

Лекция 3. Скрытые марковские модели.

Д. П. Ветров¹

¹МГУ, ВМиК, каф. ММП

Курс «Графические модели», 2012г.

План

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

1 Ликбез

Метод динамического программирования

2 Основы применения СММ

Определение СММ

Обучение СММ с учителем

Алгоритм Витерби

3 EM-алгоритм

Графические модели с неполными данными

Разделение гауссовской смеси

План

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Метод
динамического
программирова-
ния

Основы
применения
СММ

EM-алгоритм

1 Ликбез

Метод динамического программирования

2 Основы применения СММ

Определение СММ

Обучение СММ с учителем

Алгоритм Витерби

3 EM-алгоритм

Графические модели с неполными данными

Разделение гауссовской смеси

Задача объезда стран

Лекция 3.
Скрытые
марковские
модели.

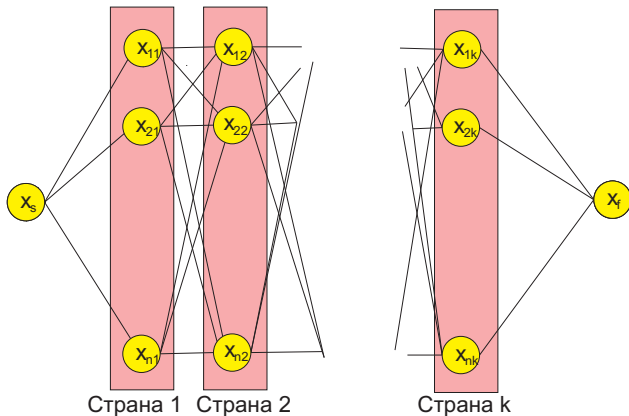
Ветров

Ликбез

Метод
динамического
программирова-
ния

Основы
применения
СММ

EM-алгоритм



- Рассмотрим такую задачу: необходимо в определенной последовательности объехать k стран, в каждой из которых провести одну ночь в отеле, потратив минимум денег
- Если в каждой стране имеется n городов, то задача сводится к перебору n^k вариантов

Функция Беллмана

Лекция 3.
Скрытые
марковские
модели.

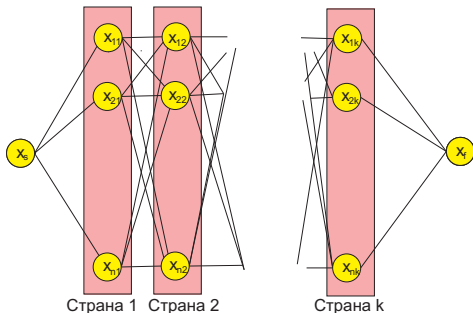
Ветров

Ликбез

Метод
динамического
программирова-
ния

Основы
применения
СММ

EM-алгоритм



- Пусть цена билета между городами x_{ij} и $x_{i+1,l}$ задается функцией $f(x_{ij}, x_{i+1,l})$, а цена ночлега в городе — функцией $h(x_{ij})$
- Подсчитаем функцию Беллмана $V(x)$, определяемую рекуррентно: $V(x_s) = 0$

$$V(x_{i+1,l}) = \min_j [V(x_{ij}) + f(x_{ij}, x_{i+1,l}) + h(x_{i+1,l})]$$

- Физический смысл функции Беллмана это наименьшая сумма денег, которую нужно потратить, чтобы добраться из города x_s в город $x_{i+1,l}$ (легко показать по индукции)

Динамическое программирование

Лекция 3.
Скрытые
марковские
модели.

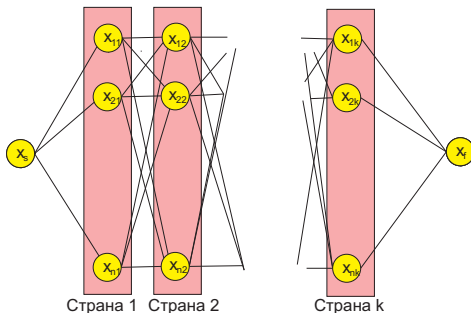
Ветров

Ликбез

Метод
динамического
программирова-
ния

Основы
применения
СММ

EM-алгоритм



- Определим также функцию $S(x)$, возвращающую город, откуда мы приехали в город x : $S(x_s) = \emptyset$

$$S(x_{i+1,l}) = \arg \min_j [V(x_{ij}) + f(x_{ij}, x_{i+1,l}) + h(x_{i+1,l})]$$

- Тогда оптимальный путь $(x_1^*, x_2^*, \dots, x_k^*)$ может быть получен путем рекуррентного вызова функции $S(x)$: $x_n^* = S(x_f)$

