ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

ЕМ-алгоритм. Обучение скрытых марковских моделей без учителя.

Д. П. Ветров[1]

[1]МГУ, ВМиК, каф. ММП

Курс «Графические модели»

План лекции

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

1. ЕМ-алгоритм в общем виде
2. Скрытая марковская модель
3. Метод максимального правдоподобия для СММ
4. Алгоритм «вперед-назад»
5. Устойчивые формулы для алгоритма «вперед-назад»
6. Модельный пример

EM-алгоритм. Разложение логарифма правдоподобия

Требуется найти максимум правдоподобия в вероятностной модели со скрытыми переменными:

$$p(X|\Theta) = \int p(X, T|\Theta)dT \to \max_{\Theta} \Leftrightarrow \log p(X|\Theta) \to \max_{\Theta}$$

$$p(X, T|\Theta) = p(T|X, \Theta)p(X|\Theta) \Rightarrow$$
$$\log p(X, T|\Theta) = \log p(T|X, \Theta) + \log p(X|\Theta)$$

$q(T)$ — произвольное распределение.

$$\log p(X|\Theta) = \int \log p(X|\Theta)q(T)dT =$$
$$\int [\log p(X, T|\Theta) - \log p(T|X, \Theta)] q(T)dT =$$
$$\int \log p(X, T|\Theta)q(T)dT - \int \log p(T|X, \Theta)q(T)dT =$$
$$\int \log \frac{p(X, T|\Theta)}{q(T)}q(T)dT - \int \log \frac{p(T|X, \Theta)}{q(T)}q(T)dT$$

Нижняя оценка для логарифма правдоподобия

$$\log p(X|\Theta) = \underbrace{\int \log \frac{p(X,T|\Theta)}{q(T)} q(T)dT}_{l(q,\Theta)} - \underbrace{\int \log \frac{p(T|X,\Theta)}{q(T)} q(T)dT}_{KL(q||p) \geq 0}$$

Дивергенция Кульбака-Лейблера $KL(q||p)$ определяет расстояние между вероятностными распределениями

- $KL(q||p) = -\int q(x) \log(p(x)/q(x))dx$
- $KL(q||p) \geq 0$ и $KL(q||p) = 0 \Leftrightarrow q \equiv p$.
- $KL(q||p) \neq KL(p||q)$

Тогда $l(q,\Theta)$ является нижней оценкой правдоподобия $\log p(X|\Theta)$:

$$\log p(X|\Theta) \geq l(q,\Theta) \text{ и равенство } \Leftrightarrow q(T) = p(T|X,\Theta)$$

Идея ЕМ-алгоритма

$$\log p(X|\Theta) = l(q,\Theta) + KL(q||p)$$

Итерационная схема. Фиксируем некоторое значение Θ_{old}. Приблизим в точке Θ_{old} правдоподобие с помощью его нижней оценки:

$$q(T) = p(T|X,\Theta_{old})$$

$$\log p(X|\Theta) \geq l(q,\Theta) = \int \log p(X,T|\Theta) p(T|X,\Theta_{old}) dT - \int \log p(T|X,\Theta_{old}) p(T|X,\Theta_{old}) dT$$

Найдем новое значение Θ с помощью максимизации нижней оценки:

$$l(q,\Theta) \to \max_{\Theta} \Leftrightarrow \mathbb{E}_{T|X,\Theta_{old}} \log p(X,T|\Theta) \to \max_{\Theta}$$

EM-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

Иллюстрация EM-алгоритма

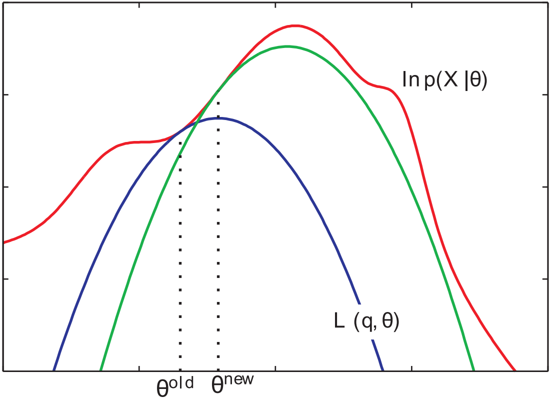


Схема EM-алгоритма

- Е-шаг. Фиксируется значение параметров Θ_{old}. Оценивается апостериорное распределение на скрытые переменные $p(T|X,\Theta_{old})$, и полное правдоподобие усредняется по полученному распределению:

$$\mathbb{E}_{T|X,\Theta_{old}} \log p(X,T|\Theta) = \int \log p(X,T|\Theta) p(T|X,\Theta_{old}) dT$$

- М-шаг. Фиксируется апостериорное распределение $p(T|X,\Theta_{old})$, и производится поиск новых значений параметров Θ_{new}:

$$\Theta_{new} = \arg\max_{\Theta} \mathbb{E}_{T|X,\Theta_{old}} \log p(X,T|\Theta)$$

- Шаги Е и М повторяются до сходимости.

