

Московский государственный университет имени М.В. Ломоносова

Факультет вычислительной математики и кибернетики

Отчёт по второму практическому заданию: ЕМ-алгоритм для детектива

Вариант для факультета ВМК

Каратыщев Дмитрий Иванович, 417 группа

Дата выполнения работы:
4 декабря 2024 г.

Содержание

1	Постановка задачи	2
2	Формулировка задания	2
3	Теоретическая часть	3
3.1	Задание 1. Вывод формул апостериорного распределения на координаты лица на изображениях.	3
3.2	Задание 2. Вывод формул точечных оценок на параметры $A, \bar{\theta} = \{F, B, s^2\}$ на М-шаге для ЕМ и МАР-ЕМ алгоритмов	3
3.2.1	Вывод точечной оценки на параметр A	4
3.2.2	Вывод точечной оценки на параметр F	5
3.2.3	Вывод точечной оценки на параметр B	5
3.2.4	Вывод точечной оценки на параметр s^2	6
3.3	Задание 3. Вывод формулы для подсчёта нижней оценки на логарифм неполного правдоподобия $\mathcal{L}(q, \theta, A)$	7
4	Практическая часть	7
4.1	Эксперимент 1. Тестирование на синтетической выборке	7
4.2	Эксперимент 2. Перебор различных значений отклонения и размера выборки	9
4.3	Эксперимент 3. Проверка ЕМ и hard ЕМ алгоритмов	10
4.4	Эксперимент 4. Применяем к реальным данным	11
4.5	Возможные улучшения ЕМ-алгоритма	12

1 Постановка задачи

Дана выборка $\mathbb{X} = \{X_k\}_{k=1}^K$ сильно зашумленных черно-белых изображений размера $H \times W$ пикселей. Каждое из этих изображений содержит один и тот же неподвижный фон и лицо преступника в неизвестных координатах, при этом лицо попадает в любое изображение целиком. Будем считать, что изображение лица имеет прямоугольную форму размера $h \times w$ пикселей, где $h = 86$, $w = 64$.

Введем следующие обозначения:

- $X_k(i, j)$ — пиксель k -ого изображения;
- $B \in \mathbb{R}^{H \times W}$ — изображение чистого фона без лица преступника, $B(i, j)$ — пиксель этого изображения;
- $F \in \mathbb{R}^{h \times w}$ — изображение лица преступника, $F(i, j)$ — пиксель этого изображения;
- $\bar{d}_k = (d_k^h, d_k^w)$ — координаты верхнего левого угла изображения лица на k -ом изображении (d_k^h — по вертикали, d_k^w — по горизонтали), $\bar{d} = (\bar{d}_1, \dots, \bar{d}_K)$ — набор координат для всех изображений выборки.

Также будем считать шум на изображении независимым для каждого пикселя и принадлежащим нормальному распределению $\mathcal{N}(0, s^2)$, где s — стандартное отклонение. Таким образом, для одного изображения имеем:

$$p(X_k | \bar{d}_k, \bar{\theta}) = \prod_{ij} \begin{cases} \mathcal{N}(X_k(i, j) | F(i - d_k^h, j - d_k^w), s^2), & \text{если } (i, j) \in \text{faceArea}(\bar{d}_k), \\ \mathcal{N}(X_k(i, j) | B(i, j), s^2) & \text{иначе} \end{cases}$$

Здесь $\bar{\theta} = \{B, F, s^2\}$, $\text{faceArea}(\bar{d}_k) = \{(i, j) | d_k^h \leq i \leq d_k^h + h - 1, d_k^w \leq j \leq d_k^w + w - 1\}$.

Распределение на неизвестные координаты лица на изображении зададим общим для всех изображений с помощью матрицы параметров $A \in \mathbb{R}^{(H-h+1) \times (W-w+1)}$ следующим образом:

$$p(\bar{d}_k | A) = A(d_k^h, d_k^w), \quad \sum_{ij} A(i, j) = 1$$

где $A(i, j)$ — элемент матрицы A .

В итоге имеем следующую совместную вероятностную модель:

$$p(\mathbb{X}, \bar{d} | \bar{\theta}, A) = \prod_k p(X_k | \bar{d}_k, \bar{\theta}) p(\bar{d}_k | A).$$

2 Формулировка задания

Требуется решить задачу

$$p(\mathbb{X} | \bar{\theta}, A) \rightarrow \max_{\bar{\theta}, A}.$$

Для этого предлагается воспользоваться ЕМ-алгоритмом, то есть перейти к следующей задаче оптимизации нижней оценки на логарифм неполного правдоподобия:

$$\mathcal{L}(q, \bar{\theta}, A) = \mathbb{E}_{q(\bar{d})} \log p(\mathbb{X}, \bar{d} | \bar{\theta}, A) - \mathbb{E}_{q(\bar{d})} \log q(\bar{d}) \rightarrow \max_{q, \bar{\theta}, A}$$

На Е-шаге вычисляется оценка на апостериорное распределение на координаты лица на изображениях:

$$q(\bar{d}) = p(\bar{d} | \mathbb{X}, \bar{\theta}, A) = \prod_k p(\bar{d}_k | X_k, \bar{\theta}, A),$$