$$x_m^* = S(x_{m+1}^*)$$

Особенности динамического программирования

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

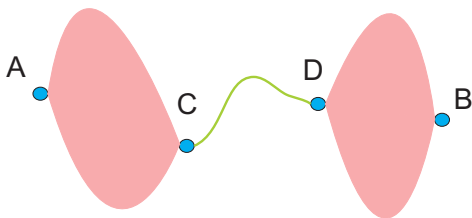
Метод
динамического
программирова-
ния

Основы
применения
СММ

EM-алгоритм

Метод динамического программирования эффективен, если имеет место

- Перекрывающиеся подзадачи, которые необходимо решить, чтобы решить исходную задачу
- Оптимальная подструктура (принцип кусочной оптимальности)
- Возможность запомнить решения подзадач



Если известно, что оптимальный путь проходит через точки C и D и известен оптимальный путь между ними, то этот путь станет частью оптимального пути между A и B

План

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

1 Ликбез

Метод динамического программирования

2 Основы применения СММ

Определение СММ

Обучение СММ с учителем

Алгоритм Витерби

3 EM-алгоритм

Графические модели с неполными данными

Разделение гауссовской смеси

Скрытая Марковская модель (СММ)

Лекция 3.
Скрытые
марковские
модели.

Ветров

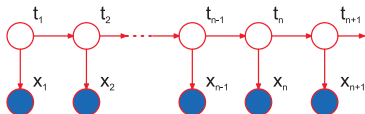
Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем
Алгоритм
Витерби

EM-алгоритм



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

- Состоит из набора наблюдаемых переменных $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, $\mathbf{x}_n \in \mathbb{R}^d$ и латентных (скрытых) переменных $T = \{t_1, \dots, t_N\}$, $t_n \in \{0, 1\}^K$, $\sum_{j=1}^K t_{nj} = 1$
- Латентные переменные T являются **бинарными** и кодируют K состояний, поэтому их иногда называют переменными состояния
- Значение наблюдаемого вектора \mathbf{x}_n , взятого в момент времени n , зависит только от скрытого состояния t_n , которое в свою очередь зависит только от скрытого состояния в предыдущий момент времени t_{n-1} .

Примеры использования СММ

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

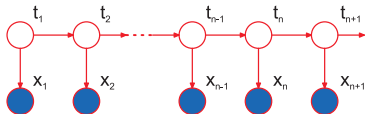
Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм



Что можно анализировать с помощью СММ

- Речь
- Видео
- Поведение
- Фондовые рынки
- Естественный язык
- ДНК
- и др.

Скрытая Марковская модель (СММ)

Лекция 3.
Скрытые
марковские
модели.

Ветров

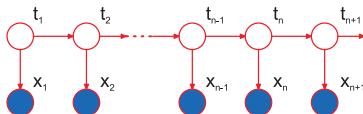
Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем
Алгоритм
Витерби

EM-алгоритм



- Скрытая марковская модель является частным случаем байесовской сети (графической модели, задаваемой ориентированным графом)
- Граф, задающий СММ, является ациклическим, поэтому для СММ существуют эффективные алгоритмы вывода
- Для полного задания модели достаточно задать все условные распределения вида $p(\mathbf{x}_n | \mathbf{t}_n)$, $p(\mathbf{t}_n | \mathbf{t}_{n-1})$ и априорное распределение $p(\mathbf{t}_1)$

Спецификация вероятностной модели

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем
Алгоритм
Витерби

EM-алгоритм

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

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

Тогда распределение $p(\mathbf{t}_n | \mathbf{t}_{n-1})$ 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.

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

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

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

Замечание

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

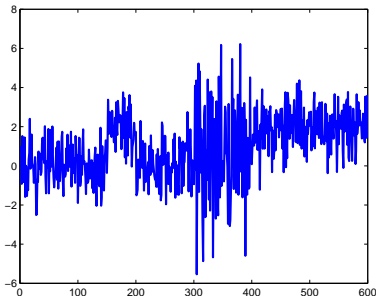
Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

- Хотя матрица A может быть произвольного вида с учетом ограничений на неотрицательность и сумму элементов строки, с точки зрения СММ представляет интерес диагональное преобладание матрицы перехода
- В этом случае можно ожидать, что процесс находится в некотором состоянии на протяжении какого-то отрезка времени
- Появляется простая физическая интерпретация СММ: имеется процесс, который иногда (относительно редко) скачкообразно меняет свои характеристики



