

Problem 1. Вывести формулы для подсчета следующих величин:

1. апостериорного распределения на координаты лица на изображениях $p(\mathbf{d}_k \mid \mathbf{X}_k, \boldsymbol{\theta}, \mathbf{A})$ на Е-шаге;
2. точечных оценок на параметры $\mathbf{A}, \boldsymbol{\theta} = \{\mathbf{F}, \mathbf{B}, s^2\}$ на М-шаге для ЕМ и МАР-ЕМ алгоритмов (учтите, что точечные оценки здесь нужно получать именно в таком порядке: $\mathbf{A}, \mathbf{F}, \mathbf{B}, s^2$);
3. нижней оценки на логарифм неполного правдоподобия $\mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A})$.

1.

$$p(\mathbf{d}_k \mid \mathbf{X}_k, \boldsymbol{\theta}, \mathbf{A}) = \frac{p(\mathbf{X}_k \mid \mathbf{d}_k, \boldsymbol{\theta}) \cdot p(\mathbf{d}_k \mid \mathbf{A})}{\sum_{\mathbf{d}_i} p(\mathbf{X}_k \mid \mathbf{d}_i, \boldsymbol{\theta}) \cdot p(\mathbf{d}_i \mid \mathbf{A})}$$

2. • \mathbf{A} :

$$\begin{aligned} \mathbb{E}_{q(d)} [\log p(\mathbf{X}, \mathbf{d} \mid \mathbf{A}, \boldsymbol{\theta})] &= \sum_{k=1}^K \mathbb{E}_{q(d)} [\log p(\mathbf{X}_k, \mathbf{d}_k \mid \mathbf{A}, \boldsymbol{\theta})] = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) [\log p(\mathbf{X}_k, (d_k^h, d_k^w) \mid \mathbf{A}, \boldsymbol{\theta})] = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) [\log p(\mathbf{X}_k \mid (d_k^h, d_k^w), \boldsymbol{\theta}) + \log p((d_k^h, d_k^w) \mid \mathbf{A})] = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) [\log p((d_k^h, d_k^w) \mid \mathbf{A})] + C = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) [\log \mathbf{A}_{d_k^h d_k^w}] \end{aligned}$$

Так как на \mathbf{A} есть условие, что оно должно суммироваться в 1, то запишем Лагранжан:

$$\begin{aligned} \frac{\partial}{\partial A_{ij}} \left(\sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \log \mathbf{A}_{d_k^h d_k^w} + \lambda \left(\sum_{ij} A_{ij} - 1 \right) \right) &= \\ = \sum_{k=1}^K q((i, j)) \frac{\partial}{\partial A_{ij}} \log \mathbf{A}_{ij} + \lambda \mathbf{A}_{ij} &= \sum_{k=1}^K q((i, j)) \frac{1}{\mathbf{A}_{ij}} + \lambda \mathbf{A}_{ij} = 0 \\ \Rightarrow \mathbf{A}_{ij} &= \frac{\sum_k q((i, j))}{\sum_{k, d_k^h, d_k^w} q((d_k^h, d_k^w))} = \left\{ \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) = 1 \right\} = \frac{\sum_k q((i, j))}{K} \end{aligned}$$

В случае hard EM

$$\mathbf{A}_{ij} = \frac{\sum_k [(i, j) = MAP(d_k)]}{K}$$

• \mathbf{F} :

$$\begin{aligned} \mathbb{E}_{q(d)} [\log p(\mathbf{X}, \mathbf{d} \mid \mathbf{A}, \boldsymbol{\theta})] &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) [\log p(\mathbf{X}_k \mid (d_k^h, d_k^w), \boldsymbol{\theta})] + C = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[\sum_{(ij) \in \text{face}} \log \mathcal{N}(\mathbf{X}_k^{ij} \mid \mathbf{F}_{i-d_k^h, j-d_k^w}, s^2) \right] + C \end{aligned}$$

