

Задание 2. EM алгоритм для детектива

Курс: Байесовские методы в машинном обучении, осень 2015

Начало выполнения задания: 24 октября.

Срок сдачи: **7 ноября, 23:59.**

Среда для выполнения задания: Python 3.

1 Введение

Леди Маргалотта обратилась к известному детективу [Никку Картеру](#) с просьбой расследовать похищение ее любимого песика. Даже огромная страховка не покроет горя хозяйки... Во время ее поездки в Лондон незнакомец выкрал любимца леди Маргалотты. Камеры видеонаблюдения, зафиксировавшие вора оказались плохими помощниками, т.к. подверглись воздействию направленного электро-магнитного постановщика помех. Тем не менее, это единственная зацепка. Под подозрением [группа Байесовских методов](#) из России.

Никку Картеру необходимо обработать снимки и восстановить лицо вора. Известно, что на разных фотографиях оно расположено в случайных координатах на неподвижном фоне. К счастью, видеолекции по курсу БММО уже выложены в сеть и пройдя его Ник Картер смог решить задачу и, с помощью EM-алгоритма, установить личность вора. Его ждал сюрприз...

Помогите детективу изобличить преступника. Зашумленные фотографии поступают из лаборатории порциями, которые будут выкладываться каждые 3 дня. Пример зашумленной фотографии показан на рисунке 1. Помните, что чем быстрее обнаружен преступник, тем проще его поймать! Определившему просьба оперативно сообщить об этом Никку (о том, как можно с ним связаться, читайте в разделе «Оформление задания»). Первые трое сообщивших правильный ответ получают благодарность от Ника и небольшие призы.

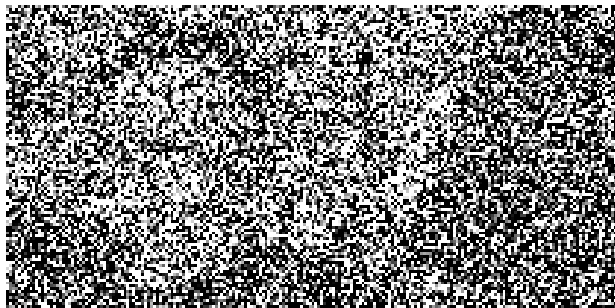


Рис. 1: Пример зашумленного изображения с камер видеонаблюдения.

2 Описание модели

Дана выборка $\mathbf{X} = \{\mathbf{X}_k\}_{k=1}^K$ 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 маска с лицом на изображении имеет прямоугольную форму размера $h \times w$ пикселей. В выданных данных $h = 100, w = 70$. Макет изображения показан на рисунке 2.

Введем следующие обозначения:

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

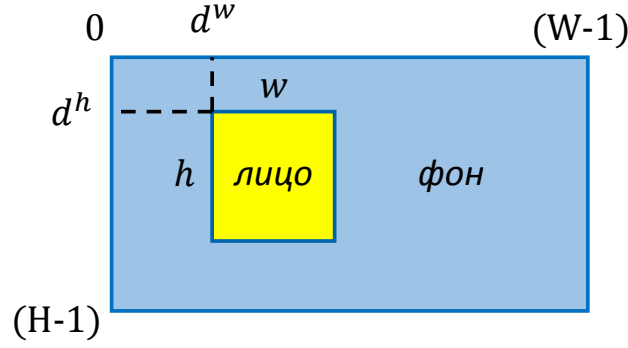


Рис. 2: Макет изображения с камеры наблюдения без шума.

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:

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

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

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

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

где $A(i, j)$ — элемент матрицы \mathbf{A} .

В итоге имеем следующую совместную вероятностную модель:

$$p(\mathbf{X}, \mathbf{d} | \boldsymbol{\theta}, \mathbf{A}) = \prod_k p(\mathbf{X}_k | \mathbf{d}_k, \boldsymbol{\theta}) p(\mathbf{d}_k | \mathbf{A}).$$

3 Формулировка задания

Требуется решить задачу

$$p(\mathbf{X} | \boldsymbol{\theta}, \mathbf{A}) \rightarrow \max_{\boldsymbol{\theta}, \mathbf{A}}.$$

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

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

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

$$q(\mathbf{d}) = p(\mathbf{d} | \mathbf{X}, \boldsymbol{\theta}, \mathbf{A}) = \prod_k p(\mathbf{d}_k | \mathbf{X}_k, \boldsymbol{\theta}, \mathbf{A}),$$

а на M-шаге вычисляется точечная оценка на параметры $\boldsymbol{\theta}, \mathbf{A}$:

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

Также далее будет рассматриваться упрощенный вариант EM-алгоритма, который называется hard EM. В нем после E шага берется не все апостериорное распределение на координаты лица на изображениях, а только MAP оценка на эти координаты (то есть после E шага $q(\mathbf{d})$ преобразовывают так, что для каждого изображения \mathbf{X}_k оценка $q(\mathbf{d}_k)$ принимает значение 1 только в одной точке — точке аргмаксимума апостериорного распределения $p(\mathbf{d}_k | \mathbf{X}_k, \boldsymbol{\theta}, \mathbf{A})$).

При выполнении итераций EM алгоритма нужно следить за значением оптимизируемого функционала $\mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A})$.

Теория.

Вывести формулы для подсчета следующих величин:

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

Программирование.

Прототипы всех основных функций выданы вместе с заданием. При оценке выполнения задания будет учитываться эффективность программного кода. Необходимо реализовать:

1. EM-алгоритм со вспомогательными функциями. В качестве критерия останова использовать следующее условие на нижнюю оценку на логарифм неполного правдоподобия:

$$\mathcal{L}(q, \boldsymbol{\theta}^{(t+1)}, \mathbf{A}^{(t+1)}) - \mathcal{L}(q, \boldsymbol{\theta}^{(t)}, \mathbf{A}^{(t)}) < tol,$$

где (t) и $(t + 1)$ означают номера итераций, а tol — небольшая константа.