Максимизация апостериорного распределения

EM-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

Задача

$$p(\Theta|X) \to \max_{\Theta} \Leftrightarrow F = \log p(X|\Theta) + \log p(\Theta) \to \max_{\Theta}$$

Справедливо разложение

$$F = L(q, \theta) + \log p(\Theta) + KL(q||p) \geq L(q, \Theta) + \log p(\Theta)$$

Е-шаг остаётся без изменений.
Модификация М-шага:

$$\mathbb{E}_{T|X,\Theta_{old}} \log p(X, T|\Theta) + \log p(\Theta) \to \max_{\Theta}$$

План лекции

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

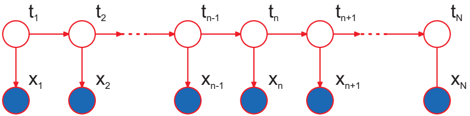
Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

1. ЕМ-алгоритм в общем виде
2. Скрытая марковская модель
3. Метод максимального правдоподобия для СММ
4. Алгоритм «вперед-назад»
5. Устойчивые формулы для алгоритма «вперед-назад»
6. Модельный пример

Скрытая марковская модель (СММ)



Скрытая марковская модель [первого порядка] — это вероятностная модель последовательности, которая

- Состоит из набора наблюдаемых переменных $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ и латентных (скрытых) переменных $T = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N\}$.
- Латентные переменные T являются дискретными, поэтому их иногда называют переменными состояния.
- Значение наблюдаемого вектора в момент времени n $\boldsymbol{x}_n$ зависит только от скрытого состояния $\boldsymbol{t}_n$, которое в свою очередь зависит только от скрытого состояния в предыдущий момент времени $\boldsymbol{t}_{n-1}$.

Спецификация вероятностной модели I

Пусть имеется K возможных состояний. Закодируем состояние в каждый момент времени n бинарным вектором $\boldsymbol{t}_n = (t_{n1}, \dots, t_{nK})$, где

$$t_{nj} = \begin{cases} 1, & \text{если в момент } n \text{ модель находится в состоянии } j \\ 0, & \text{иначе} \end{cases}$$

Тогда распределение $p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})$ можно задать матрицей перехода A размера $K \times K$, где $A_{ij} = p(t_{nj} = 1|t_{n-1,i} = 1)$, $\sum_j A_{ij} = 1$, т.е.

$$p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1}) = \prod_{i=1}^{K}\prod_{j=1}^{K} A_{ij}^{t_{n-1,i}t_{nj}}$$

Пусть в первый момент времени $p(t_{1j} = 1) = \pi_j$. Тогда

$$p(\boldsymbol{t}_1) = \prod_{j=1}^{K} \pi_j^{t_{1j}}$$

Спецификация вероятностной модели II

Пусть для каждого состояния $k \in \{1, \ldots, K\}$ в момент времени n известна модель генерации наблюдаемых данных $\boldsymbol{x}_n$ $p(\boldsymbol{x}_n|\boldsymbol{\phi}_k)$, задаваемая с помощью набора параметров $\boldsymbol{\phi}_k$. Тогда

$$p(\boldsymbol{x}_n|\boldsymbol{t}_n) = \prod_{j=1}^{K} (p(\boldsymbol{x}_n|\boldsymbol{\phi}_k))^{t_{nk}}$$

Обозначим полный набор параметров СММ через $\Theta = \{\boldsymbol{\pi}, A, \boldsymbol{\phi}\}$. Тогда правдоподобие в СММ вычисляется как

$$p(X, T|\Theta) = p(\boldsymbol{t}_1, _{\boldsymbol{\pi}}) \prod_{n=2}^{N} p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1}, _A) \prod_{n=1}^{N} p(\boldsymbol{x}_n|\boldsymbol{t}_n, _{\boldsymbol{\phi}}) =$$

$$= \prod_{j=1}^{K} \pi_j^{t_{1j}} \left(\prod_{n=2}^{N} \prod_{i=1}^{K} \prod_{j=1}^{K} A_{ij}^{t_{n-1,i} t_{nj}} \right) \left(\prod_{n=1}^{N} \prod_{k=1}^{K} (p(\boldsymbol{x}_n|\boldsymbol{\phi}_k))^{t_{nk}} \right)$$

Задачи в СММ

- Сегментация. Известна некоторая последовательность X и набор параметров Θ. Задача состоит в получении наиболее правдоподобной последовательности состояний T как $\arg\max_T p(T|X,\Theta)$ (алгоритм Витерби).
- Обучение с учителем. Известна некоторая последовательность X, для которой заданы T. Задача состоит в оценке по обучающей выборке набора параметров Θ.
- Обучение без учителя. Известна некоторая последовательность X и число состояний K. Задача состоит в оценке параметров Θ (ЕМ-алгоритм).
  - Нахождение маргинального распределения $p(\boldsymbol{t}_n|X,\Theta)$ компоненты $\boldsymbol{t}_n$ по заданным X и Θ
- Прогнозирование. Известна некоторая последовательность X. Задача состоит в оценке наблюдаемого вектора в следующий момент времени $N+1$ — $p(\boldsymbol{x}_{N+1}|X)$.