Теперь берём производную:

$$\begin{aligned} \frac{\partial}{\partial F_{ij}} \left(\sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[\sum_{(ij) \in \text{face}} \log \mathcal{N}(\mathbf{X}_k^{ij} \mid \mathbf{F}_{i-d_k^h, j-d_k^w}, s^2) \right] \right) &= \\ = \left(\sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[\frac{\partial}{\partial F_{ij}} (\mathbf{X}_k^{i+d_k^h, j+d_k^w} - \mathbf{F}_{i, j})^2 \right] \right) &= 0 \end{aligned}$$

$$\Rightarrow \mathbf{F}_{ij} = \frac{\sum_{k, d_k^h, d_k^w} q((d_k^h, d_k^w)) \mathbf{X}_k^{i+d_k^h, j+d_k^w}}{K}$$

В случае hard EM:

$$\mathbf{F}_{ij} = \frac{\sum_k [(i, j) = MAP(d_k)] \mathbf{X}_k^{i+d_k^h, j+d_k^w}}{K}$$

• B :

$$\begin{aligned} \mathbb{E}_{q(d)} [\log p(\mathbf{X}, \mathbf{d} \mid \mathbf{A}, \boldsymbol{\theta})] &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) [\log p(\mathbf{X}_k \mid (d_k^h, d_k^w), \boldsymbol{\theta})] + C = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[\sum_{(ij) \notin \text{face}} \log \mathcal{N}(\mathbf{X}_k^{ij} \mid \mathbf{B}_{ij}, s^2) \right] + C = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[\sum_{(ij) \notin \text{face}} -(\mathbf{X}_k^{ij} - \mathbf{B}_{ij})^2 \right] + C = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[\sum_{ij} -(\mathbf{X}_k^{ij} - \mathbf{B}_{ij})^2 + \sum_{(ij) \in \text{face}} (\mathbf{X}_k^{ij} - \mathbf{B}_{ij})^2 \right] + C = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \sum_{ij} (\mathbf{X}_k^{ij} - \mathbf{B}_{ij})^2 + \sum_{ij} (\mathbf{X}_k^{ij} - \mathbf{B}_{ij})^2 \sum_{d_k^h=i-h, d_k^w=j-w}^{ij} q((d_k^h, d_k^w)) + C = \\ &= \sum_{k=1}^K \sum_{ij} (\mathbf{X}_k^{ij} - \mathbf{B}_{ij})^2 \left[- \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) + \sum_{d_k^h=i-h, d_k^w=j-w}^{ij} q((d_k^h, d_k^w)) \right] + C = \\ &= \sum_{k=1}^K \sum_{ij} (\mathbf{X}_k^{ij} - \mathbf{B}_{ij})^2 \left[\sum_{d_k^h=i-h, d_k^w=j-w}^{ij} q((d_k^h, d_k^w)) - 1 \right] + C \end{aligned}$$

Теперь берём производную:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{B}_{ij}} \left(\sum_{k=1}^K \sum_{ij} (\mathbf{X}_k^{ij} - \mathbf{B}_{ij})^2 \left[\sum_{d_k^h=i-h, d_k^w=j-w}^{ij} q((d_k^h, d_k^w)) - 1 \right] \right) &= \\ &= 2 \sum_{k=1}^K (\mathbf{X}_k^{ij} - \mathbf{B}_{ij}) \left[\sum_{d_k^h=i-h, d_k^w=j-w}^{ij} q((d_k^h, d_k^w)) - 1 \right] = 0 \\ \Rightarrow \mathbf{B}_{ij} &= \frac{\sum_k \mathbf{X}_k^{ij} \left[\sum_{d_k^h=i-h, d_k^w=j-w}^{ij} q((d_k^h, d_k^w)) - 1 \right]}{\sum_k \left[\sum_{d_k^h=i-h, d_k^w=j-w}^{ij} q((d_k^h, d_k^w)) - 1 \right]} \end{aligned}$$