а на М-шаге вычисляется точечная оценка на параметры $\bar{\theta}, A$:

$$\mathbb{E}_{q(\bar{d})} \log p(\mathbb{X}, \bar{d} | \bar{\theta}, A) \rightarrow \max_{\bar{\theta}, A}.$$

Также далее будет рассматриваться упрощенный вариант ЕМ-алгоритма, который называется hard ЕМ. В нем после Е-шага берется не все апостериорное распределение на координаты лица на изображениях, а только МАР-оценка на эти координаты (то есть после Е-шага $q(\bar{d})$ преобразовывают так, что для каждого изображения X_k оценка $q(\bar{d}_k)$ принимает значение 1 только в одной точке — точке аргмаксимума апостериорного распределения $p(\bar{d}_k | X_k, \bar{\theta}, A)$). При выполнении итераций ЕМ алгоритма нужно следить за значением оптимизируемого функционала $\mathcal{L}(q, \bar{\theta}, A)$.

3 Теоретическая часть

3.1 Задание 1. Вывод формул апостериорного распределения на координаты лица на изображениях.

Выведем формулы для подсчета апостериорного распределения на координаты лица на изображениях $p(\bar{d}_k | X_k, \bar{\theta}, A)$ на Е-шаге. Определим множество $\mathcal{D} = \{(i, j) \mid i \in \overline{0, H-h}, j \in \overline{0, W-h}\}$

$$p(\bar{d}_k | X_k, \bar{\theta}, A) = [\text{Воспользуемся теоремой Байеса}] = \frac{p(X_k | \bar{d}_k, \bar{\theta})p(\bar{d}_k | A)}{\sum_{\bar{d}_k \in \mathcal{D}} p(X_k | \bar{d}_k, \bar{\theta})p(\bar{d}_k | A)}.$$

Все распределения, использующиеся в данной формуле, даны в условии задачи и могут быть спойно реализованы численно.

3.2 Задание 2. Вывод формул точечных оценок на параметры $A, \bar{\theta} = \{F, B, s^2\}$ на М-шаге для ЕМ и МАР-ЕМ алгоритмов

Выведем формулы точечных оценок на параметры $A, \bar{\theta} = \{F, B, s^2\}$ на М-шаге для ЕМ и МАР-ЕМ алгоритмов в порядке A, F, B, s^2 .

На М-шаге ЕМ-алгоритма проводится максимизация нижней оценки на неполное правдоподобие, или же максимизация математического ожидания логарифма полного правдоподобия с полученным распределением на латентные переменные на Е-шаге. Обозначим распределение на латентные переменные \bar{d} через q . Напомним, что $\mathcal{D}^K = \underbrace{\mathcal{D} \times \mathcal{D} \times \dots \times \mathcal{D}}_K$, а также $q(\bar{d}) = p(\bar{d} | \mathbb{X}, A, \bar{\theta})$. Тогда

$$\mathbb{E}_q \log p(\mathbb{X}, \bar{d} | A, \bar{\theta}) = \sum_{\bar{d} \in \mathcal{D}^K} q(\bar{d}) \log p(\mathbb{X}, \bar{d} | A, \bar{\theta})$$

Факторизуем полное правдоподобие и распределение на латентные переменные по отдельным объектам и перепишем сумму:

$$\begin{aligned} \mathbb{E}_q \log p(\mathbb{X}, \bar{d} | A, \bar{\theta}) &= \sum_{\bar{d}_1 \in \mathcal{D}} \dots \sum_{\bar{d}_K \in \mathcal{D}} q(\bar{d}_1) \cdot \dots \cdot q(\bar{d}_K) \sum_{k \in \overline{1, K}} \log p(X_k, \bar{d}_k | A, \bar{\theta}) = \\ &= \sum_{\bar{d}_1 \in \mathcal{D}} \dots \sum_{\bar{d}_K \in \mathcal{D}} q(\bar{d}_1) \cdot \dots \cdot q(\bar{d}_K) \sum_{k \in \overline{1, K}} \log p(X_k | \bar{d}_k, \bar{\theta}) \cdot p(\bar{d}_k | A) = \mathcal{L}_1(\bar{\theta}) + \mathcal{L}_2(A), \\ \mathcal{L}_1(\bar{\theta}) &= \sum_{\bar{d}_1 \in \mathcal{D}} \dots \sum_{\bar{d}_K \in \mathcal{D}} q(\bar{d}_1) \cdot \dots \cdot q(\bar{d}_K) \sum_{k \in \overline{1, K}} \log p(X_k | \bar{d}_k, \bar{\theta}) \\ \mathcal{L}_2(A) &= \sum_{\bar{d}_1 \in \mathcal{D}} \dots \sum_{\bar{d}_K \in \mathcal{D}} q(\bar{d}_1) \cdot \dots \cdot q(\bar{d}_K) \sum_{k \in \overline{1, K}} \log A(d_k^h, d_k^w) \end{aligned}$$

