иногда лучшими по нормированному log-likelihood становятся меньшие значения K . Возможно на малых выборках неполное правдоподобие ведёт себя по-другому, и для плохо восстановленных изображений log-likelihood всё-равно получается высоким.

Приведём несколько восстановленных вариантов изображений F и B .





Как можно видеть восстановление изображений при $K = 2$ и $K = 10$ далеко от истинного, но значение нормированного \log -likelihood выше, чем у варианта с $K = 100$. То есть недостаточно просто смотреть на нормированный \log -likelihood, надо ещё как минимум учитывать размер выборки и не сравнивать при малом и большом размере выборки полученные значения \log -likelihood. В случае $s = 10000$ несмотря на относительно большой размер выборки: $K = 1000$ результаты алгоритма получаются совсем странными.

2.5 Сравнение с hard-EM

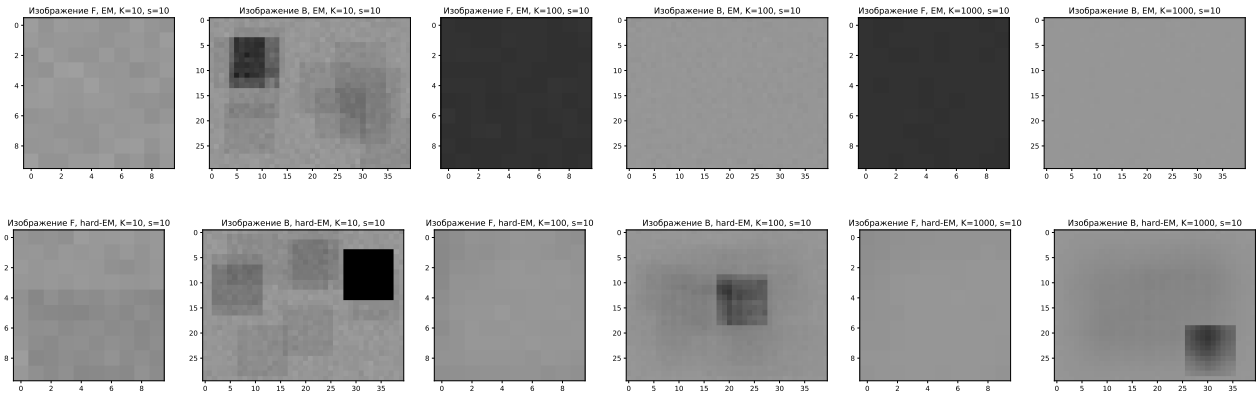
Ожидается, что hard-EM будет работать в разы быстрее за счёт того, что в нём на М-шаге функции с числом арифметических операций $c \cdot HWK$ заменены на функции с $c \cdot K$ операциями, где c - некоторая константа. По качеству ожидается, что hard-EM будет заметно хуже, так как игнорируется информация о возможном положении координат внутреннего изображения F.

Для замера времени запускался EM-алгоритм без рестартов при значении $s = 10$ и $K \in \{10, 100, 1000\}$. Для каждой пары параметров происходило 10 запусков, время усреднялось и считалось стандартное отклонение по 10 запускам. $\text{tolerance} = 10^{-3}$, максимальное количество итераций равно 50.

	$K = 10$	$K = 100$	$K = 1000$
EM	$1.12 \text{ s} \pm 357 \text{ ms}$	$4.69 \text{ s} \pm 4.15 \text{ s}$	$2\text{min } 3\text{s} \pm 1\text{min } 47\text{s}$
hard-EM	$194 \text{ ms} \pm 11.9 \text{ ms}$	$1.74 \text{ s} \pm 496 \text{ ms}$	$48.9 \text{ s} \pm 11.8 \text{ s}$