План лекции

1. ЕМ-алгоритм в общем виде
2. Скрытая марковская модель
3. Метод максимального правдоподобия для СММ
4. Алгоритм «вперед-назад»
5. Устойчивые формулы для алгоритма «вперед-назад»
6. Модельный пример

Метод максимального правдоподобия

Для оценки параметров СММ Θ воспользуемся методом максимального правдоподобия:

$$\Theta_* = \arg\max_{\Theta} p(X|\Theta) = \arg\max_{\Theta} \sum_{T} p(X, T|\Theta)$$

Прямая максимизация правдоподобия затруднительна, т.к. оптимизируемая функция не является выпуклой и, кроме того, для вычисления функции требуется суммирование N^K слагаемых. Можно воспользоваться итерационным ЕМ-алгоритмом.

ЕМ-алгоритм в общем виде

Требуется найти максимум правдоподобия в вероятностной модели со скрытыми переменными:

$$p(X|\Theta) = \sum_T p(X,T|\Theta) \to \max_\Theta \Leftrightarrow \log\left(\sum_T p(X,T|\Theta)\right) \to \max_\Theta$$

- Е-шаг. Фиксируется значение параметров Θ_{old}. Оценивается апостериорное распределение на скрытые переменные $p(T|X,\Theta_{old})$, и полное правдоподобие усредняется по полученному распределению:
$$\mathbb{E}_{T|X,\Theta_{old}} \log p(X,T|\Theta) = \sum_T \log p(X,T|\Theta) p(T|X,\Theta_{old})$$
- М-шаг. Фиксируется апостериорное распределение $p(T|X,\Theta_{old})$, и производится поиск новых значений параметров Θ_{new}:
$$\Theta_{new} = \arg\max_\Theta \mathbb{E}_{T|X,\Theta_{old}} \log p(X,T|\Theta)$$
- Шаги Е и М повторяются до сходимости.

E-шаг

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

Обозначим

$$\gamma(t_{nj}) = \mathbb{E}_{T|X,\Theta} t_{nj} = p(t_{nj} = 1|X, \Theta_{old}),$$
$$\xi(t_{n-1,i}, t_{nj}) = \mathbb{E}_{T|X,\Theta}(t_{n-1,i} t_{nj}) = p(t_{n-1,i} = 1, t_{nj} = 1|X, \Theta_{old}).$$

Тогда

$$\log p(X, T|\Theta) = \sum_{j=1}^{K} t_{1j} \log \pi_j + \sum_{n=2}^{N}\sum_{i=1}^{K}\sum_{j=1}^{K} t_{n-1,i} t_{nj} \log A_{ij} + \sum_{n=1}^{N}\sum_{j=1}^{K} t_{nj} \log p(\boldsymbol{x}_n|\phi_k)$$

$$\mathbb{E}_{T|X,\Theta_{old}} \log p(X, T|\Theta) = \sum_{j=1}^{K} \gamma(t_{1j}) \log \pi_j + \sum_{n=2}^{N}\sum_{i=1}^{K}\sum_{j=1}^{K} \xi(t_{n-1,i}, t_{nj}) \log A_{ij} + \sum_{n=1}^{N}\sum_{j=1}^{K} \gamma(t_{nj}) \log p(\boldsymbol{x}_n|\phi_k)$$

М-шаг

$$\boldsymbol{\pi}: \qquad \sum_{j=1}^{K} \gamma(t_{1j}) \log \pi_j + \lambda \left(\sum_{j=1}^{K} \pi_j - 1 \right) \to \mathop{\mathrm{extr}}_{\boldsymbol{\pi},\lambda}$$

$$\frac{\gamma(t_{1j})}{\pi_j} + \lambda = 0 \Rightarrow \pi_j = -\frac{\gamma(t_{1j})}{\lambda}$$

$$\sum_{j=1}^{K} \pi_j = 1 \Rightarrow \lambda = -\sum_{i=1}^{K} \gamma(t_{1i})$$

$$\pi_j^{new} = \frac{\gamma(t_{1j})}{\sum_{i=1}^{K} \gamma(t_{1i})}$$

Действуя аналогично для A, принимая во внимание, что $\sum_{j=1}^{K} A_{ij} = 1 \ \forall \ i$, получаем:

$$A_{ij}^{new} = \frac{\sum_{n=2}^{N} \xi(t_{n-1,i} t_{nj})}{\sum_{k=1}^{K} \sum_{n=2}^{N} \xi(t_{n-1,i} t_{nk})}$$

М-шаг для компонент ϕ

М-шаг для компонент генерации данных $p(\boldsymbol{x}_n|\phi_k)$ абсолютно аналогичен М-шагу для оценки параметров при восстановлении смесей распределений. В частности, если в качестве компонент выступают многомерные нормальные распределения

$$p(\boldsymbol{x}_n|\phi_k) = \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \Sigma_k),$$

то задача оптимизации для параметров $\boldsymbol{\mu}_k, \Sigma_k$ может быть решена в явном виде:

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^N \gamma(t_{nk})\boldsymbol{x}_n}{\sum_{n=1}^N \gamma(t_{nk})}$$

$$\Sigma_k = \frac{\sum_{n=1}^N \gamma(t_{nk})(\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T}{\sum_{n=1}^N \gamma(t_{nk})}$$

Инициализация параметров Θ

Для начала работы ЕМ-алгоритма необходимо задать начальные значения параметров $\Theta = (\boldsymbol{\pi}, A, \boldsymbol{\phi})$. Заметим, что если какой-нибудь параметр инициализирован нулем, то в процессе итераций ЕМ его значение не изменится.