Переставим знаки суммирования и заметим, что $\forall j \in \overline{1, K}$ выполнено:

$$\sum_{\bar{d}_1 \in \mathcal{D}} \dots \sum_{\bar{d}_{j-1} \in \mathcal{D}} \sum_{\bar{d}_{j+1} \in \mathcal{D}} \dots \sum_{\bar{d}_K \in \mathcal{D}} \prod_{i \neq j} q(\bar{d}_i) = 1,$$

так как маргинальные распределения тоже факторизуются по отдельным объектам и суммируются в единицу. Получим следующие равенства:

$$\mathcal{L}_1(\bar{\theta}) = \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \cdot \log p(X_k \mid \bar{d}_k, \bar{\theta})$$

$$\mathcal{L}_2(A) = \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \cdot \log A(d_k^h, d_k^w)$$

3.2.1 Вывод точечной оценки на параметр A

Введём следующие обозначения:

$$I = \overline{0, H-h}, \quad J = \overline{0, W-w}, \quad g_{ij}(A) = -A(i, j), \quad h(A) = \sum_{i \in I} \sum_{j \in J} A(i, j) - 1,$$

$$U = \{A \in \mathbb{R}^{(H-h+1) \times (W-w+1)} \mid g_{ij}(A) \leq 0 \ \forall i \in I, \ \forall j \in J; \ h(A) = 0\}$$

Для точечной оценки параметра A решим следующую задачу:

$$\hat{A} = \underset{A \in U}{\operatorname{argmax}} (\mathbb{E}_q \log p(\mathbb{X}, \bar{d} \mid A, \bar{\theta})) = \underset{A \in U}{\operatorname{argmax}} (\mathcal{L}_1(\bar{\theta}) + \mathcal{L}_2(A)) = \underset{A \in U}{\operatorname{argmax}} \mathcal{L}_2(A)$$

Пусть $f(A) = -\mathcal{L}_2(A)$. Перепишем задачу в следующем виде:

$$\underset{A \in U}{\operatorname{argmin}} f(A)$$

Функционал f является выпуклым, допустимое множество также является выпуклым. Воспользуемся условиями Каруша-Куна-Таккера для решения данной задачи. Локальный минимум \hat{A} можно найти из следующих условий на функцию Лагранжа $L(A; \lambda, \mu)$, если существуют соответствующие λ^*, μ^* :

$$\begin{cases} L(A; \lambda, \mu) = f(A) + \sum_{i \in I} \sum_{j \in J} \lambda_{ij} g_{ij}(A) + \mu h(A), \\ \nabla_A L(\hat{A}; \lambda^*, \mu^*) = 0, \\ \lambda^* \geq 0, \\ \lambda_{ij}^* g_{ij}(\hat{A}) = 0, \forall i \in I, \ \forall j \in J, \\ h(\hat{A}) = 0 \end{cases}$$

Перейдём к записи градиента функции Лагранжа по A :

$$\frac{\partial f}{\partial A(i, j)} = - \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \frac{\partial}{\partial A(i, j)} q(\bar{d}_k) \cdot \log A(d_k^h, d_k^w) = - \frac{\sum_{k \in \overline{1, K}} q_k(i, j)}{A(i, j)}$$

$$\frac{\partial L(\hat{A}; \lambda^*, \mu^*)}{\partial A(i, j)} = - \frac{\sum_{k \in \overline{1, K}} q_k(i, j)}{\hat{A}(i, j)} - \lambda_{ij}^* + \mu^* = 0$$

Учитывая, что $A(i, j) > 0$, условие дополняющей нежёсткости можно переписать следующим образом:

$$\lambda_{ij}^* A(i, j) = 0 \Rightarrow \lambda_{ij}^* = 0, \ \forall i \in I, \ \forall j \in J$$

Положим $\mu^* = \sum_{k \in \overline{1, K}} \sum_{i, j} q_k(i, j) = K$, чтобы выполнялось условие $h(\hat{A}) = 0$, и получим следующую оценку на \hat{A} :

$$\hat{A}(i, j) = \frac{\sum_{k \in \overline{1, K}} q_k(i, j)}{K}$$

Данная оценка верна как для ЕМ, так и для МАР-ЕМ алгоритмов. В случае МАР-ЕМ оценка будет вырожденной в некоторых точках.

3.2.2 Вывод точечной оценки на параметр F

Для получения точечной оценки на параметр F рассмотрим следующую задачу:

$$\hat{F} = \underset{F \in \mathbb{R}^{h \times w}}{\operatorname{argmax}} \mathcal{L}_1(F) = \underset{F \in \mathbb{R}^{h \times w}}{\operatorname{argmax}} \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \cdot \log p(X_k | \bar{d}_k, F, B, s^2)$$

$$\hat{F} = \underset{F \in \mathbb{R}^{h \times w}}{\operatorname{argmin}} \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \sum_{\substack{m \in \overline{0, H-1} \\ n \in \overline{0, W-1}}} I[(m, n) \in \operatorname{faceArea}(\bar{d}_k)] \cdot q(\bar{d}_k) \cdot (X_k(m, n) - F(m - d_k^h, n - d_k^w))^2$$