В случае hard EM:

$$\mathbf{B}_{ij} = \frac{\sum_k \mathbf{X}_k^{ij} \left[\sum_{d_k^h=i-h, d_k^w=j-w}^{ij} [(d_k, d_k) = MAP(d_k)] - 1 \right]}{\sum_k \left[\sum_{d_k^h=i-h, d_k^w=j-w}^{ij} [(d_k, d_k) = MAP(d_k)] - 1 \right]}$$

• s^2 :

$$\begin{aligned} \mathbb{E}_{q(d)} [\log p(\mathbf{X}, \mathbf{d} \mid \mathbf{A}, \boldsymbol{\theta})] &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) [\log p(\mathbf{X}_k \mid (d_k^h, d_k^w), \boldsymbol{\theta})] + C = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[\sum_{(ij) \in \text{face}} \log \mathcal{N}(\mathbf{X}_k \mid \mathbf{F}_{i-d_k^h, j-d_k^w}, s^2) + \sum_{(ij) \notin \text{face}} \log \mathcal{N}(\mathbf{X}_k \mid \mathbf{B}_{ij}, s^2) \right] + C = \end{aligned}$$

$$= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[-\frac{HW}{2} \log s^2 - \frac{1}{2s^2} \sum_{(ij) \in \text{face}} + \sum_{(ij) \notin \text{face}} \right] + C$$

Теперь берём производную:

$$\begin{aligned} \frac{\partial}{\partial s^2} \left(\sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[-\frac{HW}{2} \log s^2 - \frac{1}{2s^2} \sum_{(ij) \in \text{face}} + \sum_{(ij) \notin \text{face}} \right] \right) &= \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[-\frac{HW}{2s^2} + \frac{1}{2s^4} \sum_{(ij) \in \text{face}} + \sum_{(ij) \notin \text{face}} \right] = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[-HWs^2 + \sum_{(ij) \in \text{face}} + \sum_{(ij) \notin \text{face}} \right] = 0 \\ \Rightarrow s^2 &= \frac{\sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[\sum_{(ij) \in \text{face}} + \sum_{(ij) \notin \text{face}} \right]}{HWK}, \end{aligned}$$

где

$$\begin{aligned} \sum_{(ij) \in \text{face}} &= \sum_{(ij) \in \text{face}} \left(X_k^{ij} - F^{i-d_k^h, j-d_k^w} \right)^2 \\ \sum_{(ij) \notin \text{face}} &= \sum_{(ij) \notin \text{face}} \left(X_k^{ij} - B^{ij} \right)^2 \end{aligned}$$

В случае hard EM:

$$s^2 = \frac{\sum_{k=1}^K \sum_{d_k^h, d_k^w} [(d_k^h, d_k^w) = \text{MAP}(d_k)] \left[\sum_{(ij) \in \text{face}} + \sum_{(ij) \notin \text{face}} \right]}{HWK}$$

3.

$$\begin{aligned} \mathcal{L}(q, \theta, A) &= \mathbb{E}_{q(d)} [\log p(X, d | A, \theta) - \log q(d)] = \\ &= \sum_{k=1}^K \sum_{d_k^h, d_k^w} q((d_k^h, d_k^w)) \left[-\frac{HW}{2} \log(s^2 \cdot 2 \cdot \pi) - \frac{1}{2s^2} \sum_{(ij) \in \text{face}} + \sum_{(ij) \notin \text{face}} - \log q((d_k^h, d_k^w)) \right] \end{aligned}$$

Problem 2. Приведите примеры исходных данных.

Перед началом работы я нормирую изображения, чтобы они лежали в $[0, 1]$.