- Значения параметров $\boldsymbol{\pi}$ и A обычно выбираются случайными при соблюдении ограничений $\sum_j \pi_j = 1$ и $\sum_j A_{ij} = 1 \ \forall i$.
- Инициализация $\boldsymbol{\phi}$ зависит от формы распределений $p(\mathbf{x}|\boldsymbol{\phi})$. В случае нормальных распределений можно провести кластеризацию данных на K кластеров и выбрать в качестве $\boldsymbol{\mu}_k$ и Σ_k центр и разброс соответствующего кластера.

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

План лекции

1. ЕМ-алгоритм в общем виде
2. Скрытая марковская модель
3. Метод максимального правдоподобия для СММ
4. Алгоритм «вперед-назад»
5. Устойчивые формулы для алгоритма «вперед-назад»
6. Модельный пример

Вычисление апостериорных распределений на скрытые компоненты

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

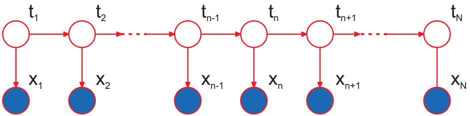
Модельный пример

На Е-шаге алгоритма обучения СММ требуется вычисление апостериорных распределений на скрытые компоненты

$$\gamma(t_{nj}) = p(t_{nj} = 1|X, \Theta), \quad \xi(t_{n-1,i}, t_{nj}) = p(t_{n-1,i} = 1, t_{nj} = 1|X, \Theta)$$

Алгоритм «вперед-назад» (Баума-Уэлша) позволяет эффективно вычислять эти величины для всех n, i, j за линейное по N время.
В дальнейшем для удобства будем опускать Θ во всех формулах, считая набор параметров фиксированным.

Свойства условной независимости для СММ



$$
\begin{aligned}
p(X|\boldsymbol{t}_n) &= p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n|\boldsymbol{t}_n)p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_n)\\
p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1}|\boldsymbol{x}_n,\boldsymbol{t}_n) &= p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1}|\boldsymbol{t}_n)\\
p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1}|\boldsymbol{t}_{n-1},\boldsymbol{t}_n) &= p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1}|\boldsymbol{t}_{n-1})\\
p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_n,\boldsymbol{t}_{n+1}) &= p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_{n+1})\\
p(\boldsymbol{x}_{n+2},\dots,\boldsymbol{x}_N|\boldsymbol{t}_{n+1},\boldsymbol{x}_{n+1}) &= p(\boldsymbol{x}_{n+2},\dots,\boldsymbol{x}_N|\boldsymbol{t}_{n+1})\\
p(X|\boldsymbol{t}_{n-1},\boldsymbol{t}_n) &= p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1}|\boldsymbol{t}_{n-1})p(\boldsymbol{x}_n|\boldsymbol{t}_n)\times\\
&\times p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_n)\\
p(\boldsymbol{x}_{N+1}|X,\boldsymbol{t}_{N+1}) &= p(\boldsymbol{x}_{N+1}|\boldsymbol{t}_{N+1})\\
p(\boldsymbol{t}_{N+1}|\boldsymbol{t}_N,X) &= p(\boldsymbol{t}_{N+1}|\boldsymbol{t}_N)
\end{aligned}
$$

Вычисление $\gamma(t_{nj})$

По формуле Байеса

$$\gamma(\boldsymbol{t}_n) = p(\boldsymbol{t}_n|X) = \frac{p(X|\boldsymbol{t}_n)p(\boldsymbol{t}_n)}{p(X)}$$

Пользуясь свойствами условной независимости для СММ, получаем

$$\gamma(\boldsymbol{t}_n) = \frac{p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{t}_n)p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_n)}{p(X)} = \frac{\boldsymbol{\alpha}(\boldsymbol{t}_n)\boldsymbol{\beta}(\boldsymbol{t}_n)}{p(X)}$$

Здесь

$$\boldsymbol{\alpha}(\boldsymbol{t}_n) = p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{t}_n)$$
$$\boldsymbol{\beta}(\boldsymbol{t}_n) = p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_n)$$

Алгоритм «вперед-назад» позволяет быстро рекуррентно вычислять значение $\boldsymbol{\alpha}(\boldsymbol{t}_n)$ через $\boldsymbol{\alpha}(\boldsymbol{t}_{n-1})$ (проход вперед) и $\boldsymbol{\beta}(\boldsymbol{t}_n)$ через $\boldsymbol{\beta}(\boldsymbol{t}_{n+1})$ (проход назад).