Таблица 2: Время работы EM-алгоритма без рестартов, $s = 10$

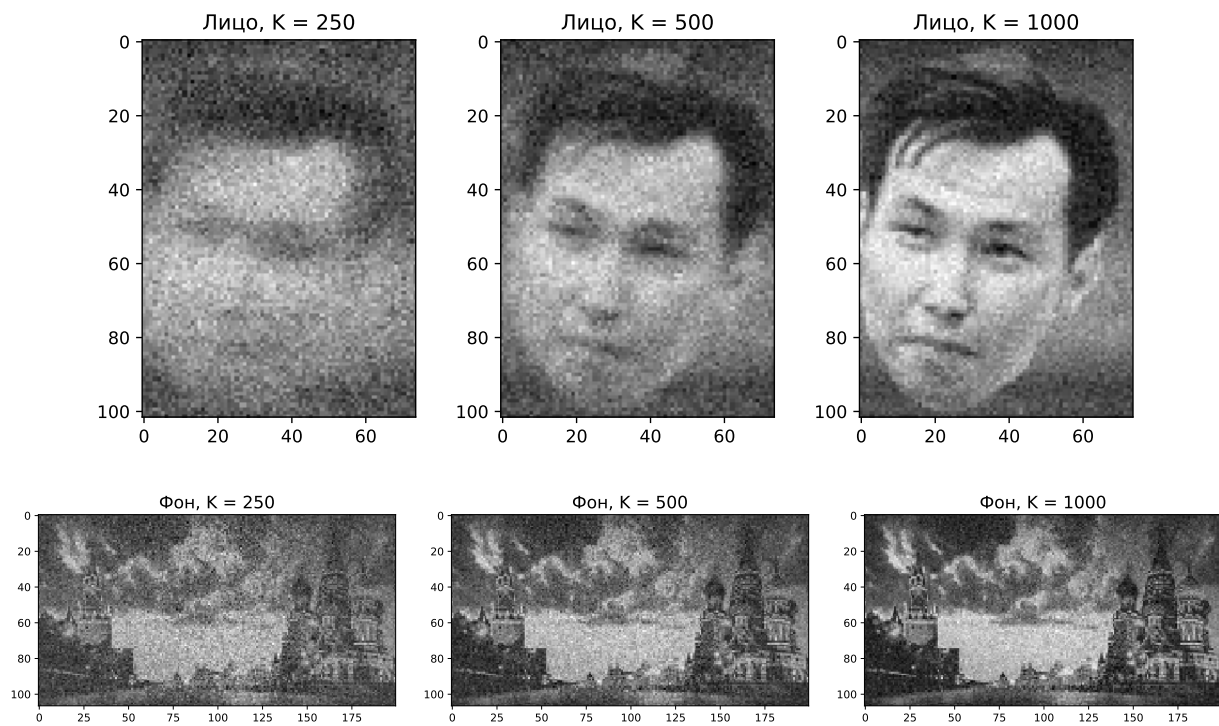
Результаты подтвердили ожидания. Также было замечено, что в случае обычного EM и $K = 1000$ наблюдается большой разброс во времени работы, что говорит о том, что не только результат неустойчиво зависит от начального приближения, но и количество итераций, необходимое для сходимости. Приведём несколько примеров восстановленных изображений для тех же значений параметров.



Даже при достаточно большом размере выборки ($K = 1000$) и при малой стандартном отклонении шума ($s = 10$) hard-EM плохо справляется с восстановлением изображений (F серое вместо чёрного, а на изображении фона есть следы чёрного квадрата), при этом обычный EM-алгоритм даже без рестартов хорошо справляется с задачей при $K \in \{100, 1000\}$.

3 Разоблачение преступника

EM-алгоритм запускался на подвыборках размеров $\{250, 500, 1000\}$ без рестартов. Результаты восстановления лица и фона представлены ниже. Преступником оказался Айбек Аланов.



4 Возможные пути улучшения схемы

Самое больше время занимает М-шаг, в котором операции имеют линейную сложность от объёма выборки. Для ускорения М-шага можно брать батч изображений и искать $\arg \max_{\mathbf{A}, \boldsymbol{\theta}} \mathcal{L}(q, \mathbf{A}, \boldsymbol{\theta})$ не по полной, а по батчу. Разве что в таком случае нет гарантий, что исходный log-likelihood будет монотонно возрастать.