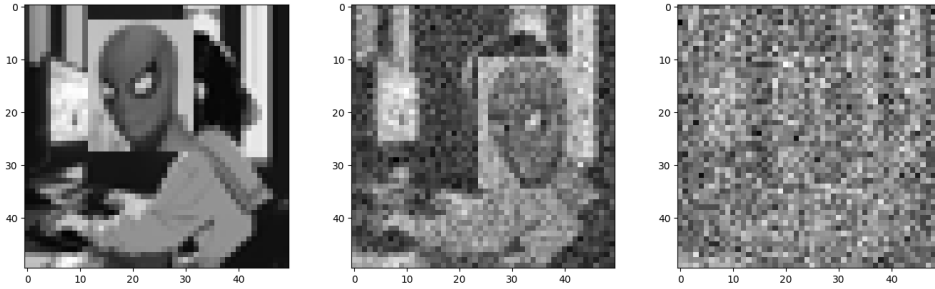


Рис. 1: Примеры данных при разных s : 0, 0.1, 0.5.

Problem 3. Протестируйте полученный EM алгоритм на сгенерированных данных. Сильно ли влияет начальное приближение на параметры на результаты работы? Стоит ли для данной задачи запускать EM алгоритм из разных начальных приближений?

Видно, что изображения немного, но отличаются; третье изображение выбивается из остальных (чуть хуже получилось). Если смотреть на значения L, то там тоже примерно всё одинаково, кроме третьего изображения (на ней чуть меньше). Поэтому можно сделать вывод, что начальное приближение влияет, но не сильно. Использовать EM с restart конечно стоит, не зря же мы его писали.

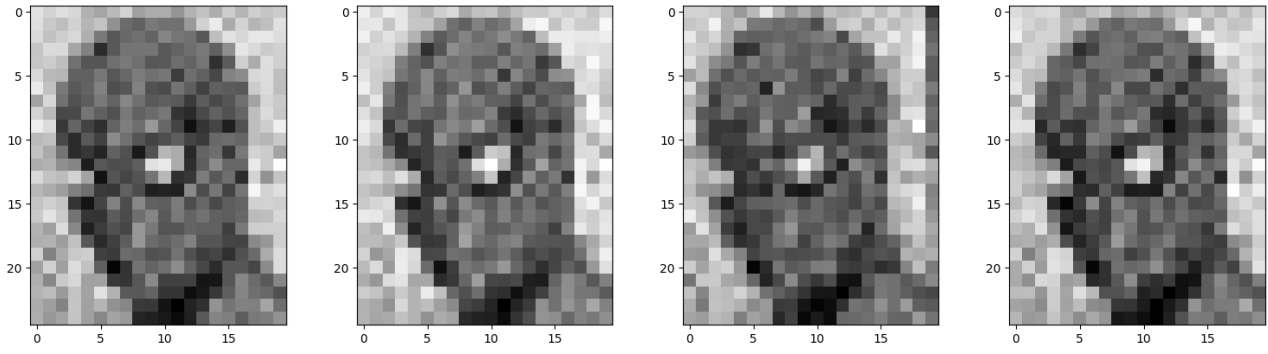


Рис. 2: Результаты запуска ЕМ с разных начальных приближений. При зашумлении $s=0.5$.

Problem 4. Запустите ЕМ алгоритм на сгенерированных выборках разных размеров и с разным уровнем зашумления. Как изменения в обучающей выборке влияют на результаты работы (получаемые F, B и $\mathcal{L}(q, \theta, \mathbf{A})$)? При каком уровне шума ЕМ-алгоритм перестает выдавать вменяемые результаты? В данном пункте учтите, что для сравнения значения $\mathcal{L}(q, \theta, \mathbf{A})$ для выборок разного размера стоит нормировать его на объем выборки.

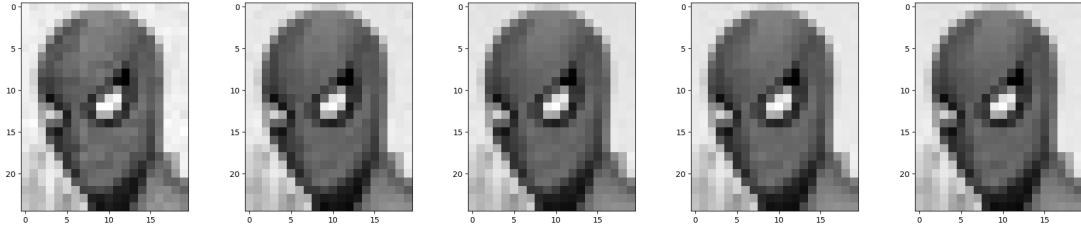


Рис. 3: $s=0.1$, $K=50$, $L=2235$ Рис. 4: $s=0.1$, $K=250$, $L=2210$ Рис. 5: $s=0.1$, $K=500$, $L=2205$ Рис. 6: $s=0.1$, $K=750$, $L=2204$ Рис. 7: $s=0.1$, $K=1k$, $L=2203$