Рекуррентная формула для $\boldsymbol{\alpha}(\boldsymbol{t}_n)$

$$\begin{aligned}
\alpha(\boldsymbol{t}_n) &= p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n,\boldsymbol{t}_n) = \\
&= p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n|\boldsymbol{t}_n)p(\boldsymbol{t}_n) = \\
&= p(\boldsymbol{x}_n|\boldsymbol{t}_n)p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1}|\boldsymbol{t}_n)p(\boldsymbol{t}_n) = \\
&= p(\boldsymbol{x}_n|\boldsymbol{t}_n)p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1},\boldsymbol{t}_n) = \\
&= p(\boldsymbol{x}_n|\boldsymbol{t}_n)\sum_{\boldsymbol{t}_{n-1}} p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1},\boldsymbol{t}_{n-1},\boldsymbol{t}_n) = \\
&= p(\boldsymbol{x}_n|\boldsymbol{t}_n)\sum_{\boldsymbol{t}_{n-1}} p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1},\boldsymbol{t}_n|\boldsymbol{t}_{n-1})p(\boldsymbol{t}_{n-1}) = \\
&= p(\boldsymbol{x}_n|\boldsymbol{t}_n)\sum_{\boldsymbol{t}_{n-1}} p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1}|\boldsymbol{t}_{n-1})p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})p(\boldsymbol{t}_{n-1}) = \\
&= p(\boldsymbol{x}_n|\boldsymbol{t}_n)\sum_{\boldsymbol{t}_{n-1}} p(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{n-1},\boldsymbol{t}_{n-1})p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1}) = \\
&= p(\boldsymbol{x}_n|\boldsymbol{t}_n)\sum_{\boldsymbol{t}_{n-1}} \alpha(\boldsymbol{t}_{n-1})p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})
\end{aligned}$$

Вычисление $\boldsymbol{\alpha}(\boldsymbol{t}_n)$

Для запуска рекуррентного процесса необходимо вычислить

$$\boldsymbol{\alpha}(\boldsymbol{t}_1) = p(\boldsymbol{x}_1, \boldsymbol{t}_1) = p(\boldsymbol{t}_1)p(\boldsymbol{x}_1|\boldsymbol{t}_1) = \prod_{j=1}^{K}(\pi_j p(\boldsymbol{x}_1|\phi_k))^{t_{1j}}$$

На каждом шаге рекурсии вектор $\boldsymbol{\alpha}(\boldsymbol{t}_n)$ длины K умножается на матрицу $p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})$ размера $K \times K$. Поэтому сложность всего рекуррентного процесса составляет $O(NK^2)$.

Рекуррентная формула для $\beta(\boldsymbol{t}_n)$

$$
\begin{aligned}
\beta(\boldsymbol{t}_n) =& p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_n) = \\
&= \sum_{\boldsymbol{t}_{n+1}} p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N,\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) = \\
&= \sum_{\boldsymbol{t}_{n+1}} p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_{n+1})p(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) = \\
&= \sum_{\boldsymbol{t}_{n+1}} p(\boldsymbol{x}_{n+2},\dots,\boldsymbol{x}_N|\boldsymbol{t}_{n+1})p(\boldsymbol{x}_{n+1}|\boldsymbol{t}_{n+1})p(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) = \\
&= \sum_{\boldsymbol{t}_{n+1}} \beta(\boldsymbol{t}_{n+1})p(\boldsymbol{x}_{n+1}|\boldsymbol{t}_{n+1})p(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n)
\end{aligned}
$$

Инициализация рекуррентного процесса для $\beta(t_n)$

Вспомним формулу, связывающую значение $\gamma(t_n)$ с $\alpha(t_n)$ и $\beta(t_n)$:

$$\gamma(t_n) = \frac{p(x_1, \ldots, x_n, t_n)p(x_{n+1}, \ldots, x_N|t_n)}{p(X)} = \frac{\alpha(t_n)\beta(t_n)}{p(X)}$$

Подставляя в формулу значение $n = N$, получаем

$$\gamma(t_N) = p(t_N|X) = \frac{p(X, t_N)\beta(t_N)}{p(X)}$$

Таким образом, $\beta(t_N) = 1$.

Нормировочная константа $p(X)$

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

Значение $\boldsymbol{\gamma}(\boldsymbol{t}_n)$ определено с точностью до нормировочной константы $p(X)$. Однако, на М-шаге значение $\boldsymbol{\gamma}(\boldsymbol{t}_n)$, как правило, входит в числитель и в знаменатель. Таким образом, нормировочная константа сокращается. Например, при вычислении центра нормального распределения:

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} \gamma(t_{nk})\boldsymbol{x}_n}{\sum_{n=1}^{N} \gamma(t_{nk})} = \frac{\sum_{n=1}^{N} \alpha(t_{nk})\beta(t_{nk})\boldsymbol{x}_k}{\sum_{n=1}^{N} \alpha(t_{nk})\beta(t_{nk})}$$