Спецификация вероятностной модели

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем
Алгоритм
Витерби

EM-алгоритм

- Условное распределение $p(\mathbf{x}_n | \mathbf{t}_n)$ определяется текущим состоянием \mathbf{t}_n
- Обычно предполагают, что оно нам известно с точностью до параметров ϕ_k , $k \in \{1, \dots, K\}$, т.е. если $t_{n1} = 1$, то \mathbf{x}_n взят из распределения $p(\mathbf{x}_n | \phi_1)$, если $t_{n2} = 1$, то \mathbf{x}_n взят из распределения $p(\mathbf{x}_n | \phi_2)$, и т.д.
- Таким образом

$$p(\mathbf{x}_n | \mathbf{t}_n) = \prod_{j=1}^K (p(\mathbf{x}_n | \phi_k))^{t_{nk}}$$

Задачи в СММ

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем
Алгоритм
Витерби

EM-алгоритм

Обозначим полный набор параметров $\Theta = \{\pi, A, \phi\}$. Тогда основные задачи, возникающие в СММ, можно сформулировать следующим образом:

- **Обучение с учителем.** Известна некоторая последовательность X , для которой заданы T . Задача состоит в оценке по обучающей выборке набора параметров Θ .
- **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 без учителя.** Известна некоторая последовательность X и число состояний K . Задача состоит в оценке параметров Θ (EM-алгоритм).
 - **Нахождение маргинального распределения** $p(t_n|X, \Theta)$ компоненты t_n по заданным X и Θ
- **Прогнозирование.** Известна некоторая последовательность X . Задача состоит в оценке наблюдаемого вектора в следующий момент времени $N + 1 - p(x_{N+1}|X)$.

План

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

1 Ликбез

Метод динамического программирования

2 Основы применения СММ

Определение СММ

Обучение СММ с учителем

Алгоритм Витерби

3 EM-алгоритм

Графические модели с неполными данными

Разделение гауссовской смеси

Совместное распределение переменных СММ

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

- Предположим, нам задана **обучающая выборка** (X, T) , представляющая собой одну или несколько последовательностей, в которых известны значения скрытых компонент
- Требуется оценить вектор параметров Θ
- По построению байесовской сети совместное распределение переменных задается формулой

$$\begin{aligned} p(X, T | \Theta) &= p_{\pi}(t_1) \prod_{n=1}^N p_{\phi}(\mathbf{x}_n | t_n) \prod_{n=2}^N p_A(t_n | t_{n-1}) = \\ &= \prod_{j=1}^K \pi_j^{t_{1j}} \left(\prod_{n=2}^N \prod_{i=1}^K \prod_{j=1}^K A_{ij}^{t_{n-1}, t_{nj}} \right) \left(\prod_{n=1}^N \prod_{k=1}^K (p(\mathbf{x}_n | \phi_k))^{t_{nk}} \right) \end{aligned}$$

Метод максимального правдоподобия

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

- Для оценки параметров Θ воспользуемся методом максимального правдоподобия

$$\Theta_{ML} = \arg \max p(X, T | \Theta) = \arg \max \log p(X, T | \Theta)$$

- Для удобства перейдем к логарифму функции правдоподобия (положение максимума, очевидно, не изменится)

$$\begin{aligned} \log p(X, T | \Theta) = & \left(\sum_{j=1}^K t_{1j} \log \pi_j \right) + \\ & + \left(\sum_{n=2}^N \sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \left(\sum_{n=1}^N \sum_{k=1}^K t_{nk} \log p(\mathbf{x}_n | \phi_k) \right) \end{aligned}$$

Функция Лагранжа

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

- Параметры, входящие в Θ , не могут принимать произвольные значения, следовательно необходима оптимизация при ограничениях

$$\sum_{j=1}^K \pi_j = 1, \quad \sum_{j=1}^K A_{ij} = 1, \quad \forall i = \overline{1, K}$$

- Воспользуемся правилом множителей Лагранжа и выпишем лагранжиан

$$\mathcal{L}(\Theta, \lambda, \mu) = \log p(X, T | \Theta) + \lambda \left(\sum_{j=1}^K \pi_j - 1 \right) + \sum_{i=1}^K \mu_i \left(\sum_{j=1}^K A_{ij} - 1 \right) \rightarrow \text{extr}$$

Оценка π

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

$$\begin{aligned}\log p(X, T|\Theta) &= \left(\sum_{j=1}^K t_{1j} \log \pi_j \right) + \\ &+ \left(\sum_{n=2}^N \sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \left(\sum_{n=1}^N \sum_{k=1}^K t_{nk} \log p(x_n | \phi_k) \right) \\ \mathcal{L}(\Theta, \lambda, \mu) &= \log p(X, T|\Theta) + \lambda \left(\sum_{j=1}^K \pi_j - 1 \right) + \sum_{i=1}^K \mu_i \left(\sum_{j=1}^K A_{ij} - 1 \right) \rightarrow \text{extr} \\ \frac{\partial \mathcal{L}(\Theta, \lambda, \mu)}{\partial \pi_j} &= \frac{t_{1j}}{\pi_j} + \lambda = 0 \Rightarrow \pi_j = -\frac{t_{1j}}{\lambda} \\ \sum_{j=1}^K \pi_j &= 1 \Rightarrow \lambda = -\sum_{j=1}^K t_{1j} = -1 \\ \pi_j &= t_{1j}\end{aligned}$$