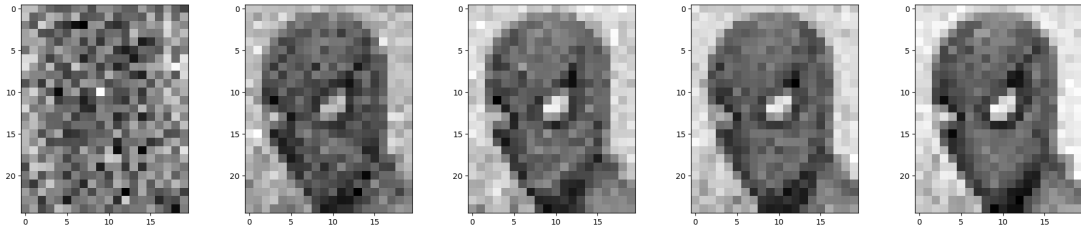


Рис. 8: $s=1.0$, $K=50$, $L=3531$ Рис. 9: $s=1.0$, $K=250$, $L=3545$ Рис. 10: $s=1.0$, $K=500$, $L=3549$ Рис. 11: $s=1.0$, $K=750$, $L=3550$ Рис. 12: $s=1.0$, $K=1k$, $L=3551$

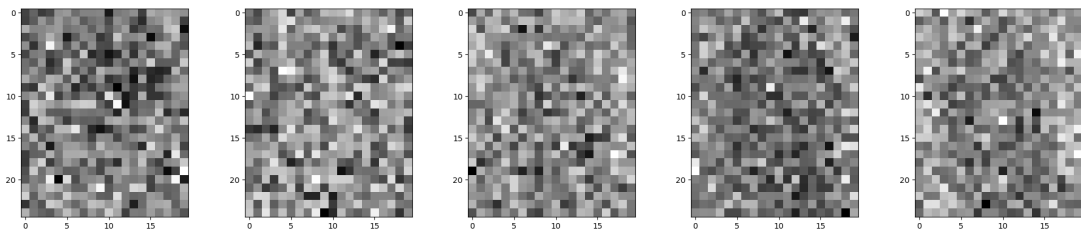
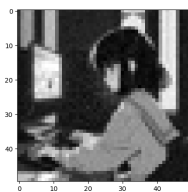
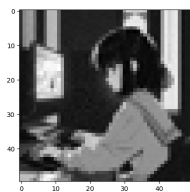
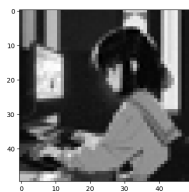
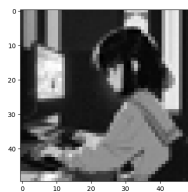
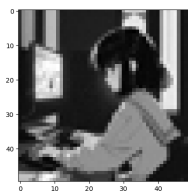
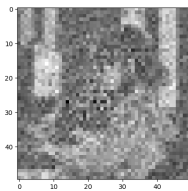
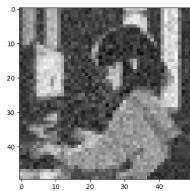
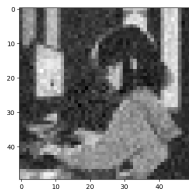
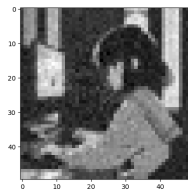
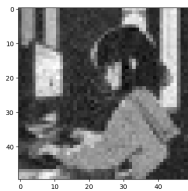
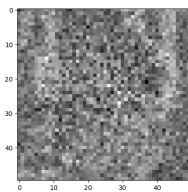
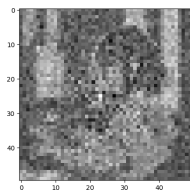
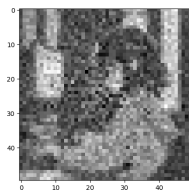
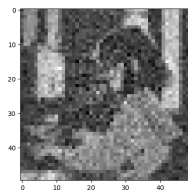
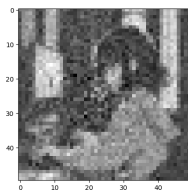