Обозначим выражение под argmin через $\hat{f}(F)$ и найдём градиент $\nabla_F \hat{f}$.

$$\frac{\partial \hat{f}}{\partial F(i, j)} = \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \sum_{\substack{m \in \overline{0, H-1} \\ n \in \overline{0, W-1}}} 2 \cdot \mathcal{I}(m, n, i, j, \bar{d}_k) \cdot q(\bar{d}_k) \cdot (F(m - d_k^h, n - d_k^w) - X_k(m, n))$$

Здесь $\mathcal{I}(m, n, i, j, \bar{d}_k) = I[(m, n) \in \operatorname{faceArea}(\bar{d}_k)] \cdot I[m - d_k^h = i] \cdot I[n - d_k^w = j]$. Данный индикатор аналогичен следующим условиям: $m = i + d_k^h$, $n = j + d_k^w$, $0 \leq i \leq h - 1$, $0 \leq j \leq w - 1$. Ограничения на i и j выполняются в силу размерности F . Обозначим новый индикатор $\bar{\mathcal{I}}(i, j, \bar{d}_k) = I[i + d_k^h < H] \cdot I[j + d_k^w < W]$. Тогда, используя тот факт, что градиент в точке экстремума равен нулю, получим следующее выражение:

$$\begin{aligned} \frac{\partial \hat{f}(\hat{F})}{\partial F(i, j)} &= \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} 2 \cdot \bar{\mathcal{I}}(i, j, \bar{d}_k) \cdot q(\bar{d}_k) \cdot (\hat{F}(i, j) - X_k(i + d_k^h, j + d_k^w)) = 0 \\ \hat{F}(i, j) &= \frac{\sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \bar{\mathcal{I}}(i, j, \bar{d}_k) \cdot q(\bar{d}_k) \cdot X_k(i + d_k^h, j + d_k^w)}{\sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \bar{\mathcal{I}}(i, j, \bar{d}_k) \cdot q(\bar{d}_k)} \end{aligned}$$

Заметим, что ограничения, накладываемые индикатором $\bar{\mathcal{I}}$, выполняются в силу ограничений, которые определяются множеством \mathcal{D} и возможными значениями i , j . Таким образом, оценку на F можно переписать следующим образом:

$$\bar{F}(i, j) = \frac{1}{K} \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \cdot X_k(i + d_k^h, j + d_k^w)$$

В случае MAP-ЕМ алгоритма распределение вырождается и оценка на F принимает следующий вид:

$$\bar{F}(i, j) = \frac{1}{K} \sum_{k \in \overline{1, K}} X_k(i + \hat{d}_k^h, j + \hat{d}_k^w)$$

3.2.3 Вывод точечной оценки на параметр B

Для получения точечной оценки на параметр B рассмотрим следующую задачу:

$$\hat{B} = \underset{B \in \mathbb{R}^{H \times W}}{\operatorname{argmax}} \mathcal{L}_1(B) = \underset{B \in \mathbb{R}^{H \times W}}{\operatorname{argmax}} \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \cdot \log p(X_k | \bar{d}_k, F, B, s^2)$$

$$\hat{B} = \underset{B \in \mathbb{R}^{H \times W}}{\operatorname{argmin}} \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \sum_{\substack{m \in \overline{0, H-1} \\ n \in \overline{0, W-1}}} I[(m, n) \notin \operatorname{faceArea}(\bar{d}_k)] \cdot q(\bar{d}_k) \cdot (X_k(m, n) - B(m, n))^2$$

Обозначим выражение под argmin через $\tilde{f}(B)$ и найдём градиент $\nabla_B \tilde{f}$.

$$\frac{\partial \tilde{f}}{\partial B(i, j)} = \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \sum_{\substack{m \in \overline{0, H-1} \\ n \in \overline{0, W-1}}} 2 \cdot \tilde{\mathcal{I}}(m, n, i, j, \bar{d}_k) \cdot q(\bar{d}_k) \cdot (B(m, n) - X_k(m, n))$$

Здесь $\tilde{I}(m, n, i, j, \bar{d}_k) = I[(m, n) \notin \text{faceArea}(\bar{d}_k)] \cdot I[m = i] \cdot I[n = j]$. Произведение данных индикаторов аналогично следующему произведению: $I[(i, j) \notin \text{faceArea}(\bar{d}_k)] \cdot I[m = i] \cdot I[n = j]$. Учтём, что $m = i$ и $n = j$:

$$\frac{\partial \tilde{f}}{\partial B(i, j)} = \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} 2 \cdot I[(i, j) \notin \text{faceArea}(\bar{d}_k)] \cdot q(\bar{d}_k) \cdot (B(i, j) - X_k(i, j))$$

Тогда, используя тот факт, что градиент в точке экстремума равен нулю, получим следующее выражение:

$$\hat{B}(i, j) = \frac{\sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} I[(i, j) \notin \text{faceArea}(\bar{d}_k)] \cdot q(\bar{d}_k) \cdot X_k(i, j)}{\sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} I[(i, j) \notin \text{faceArea}(\bar{d}_k)] \cdot q(\bar{d}_k)}$$