Тем не менее, сама нормировочная константа $p(X)$ — это значение правдоподобия, которое может представлять отдельный интерес (например, можно отслеживать возрастание правдоподобия при итерациях ЕМ-алгоритма). Зная $\boldsymbol{\alpha}(\boldsymbol{t}_n)$ и $\boldsymbol{\beta}(\boldsymbol{t}_n)$, правдоподобие может быть легко вычислено как

$$p(X) = \sum_{\boldsymbol{t}_n} \boldsymbol{\alpha}(\boldsymbol{t}_n)\boldsymbol{\beta}(\boldsymbol{t}_n),\ \forall n \ \Rightarrow\ p(X) = \sum_{\boldsymbol{t}_N} \boldsymbol{\alpha}(\boldsymbol{t}_N)$$

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

Формула для $\xi(\boldsymbol{t}_{n-1}, \boldsymbol{t}_n)$

$$\xi(\boldsymbol{t}_{n-1}, \boldsymbol{t}_n) = p(\boldsymbol{t}_{n-1}, \boldsymbol{t}_n | X) =$$

$$= \frac{p(X|\boldsymbol{t}_{n-1}, \boldsymbol{t}_n)p(\boldsymbol{t}_{n-1}, \boldsymbol{t}_n)}{p(X)} =$$

$$= \frac{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}|\boldsymbol{t}_{n-1})p(\boldsymbol{x}_n|\boldsymbol{t}_n)p(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_N|\boldsymbol{t}_n)p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})p(\boldsymbol{t}_{n-1})}{p(X)} =$$

$$= \frac{\boldsymbol{\alpha}(\boldsymbol{t}_{n-1})p(\boldsymbol{x}_n|\boldsymbol{t}_n)p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})\boldsymbol{\beta}(\boldsymbol{t}_n)}{p(X)}$$

Итоговый EM-алгоритм для случая, когда $p(\mathbf{x}_n|\mathbf{t}_n)$ — нормальные распределения

- Начальная инициализация параметров $\Theta = (\boldsymbol{\pi}, A, \boldsymbol{\phi})$. Параметры $\boldsymbol{\pi}$ и A можно инициализировать случайно, а $\boldsymbol{\phi}$ с помощью кластеризации данных.
- Е-шаг. Вычисление $\boldsymbol{\alpha}(\mathbf{t}_n)$ и $\boldsymbol{\beta}(\mathbf{t}_n)$ с помощью рекуррентного алгоритма «вперед-назад». Вычисление величин $\gamma(t_{nj})$ и $\xi(t_{n-1,i}, t_{nj})$ и, возможно, правдоподобия $p(X)$.
- М-шаг. Вычисление новых значений параметров:

$$\pi_j^{new} = \frac{\gamma(t_{1j})}{\sum_{i=1}^K \gamma(t_{1i})}, \qquad A_{ij}^{new} = \frac{\sum_{n=2}^N \xi(t_{n-1,i} t_{nj})}{\sum_{k=1}^K \sum_{n=2}^N \xi(t_{n-1,i} t_{nk})}$$

$$\boldsymbol{\mu}_k^{new} = \frac{\sum_{n=1}^N \gamma(t_{nk})\mathbf{x}_n}{\sum_{n=1}^N \gamma(t_{nk})}, \quad \Sigma_k^{new} = \frac{\sum_{n=1}^N \gamma(t_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T}{\sum_{n=1}^N \gamma(t_{nk})}$$

- Повторять шаги Е и М до сходимости (пока значение $p(X)$ или Θ не стабилизируется).

Задача прогнозирования

$$p(\boldsymbol{x}_{N+1}|X) = \sum_{\boldsymbol{t}_{N+1}} p(\boldsymbol{x}_{N+1}, \boldsymbol{t}_{N+1}|X) = \sum_{\boldsymbol{t}_{N+1}} p(\boldsymbol{x}_{N+1}|\boldsymbol{t}_{N+1}) p(\boldsymbol{t}_{N+1}|X) =$$

$$= \sum_{\boldsymbol{t}_{N+1}} p(\boldsymbol{x}_{N+1}|\boldsymbol{t}_{N+1}) \left(\sum_{\boldsymbol{t}_N} p(\boldsymbol{t}_{N+1}, \boldsymbol{t}_N|X) \right) =$$

$$= \sum_{\boldsymbol{t}_{N+1}} p(\boldsymbol{x}_{N+1}|\boldsymbol{t}_{N+1}) \left(\sum_{\boldsymbol{t}_N} p(\boldsymbol{t}_N|X) p(\boldsymbol{t}_{N+1}|\boldsymbol{t}_N) \right) =$$

$$= \sum_{\boldsymbol{t}_{N+1}} p(\boldsymbol{x}_{N+1}|\boldsymbol{t}_{N+1}) \left(\sum_{\boldsymbol{t}_N} p(\boldsymbol{t}_{N+1}|\boldsymbol{t}_N) \frac{p(\boldsymbol{t}_N, X)}{p(X)} \right) =$$

$$= \frac{1}{p(X)} \sum_{\boldsymbol{t}_{N+1}} p(\boldsymbol{x}_{N+1}|\boldsymbol{t}_{N+1}) \left(\sum_{\boldsymbol{t}_N} p(\boldsymbol{t}_{N+1}|\boldsymbol{t}_N) \boldsymbol{\alpha}(\boldsymbol{t}_N) \right)$$

Это фактически смесь распределений с компонентами $p(\boldsymbol{x}_{N+1}|\boldsymbol{t}_{N+1})$ и весами $w_k = \frac{1}{p(X)} \sum_{\boldsymbol{t}_N} p(\boldsymbol{t}_{N+1}|\boldsymbol{t}_N) \boldsymbol{\alpha}(\boldsymbol{t}_N)$. Для получения точечного прогноза $\boldsymbol{x}^*_{N+1}$ можно воспользоваться MCMC.