2. Дополнить функции для выполнения M шага и EM алгоритма на случай hard EM алгоритма.
3. Функцию, запускающую EM-алгоритм несколько раз из разных начальных приближений.
4. Для проверки работы алгоритма сгенерировать выборку из небольших зашумленных нормальным шумом черно-белых изображений с одинаковым фоном и каким-то объектом в случайной позиции. Пример таких изображений до и после зашумления показан на рисунке 3. Эти изображения должны быть достаточно маленькими для быстрой отладки кода.

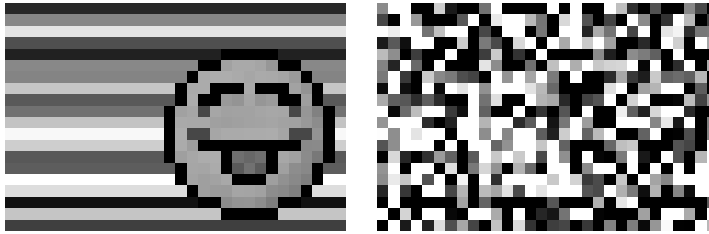


Рис. 3: Пример изображений для тестирования работы алгоритма до и после зашумления. Здесь $h = w = 15, H = 20, W = 30$.

Рекомендации по реализации:

- Значение $\mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A})$ должно возрастать с течением итераций. Если это не так, в реализации или в выводе формул ошибка. Это хороший способ отладки вашей программы.
- Для ускорения можно вычислять $\mathcal{L}(q, \boldsymbol{\theta}, \mathbf{A})$ не на каждой итерации алгоритма. Используйте этот совет только в своих экспериментах, а на проверку код отправляйте с подсчетом этого функционала на каждой итерации, как и написано в прототипе.
- Для того, чтобы избежать проблем с точностью вычислений, следует везде, где это возможно, переходить от произведений к суммированию логарифмов. Этот момент частично учтен в прототипах функций: функция для подсчета $p(\mathbf{X}_k | \mathbf{d}_k, \boldsymbol{\theta})$ должна выдавать прологарифмированные значения вероятностей.
- При подсчете $q(\mathbf{d})$ для более устойчивой нормировки также стоит перейти к логарифмам вероятностей и воспользоваться следующим трюком:

$$\alpha_i = \log p_i(\dots) \quad \rightarrow \quad \frac{e^{\alpha_i}}{\sum_k e^{\alpha_k}} = \frac{e^{(\alpha_i - \max_j \alpha_j)}}{\sum_k e^{(\alpha_k - \max_j \alpha_j)}}.$$

Анализ.

Обязательно приведите в отчете примеры исходных данных и полученных результатов (F, B).

1. 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?
2. Запустите EM 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 EM-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.
3. Сравните качество и время работы EM и hard EM на сгенерированных данных. Как Вы думаете, почему разница в результатах работы так заметна?
4. Примените EM алгоритм к данным с зашумленными снимками преступника. Приведите результаты работы алгоритма на выборках разного размера.
5. Предложите какую-нибудь модификацию полученного EM алгоритма, которая бы работала на данной задаче качественнее и/или быстрее.
6. * Реализуйте предложенную модификацию EM алгоритма и сравните ее с исходной по качеству результатов и времени работы.

Пункт со * является необязательным. За его выполнение можно получить дополнительные баллы. Также дополнительные баллы будут выставлены за наиболее интересные ответы на пункт 5.

Формат данных

Данные с зашумленными фотографиями преступника будут выкладываться каждые 3 дня начиная с дня выдачи задания (всего 5 раз). Каждый раз будет выкладываться файл, который будет содержать все ранее выданные данные и новую порцию данных. Файл будет иметь формат .pnu, то есть это будет сохраненный numpy array размера $H \times W \times N$, где $H \times W$ — размер каждого изображения, а N — их число. В выданных данных $h = 100, w = 70$.

4 Оформление задания

Срочное сообщение Нику. Сообщение Нику является необязательным и не влияет на оценку за задание. Сообщение следует отправить письмом по адресу bayesml@gmail.com с заголовком письма

«[БММО15] Практика 2, Фамилия Имя, Имя преступника».

В письме нужно указать кого Вы подозреваете в преступлении. Также свои подозрения нужно подтвердить, например, фотографией преступника, которую Вам удалось получить. Присылать сообщение Нику можно **только один раз**. Сдача полного задания не считается сообщением Нику. Первые трое приславших сообщение с правильным ответом получают благодарность от Ника и небольшие призы.

Полная сдача задания. Выполненное задание следует отправить письмом по адресу bayesml@gmail.com с заголовком письма

«[БММО15] Практика 2, Фамилия Имя».

Убедительная просьба присылать выполненное задание только один раз с окончательным вариантом. Также убедительная просьба строго придерживаться заданных прототипов реализуемых функций (для проверки задания используются, в том числе, автоматические процедуры, которые являются чувствительными к неверным прототипам).

Присланный вариант задания должен содержать в себе:

- Текстовый файл в формате PDF с указанием ФИО, содержащий описание всех проведенных исследований (вывод всех формул, описание экспериментов, анализ и выводы).
- Файл .py с реализацией всех обязательных функций. К обязательным функциям относятся все функции, для которых были выданы прототипы.
- Оба файла должны быть названы фамилией студента на английском языке с заглавной буквы. Например, Ivanov.pdf и Ivanov.py. Если Вы присылаете какие-либо дополнительные файлы, то их название должно также начинаться с фамилии студента на английском языке с заглавной буквы.