Обозначим за $w_k(i, j) = \sum_{\bar{d}_k \in \mathcal{D}} I[(i, j) \notin \text{faceArea}(\bar{d}_k)] \cdot q(\bar{d}_k)$. Тогда оценка на параметр B принимает следующий вид:

$$\hat{B}(i, j) = \frac{\sum_{k \in \overline{1, K}} w_k(i, j) X_k(i, j)}{\sum_{k \in \overline{1, K}} w_k(i, j)}$$

В случае MAP-ЕМ алгоритма распределение вырождается, однако изменяется функциональный вид $w_k(i, j)$, а не B : $w_k(i, j) = I[(i, j) \notin \text{faceArea}(\bar{d}_k^*)]$, где \bar{d}_k^* - латентные переменные, на которых распределение вырождается.

3.2.4 Вывод точечной оценки на параметр s^2

Для получения точечной оценки на параметр B рассмотрим следующую задачу:

$$\hat{s} = \underset{s > 0}{\operatorname{argmax}} \mathcal{L}_1(s) = \underset{s > 0}{\operatorname{argmax}} \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \cdot \log p(X_k | \bar{d}_k, F, B, s^2)$$

$$\hat{s} = \underset{s > 0}{\operatorname{argmax}} \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \cdot \sum_{\substack{i \in \overline{0, H-1} \\ j \in \overline{0, W-1}}} I[(i, j) \in \text{faceArea}(\bar{d}_k)] \cdot h_1(s) + I[(i, j) \notin \text{faceArea}(\bar{d}_k)] \cdot h_2(s)$$

$$\text{Здесь } h_1(s) = -\log(s) - \frac{(X_k(i, j) - F(i - d_k^h, j - d_k^w))^2}{2s^2} \text{ и } h_2(s) = -\log(s) - \frac{(X_k(i, j) - B(i, j))^2}{2s^2}.$$

Обозначим всё, что под argmax , через $g(s)$. Тогда запишем $\frac{\partial g(s)}{\partial s}$:

$$\frac{\partial g(s)}{\partial s} = \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \cdot \sum_{\substack{i \in \overline{0, H-1} \\ j \in \overline{0, W-1}}} I[(i, j) \in \text{faceArea}(\bar{d}_k)] \cdot \frac{\partial h_1(s)}{\partial s} + I[(i, j) \notin \text{faceArea}(\bar{d}_k)] \cdot \frac{\partial h_2(s)}{\partial s}$$

$$\text{В свою очередь, } \frac{\partial h_1(s)}{\partial s} = -\frac{1}{s} + \frac{(X_k(i, j) - F(i - d_k^h, j - d_k^w))^2}{s^3} \text{ и } \frac{\partial h_2(s)}{\partial s} = -\frac{1}{s} + \frac{(X_k(i, j) - B(i, j))^2}{s^3}$$

Приравняем производную к нулю и найдём s^2 . Изначально домножим обе части равенства на s^3 :

$$\begin{aligned} HWK \cdot \hat{s}^2 &= \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \sum_{\substack{i \in \overline{0, H-1} \\ j \in \overline{0, W-1}}} f_1(i, j, \bar{d}_k) + f_2(i, j, \bar{d}_k) \\ \hat{s}^2 &= \frac{1}{HWK} \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} \sum_{\substack{i \in \overline{0, H-1} \\ j \in \overline{0, W-1}}} f_1(i, j, \bar{d}_k) + f_2(i, j, \bar{d}_k) \end{aligned}$$

Здесь функции f_1 и f_2 определяются следующим образом:

$$f_1(i, j, \bar{d}_k) = I[(i, j) \in \text{faceArea}(\bar{d}_k)] \cdot q(\bar{d}_k) \cdot (X_k(i, j) - F(i - d_k^h, j - d_k^w))^2$$

$$f_2(i, j, \bar{d}_k) = I[(i, j) \notin \text{faceArea}(\bar{d}_k)] \cdot q(\bar{d}_k) \cdot (X_k(i, j) - B(i, j))^2$$

В случае MAP-ЕМ алгоритма распределение вырождается, и оценка на s^2 имеет следующий вид:

$$\hat{s}^2 = \frac{1}{HWK} \sum_{k \in \overline{1, K}} \sum_{\substack{i \in \overline{0, H-1} \\ j \in \overline{0, W-1}}} \tilde{f}_1(i, j, \bar{d}_k^*) + \tilde{f}_2(i, j, \bar{d}_k^*)$$

$$\tilde{f}_1(i, j, \bar{d}_k^*) = I[(i, j) \in \text{faceArea}(\bar{d}_k^*)] \cdot (X_k(i, j) - F(i - d_k^{*h}, j - d_k^{*w}))^2$$

$$\tilde{f}_2(i, j, \bar{d}_k^*) = I[(i, j) \notin \text{faceArea}(\bar{d}_k^*)] \cdot (X_k(i, j) - B(i, j))^2$$

3.3 Задание 3. Вывод формулы для подсчёта нижней оценки на логарифм неполного правдоподобия $\mathcal{L}(q, \theta, A)$