План лекции

1. ЕМ-алгоритм в общем виде
2. Скрытая марковская модель
3. Метод максимального правдоподобия для СММ
4. Алгоритм «вперед-назад»
5. **Устойчивые формулы для алгоритма «вперед-назад»**
6. Модельный пример

Необходимость устойчивых вычислений для $\boldsymbol{\alpha}(\boldsymbol{t}_n)$ и $\boldsymbol{\beta}(\boldsymbol{t}_n)$

ЕМ-алгоритм. Обучение скрытых марковских моделей без учителя.

Ветров

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

Формулы пересчета для $\boldsymbol{\alpha}(\boldsymbol{t}_n)$ и $\boldsymbol{\beta}(\boldsymbol{t}_n)$:

$$\begin{aligned}\boldsymbol{\alpha}(\boldsymbol{t}_n) &= p(\boldsymbol{x}_n|\boldsymbol{t}_n)\sum_{\boldsymbol{t}_{n-1}}\boldsymbol{\alpha}(\boldsymbol{t}_{n-1})p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})\\ \boldsymbol{\beta}(\boldsymbol{t}_n) &= \sum_{\boldsymbol{t}_{n+1}}\boldsymbol{\beta}(\boldsymbol{t}_{n+1})p(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n)p(\boldsymbol{x}_{n+1}|\boldsymbol{t}_{n+1})\end{aligned}$$

На практике значения вероятностей $p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})$ и $p(\boldsymbol{x}_n|\boldsymbol{t}_n)$ могут быть существенно меньше единицы. В процессе пересчета эти вероятности умножаются друг на друга, и получающиеся значения перестают укладываться в машинную точность.

Устойчивые формулы для $\boldsymbol{\alpha}(\boldsymbol{t}_n)$ I

ЕМ-алгоритм. Обучение скрытых марковских моделей без учителя.

Ветров

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

Предлагается вместо $\boldsymbol{\alpha}(\boldsymbol{t}_n)$ рассмотреть следующую величину:

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{t}_n) = p(\boldsymbol{t}_n | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{\boldsymbol{\alpha}(\boldsymbol{t}_n)}{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)}$$

Можно надеяться, что значения $\hat{\boldsymbol{\alpha}}(\boldsymbol{t}_n)$ будут существенно отличны от нуля, т.к. $\sum_{\boldsymbol{t}_n} \hat{\boldsymbol{\alpha}}(\boldsymbol{t}_n) = 1$.
Рассмотрим также величины

$$c_n = p(\boldsymbol{x}_n | \boldsymbol{x}_{n-1}, \ldots, \boldsymbol{x}_1).$$

Вычисление этих величин также будет устойчивым, т.к. они имеют смысл одномерных вероятностных распределений.

Устойчивые формулы для $\boldsymbol{\alpha}(\boldsymbol{t}_n)$ II

Очевидно, что

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \prod_{i=1}^{n} c_i$$

$$\boldsymbol{\alpha}(\boldsymbol{t}_n) = p(\boldsymbol{t}_n | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \hat{\boldsymbol{\alpha}}(\boldsymbol{t}_n) \left(\prod_{i=1}^{n} c_i \right)$$

Подставляя это выражение в формулу пересчета для $\boldsymbol{\alpha}(\boldsymbol{t}_n)$

$$\boldsymbol{\alpha}(\boldsymbol{t}_n) = p(\boldsymbol{x}_n | \boldsymbol{t}_n) \sum_{\boldsymbol{t}_{n-1}} \boldsymbol{\alpha}(\boldsymbol{t}_{n-1}) p(\boldsymbol{t}_n | \boldsymbol{t}_{n-1}),$$

получаем

$$c_n \hat{\boldsymbol{\alpha}}(\boldsymbol{t}_n) = p(\boldsymbol{x}_n | \boldsymbol{t}_n) \sum_{\boldsymbol{t}_{n-1}} \hat{\boldsymbol{\alpha}}(\boldsymbol{t}_{n-1}) p(\boldsymbol{t}_n | \boldsymbol{t}_{n-1})$$

Значение c_n определяется из условия нормировки для $\hat{\boldsymbol{\alpha}}(\boldsymbol{t}_n)$.