Рис. 13: $s=2.5$, $K=50$, $L=5817$ Рис. 14: $s=2.5$, $K=250$, $L=5838$ Рис. 15: $s=2.5$, $K=500$, $L=5836$ Рис. 16: $s=2.5$, $K=750$, $L=5838$ Рис. 17: $s=2.5$, $K=1k$, $L=5837$

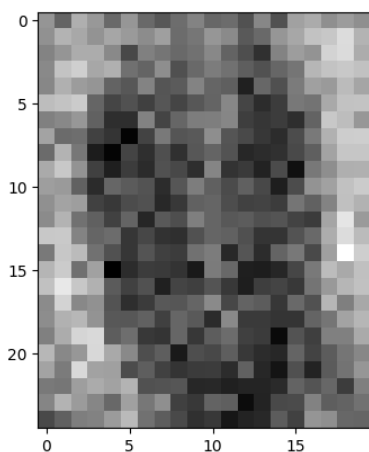
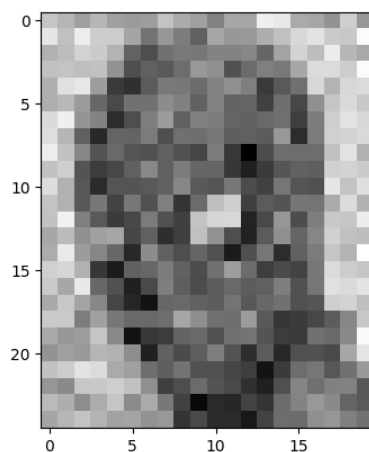
Как видно, при увеличении размера выборки, изображения расшумляются лучшим образом. При этом значение \mathcal{L} не всегда отражает действительность: в первой строчке оно падает с увеличением размера выборки, во второй последние изображения имеют меньшее значение, хотя очевидно, что она выглядит

Рис. 18: $s=0.1$, $K=50$ Рис. 19: $s=0.1$, $K=250$ Рис. 20: $s=0.1$, $K=500$ Рис. 21: $s=0.1$, $K=750$ Рис. 22: $s=0.1$, $K=1000$ Рис. 23: $s=1.0$, $K=50$ Рис. 24: $s=1.0$, $K=250$ Рис. 25: $s=1.0$, $K=500$ Рис. 26: $s=1.0$, $K=750$ Рис. 27: $s=1.0$, $K=1000$ Рис. 28: $s=2.5$, $K=50$ Рис. 29: $s=2.5$, $K=250$ Рис. 30: $s=2.5$, $K=500$ Рис. 31: $s=2.5$, $K=750$ Рис. 32: $s=2.5$, $K=1000$

лучше. Последняя строка вообще не расшумилась, хотя в последнем изображении виден тёмный овал (силуэт лица), то есть делаем вывод, что s надо брать меньше 2.5.