В начале задания 2 была получена часть формулы для нижней оценки на логарифм неполного правдоподобия. Из неё осталось лишь вычесть энтропию распределения на латентные переменные:

$$\mathcal{L}(q, \theta, A) = \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \log p(X_k | \bar{d}_k, \bar{\theta}) + \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \log A(d_k^h, d_k^w) - \sum_{k \in \overline{1, K}} \sum_{\bar{d}_k \in \mathcal{D}} q(\bar{d}_k) \log q(\bar{d}_k)$$

4 Практическая часть

Данное задание предполагает наличие двух наборов данных - синтетически сгенерированный малого размера, а также предоставленный заранее с размером $H \times W \times K = 105 \times 245 \times 1000$

Требуется реализовать ЕМ и MAP-ЕМ алгоритмы и проверить их на исходных данных.

4.1 Эксперимент 1. Тестирование на синтетической выборке

Для данного эксперимента было выбрано небольшое изображение объекта размера 24×24 :

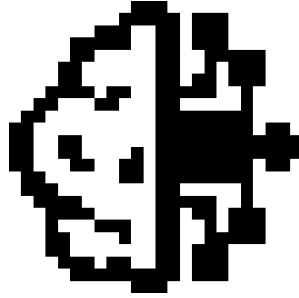


Рис. 1: Изображение объекта

В качестве изображения фона была создана чёрно-белая рябь размеров 40×50 следующего вида:

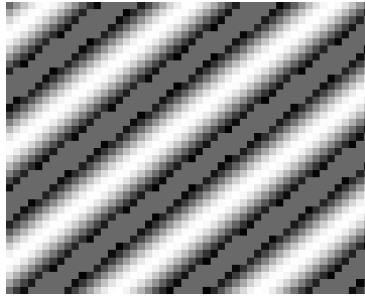


Рис. 2: Изображение фона

Изображение объекта было помещено в случайных позициях для генерации выборки. Пример изображений полученной выборки можно видеть на рисунке 3

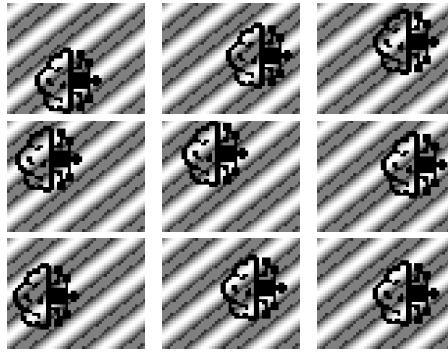


Рис. 3: Девять изображений из сгенерированных данных

К данным изображениям был добавлен шум из $\mathcal{N}(0, 0.1)$. Их можно увидеть на изображении 4:

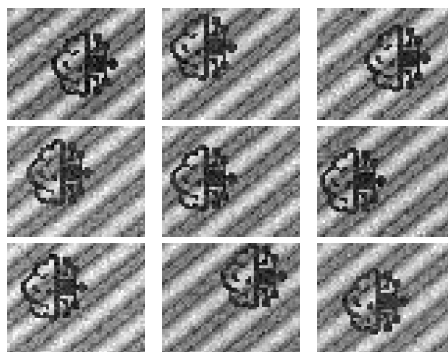
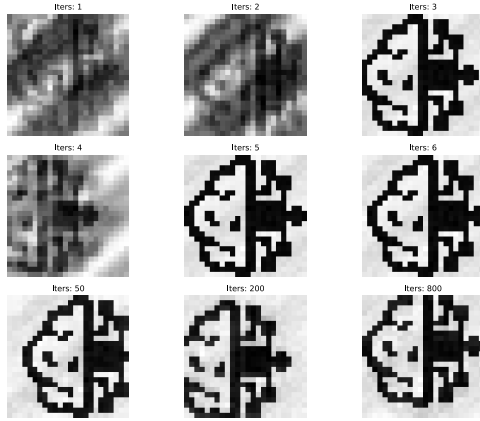
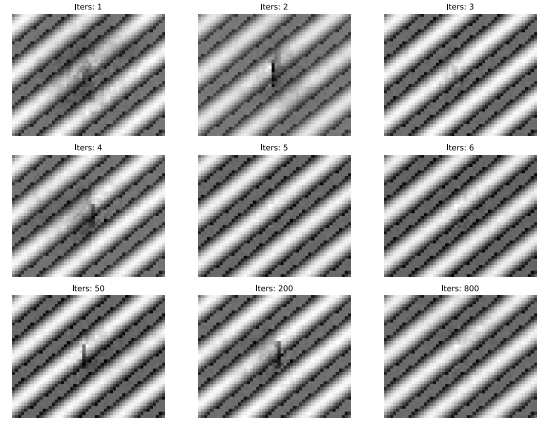


Рис. 4: Зашумлённые изображения

Начальное приближение для данного эксперимента было выбрано как равномерное на все параметры, кроме параметра A , для которого использовалось распределение Дирихле. Проводились также эксперименты с инициализацией через нормальное распределение, но результаты приведём для случая с равномерным распределением.