Устойчивые формулы для $\beta(\boldsymbol{t}_n)$

Рассмотрим следующую величину

$$\hat{\beta}(\boldsymbol{t}_n) = \frac{p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{t}_n)}{p(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N|\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)} = \frac{\beta(\boldsymbol{t}_n)}{\prod_{i=n+1}^{N} c_i}$$

Вычисление данной величины будет устойчивым, т.к. $\hat{\beta}(\boldsymbol{t}_n)$ является отношением двух распределений на $(\boldsymbol{x}_{n+1},\dots,\boldsymbol{x}_N)$. Подставляя выражение для $\hat{\beta}(\boldsymbol{t}_n)$ в формулу пересчета

$$\beta(\boldsymbol{t}_n) = \sum_{\boldsymbol{t}_{n+1}} \beta(\boldsymbol{t}_{n+1}) p(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) p(\boldsymbol{x}_{n+1}|\boldsymbol{t}_{n+1}),$$

получаем

$$c_{n+1}\hat{\beta}(\boldsymbol{t}_n) = \sum_{\boldsymbol{t}_{n+1}} \hat{\beta}(\boldsymbol{t}_{n+1}) p(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) p(\boldsymbol{x}_{n+1}|\boldsymbol{t}_{n+1})$$

Значения c_n определяются из формул пересчета для $\hat{\alpha}(\boldsymbol{t}_n)$.

Окончательные выражения для E-шага

Правдоподобие $p(X)$ вычисляется как

$$p(X) = \prod_{n=1}^{N} c_n$$

Другие необходимые величины вычисляются как

$$\begin{aligned}
\gamma(\boldsymbol{t}_n) &= \hat{\alpha}(\boldsymbol{t}_n)\hat{\beta}(\boldsymbol{t}_n) \\
\xi(\boldsymbol{t}_{n-1}, \boldsymbol{t}_n) &= \frac{1}{c_n}\hat{\alpha}(\boldsymbol{t}_{n-1})p(\boldsymbol{x}_n|\boldsymbol{t}_n)p(\boldsymbol{t}_n|\boldsymbol{t}_{n-1})\hat{\beta}(\boldsymbol{t}_n)
\end{aligned}$$

План лекции

1. ЕМ-алгоритм в общем виде
2. Скрытая марковская модель
3. Метод максимального правдоподобия для СММ
4. Алгоритм «вперед-назад»
5. Устойчивые формулы для алгоритма «вперед-назад»
6. Модельный пример

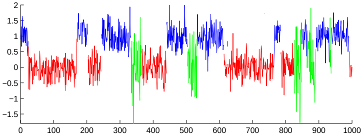
Модельный пример I

В качестве иллюстрации работы ЕМ-алгоритма обучения СММ рассмотрим простой модельный пример. $X \in \mathbb{R}$, $K = 3$, данные сгенерированы из нормальных распределений со следующими параметрами:

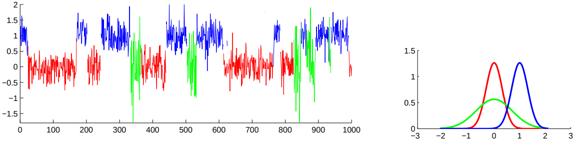
$$\mu_1 = 0, \quad \mu_2 = 0, \quad \mu_3 = 1$$
$$\sigma_1^2 = 0.1, \; \sigma_2^2 = 0.5, \; \sigma_3^2 = 0.1$$

Априорные вероятности и матрица перехода выбраны следующим образом:

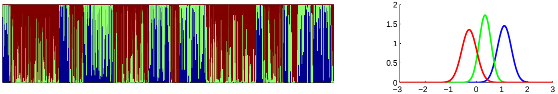
$$\boldsymbol{\pi} = [0.3, \; 0.2, \; 0.5], \qquad A = \begin{bmatrix} 0.98 & 0.01 & 0.01 \\ 0.01 & 0.97 & 0.02 \\ 0.01 & 0.01 & 0.98 \end{bmatrix}$$



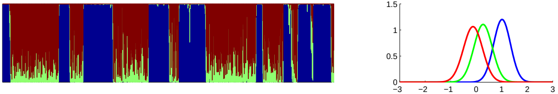
Модельный пример II



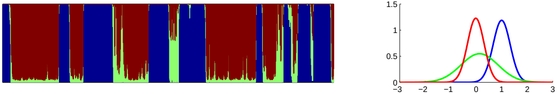
Ит. 1:



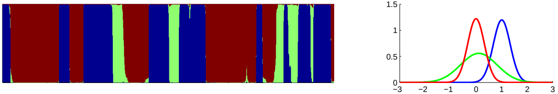
Ит. 5:



Ит. 20:



Ит. 54:



Модельный пример III

ЕМ-алгоритм в общем виде

Скрытая марковская модель

Метод максимального правдоподобия для СММ

Алгоритм «вперед-назад»

Устойчивые формулы для алгоритма «вперед-назад»

Модельный пример

После 54-ой итерации ЕМ-алгоритма значения параметров были следующие:

$$\boldsymbol{\pi} = [10^{-190},\ 10^{-125},\ 1], \qquad A = \begin{bmatrix} 0.984 & 0.004 & 0.012 \\ 0.013 & 0.949 & 0.038 \\ 0.011 & 0.011 & 0.978 \end{bmatrix}$$

$$\mu_1 = -0.01,\ \sigma_1^2 = 0.11$$
$$\mu_2 = 0.1,\ \sigma_2^2 = 0.51$$
$$\mu_3 = 1,\ \sigma_3^2 = 0.11$$