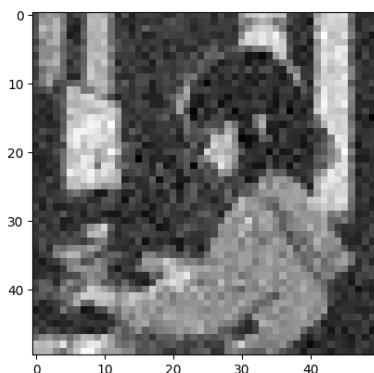
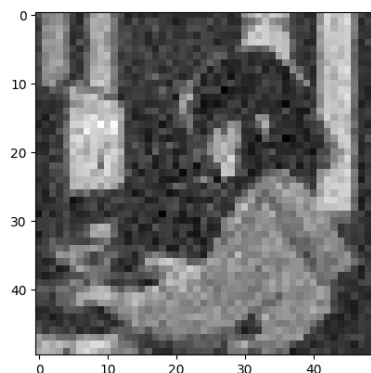
С задним фоном дела обстоят намного лучше, но это и неудивительно, ведь он статичный.

Problem 5. Сравните качество и время работы EM и hard EM на сгенерированных данных. Как Вы думаете, почему разница в результатах работы так заметна?

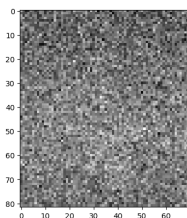
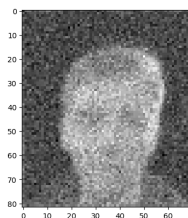
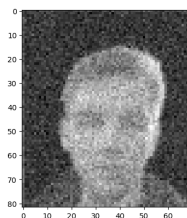
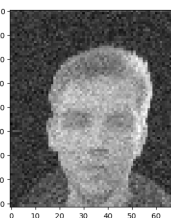
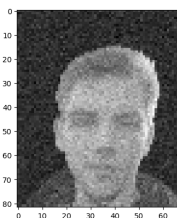
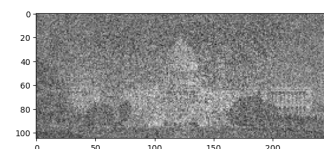
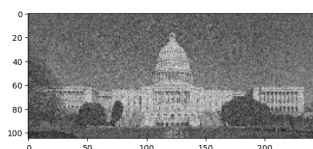
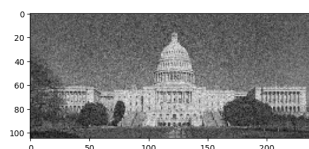
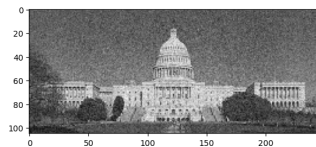
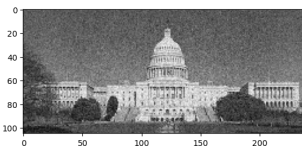
Рис. 33: $s=1.0$, $K=250$, $L=-888k$,
hard EMРис. 34: $s=1.0$, $K=250$, $L=-885k$,
EM

Если смотреть на качество, то изображение лица у hard EM получается довольно плохо (виден лишь силуэт), тогда как EM неплохо расшумляет лицо; бэкграунд у методов почти одинаковый. Если смотреть на время работы, то hard EM в 6 раз быстрее обычный EM.

На мой такое происходит, потому что мода распределения довольно часто нерепрезентативна. Поэтому тут встанет trade-off: либо качество, либо скорость.

Рис. 35: $s=1.0$, $K=250$, hard EMРис. 36: $s=1.0$, $K=250$, EM

Problem 6. Примените EM алгоритм к данным с зашумленными снимками преступника. Приведите результаты работы алгоритма на выборках разного размера.

Рис. 37: $K=50$ Рис. 38: $K=250$ Рис. 39: $K=500$ Рис. 40: $K=750$ Рис. 41: $K=1000$ Рис. 42: $K=50$ Рис. 43: $K=250$ Рис. 44: $K=500$ Рис. 45: $K=750$ Рис. 46: $K=1000$

Problem 7. Предложите какую-нибудь модификацию полученного EM алгоритма, которая бы работала на данной задаче качественнее и/или быстрее.

На мой взгляд тут сильно может помочь [Винеровская фильтрация](#), которую используют для расшумления изображений. Комбинируя EM и фильтрацию, можно с одной стороны расшумлять довольно шумные изображения, а с другой стороны оценивать координаты лица на изображениях.