(a) Матрица F для разного максимального количества итераций



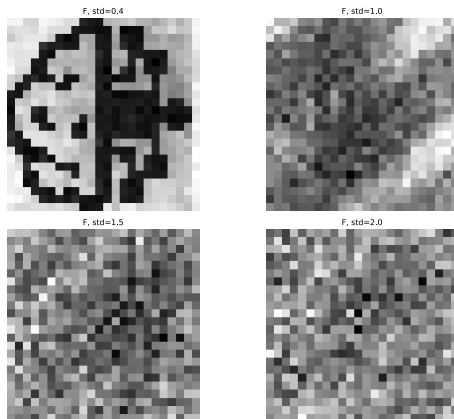
(b) Матрица F для разного максимального количества итераций

Видно, что ЕМ-алгоритм сходится достаточно быстро по количеству итераций, а также хорошо восстанавливает исходные изображения. Алгоритм запускался из разных приближений, было видно, что разные начальные приближения могут сильно влиять на результат алгоритма, что даёт основание использовать несколько запусков алгоритма для получения лучших результатов.

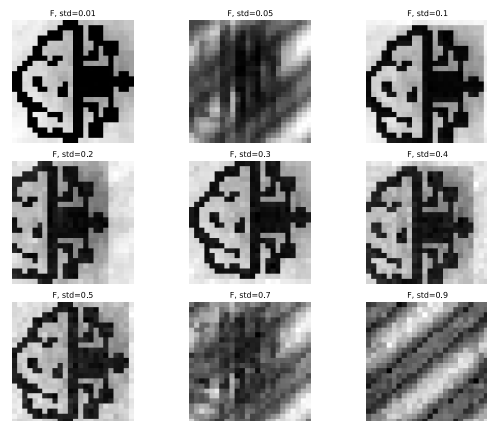
4.2 Эксперимент 2. Перебор различных значений отклонения и размера выборки

В качестве размеров выборки использовались следующие значения: 50, 100, 200, 300, 500, 1000, а для стандартного отклонения - 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.9, 1.0, 1.2, 1.5, 2.0. По результатам эксперимента было получено, что, начиная со стандартного отклонения, равного 1.0, результаты были уже плохо интерпретируемыми.

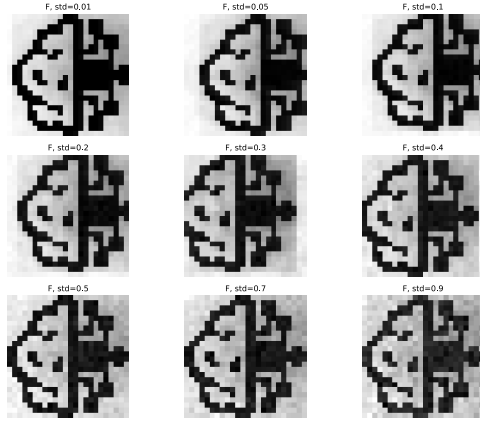
Также получено, что при большем объёме выборки результаты более стабильные, а при меньшем значении зашумления - менее зашумлённые. Также после применения алгоритма для разных зашумлений на выборке размера 300 видно, что значения $ELBO$ тем больше, чем меньше выборка была зашумлена.



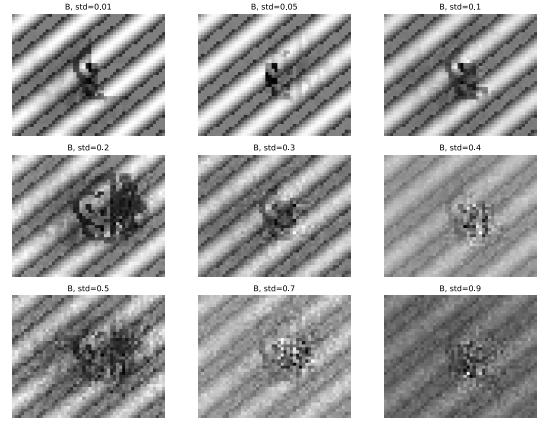
(a) Изображения объекта, пограничное значение зашумлённости



(b) Матрица F для выборки размера 300



(a) Матрица F для выборки размера 500



(b) Матрица B (фон) для выборки размера 50

Результаты для значений $ELBO$ при различной зашумлённости приведены на рисунке 8:

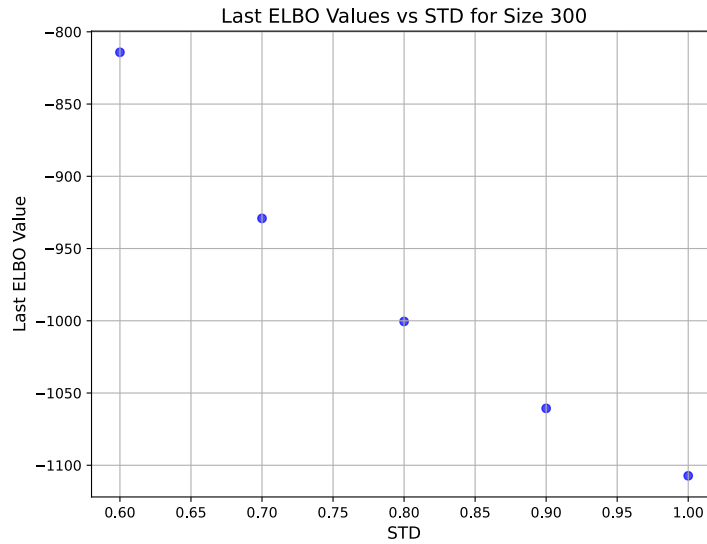
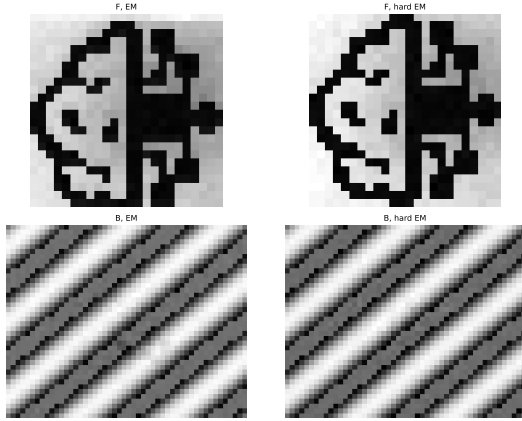


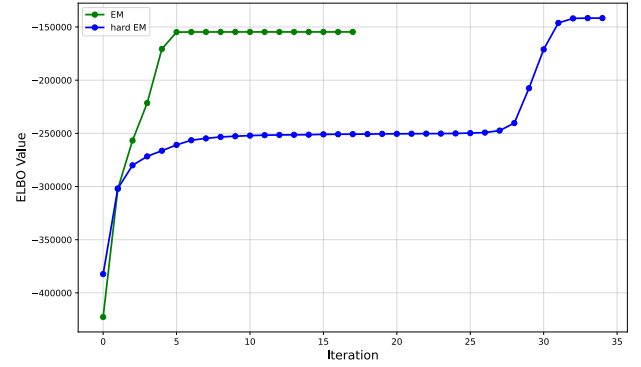
Рис. 8: Scatterplot значений $ELBO$ для выборки из 300 объектов

4.3 Эксперимент 3. Проверка EM и hard EM алгоритмов

Для сравнения алгоритмов EM и $hardEM$ была использована выборка из 500 с уровнем зашумления $s = 0.3$. Итоговые результаты отличаются не сильно, однако разница всё же есть, стандартный EM более стабильный. Приведём результаты на графиках.



(a) Матрицы F, B для выборки размера 500



(b) Значения ELBO в 3 эксперименте

4.4 Эксперимент 4. Применяем к реальным данным

Для расшумления предоставленной выборки данных были выбраны следующие размеры подвыборок: 50, 100, 300, 500. Запуск проводился на 30 итерациях. Результаты приведены на следующих графиках:

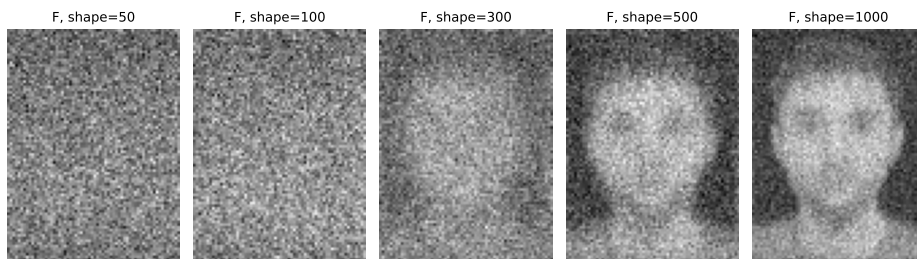


Рис. 10: Расшумлённые лица преступника для разного размера подвыборки

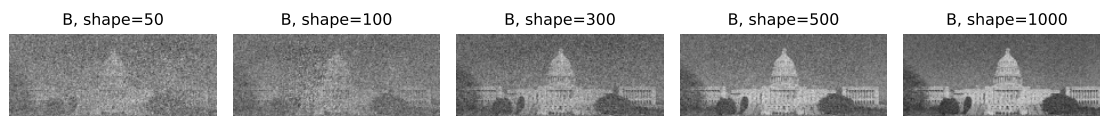


Рис. 11: Полученные изображения фона

Легко видеть, что хорошее расшумление получается при использовании всей выборки. Изображение фона - это изображение Капитолия в Вашингтоне. Изображение преступника - это, вероятно, Александр Оганов.

4.5 Возможные улучшения ЕМ-алгоритма

Приведём возможные улучшения ЕМ-алгоритма для данной задачи.

1. Улучшенная инициализация. Можно инициализировать параметры не из равномерного или нормального распределения, а брать среднее по выборке
2. Стохастический ЕМ. Можно делать шаги ЕМ алгоритма не на всей выборке, а на случайных батчах данных
3. Soft ЕМ. Вместо того чтобы брать максимум по распределению в $MAP - EM$, будем брать TOP-K значений и их оставлять, предварительно нормируя.