Оценка матрицы A

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

$$\log p(X, T|\Theta) = \left(\sum_{j=1}^K t_{1j} \log \pi_j \right) +$$
$$+ \left(\sum_{n=2}^N \sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \left(\sum_{n=1}^N \sum_{k=1}^K t_{nk} \log p(\mathbf{x}_n | \phi_k) \right)$$
$$\mathcal{L}(\Theta, \lambda, \boldsymbol{\mu}) = \log p(X, T|\Theta) + \lambda \left(\sum_{j=1}^K \pi_j - 1 \right) + \sum_{i=1}^K \mu_i \left(\sum_{j=1}^K A_{ij} - 1 \right) \rightarrow \text{extr}$$

$$\frac{\partial \mathcal{L}(\Theta, \lambda, \boldsymbol{\mu})}{\partial A_{ij}} = \sum_{n=2}^N \frac{t_{n-1,i} t_{nj}}{A_{ij}} + \mu_i = 0 \Rightarrow A_{ij} = - \sum_{n=2}^N \frac{t_{n-1,i} t_{nj}}{\mu_i}$$

$$\sum_{j=1}^K A_{ij} = 1 \Rightarrow \mu_i = - \sum_{n=2}^N \sum_{j=1}^K t_{n-1,i} t_{nj} = - \sum_{n=2}^N t_{n-1,i}$$

$$A_{ij} = \frac{\sum_{n=2}^N t_{n-1,i} t_{nj}}{\sum_{n=2}^N t_{n-1,i}}$$

Оценка условной плотности $p(\mathbf{x}|\mathbf{t})$

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

$$\begin{aligned}\log p(X, T|\Theta) &= \left(\sum_{j=1}^K t_{1j} \log \pi_j \right) + \\ &+ \left(\sum_{n=2}^N \sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \left(\sum_{n=1}^N \sum_{k=1}^K t_{nk} \log p(\mathbf{x}_n|\phi_k) \right) \\ \mathcal{L}(\Theta, \lambda, \mu) &= \log p(X, T|\Theta) + \lambda \left(\sum_{j=1}^K \pi_j - 1 \right) + \sum_{i=1}^K \mu_i \left(\sum_{j=1}^K A_{ij} - 1 \right) \rightarrow \text{extr} \\ \frac{\partial \mathcal{L}(\Theta, \lambda, \mu)}{\partial \phi_k} &= \sum_{n=1}^N \sum_{k=1}^K t_{nk} \frac{\partial \log p(\mathbf{x}_n|\phi_k)}{\phi_k} = \sum_{\{n: t_{nk}=1\}} \frac{\partial \log p(\mathbf{x}_n|\phi_k)}{\phi_k} = 0\end{aligned}$$

Получили стандартную задачу максимизации правдоподобия по выборке независимых одинаково-распределенных объектов

$$\phi_k = \arg \max \sum_{\{n: t_{nk}=1\}} \log p(\mathbf{x}_n|\phi_k)$$

Для оценки параметров ϕ_k можно воспользоваться методами восстановления плотностей, например, EM-алгоритмом

План

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

1 Ликбез

Метод динамического программирования

2 Основы применения СММ

Определение СММ

Обучение СММ с учителем

Алгоритм Витерби

3 EM-алгоритм

Графические модели с неполными данными

Разделение гауссовской смеси

Сегментация сигнала

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

- Пусть известна некоторая последовательность наблюдений X и набор параметров СММ Θ . Требуется определить наиболее вероятную последовательность состояний T , т.е. найти $\arg \max_T p(T|X, \Theta)$
- Заметим, что $p(X|\Theta)$ не зависит от T , поэтому

$$\begin{aligned}\arg \max_T p(T|X, \Theta) &= \arg \max_T \frac{p(X, T|\Theta)}{p(X|\Theta)} = \\ &= \arg \max_T p(X, T|\Theta) = \arg \max_T \log p(X, T|\Theta)\end{aligned}$$

- Но это же классическая задача динамического программирования!

Аналогия с задачей объезда стран

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

- В самом деле логарифм совместной плотности по определению

$$\begin{aligned} \log p(X, T | \Theta) = & \left(\sum_{j=1}^K t_{1j} \log \pi_j \right) + \\ & + \left(\sum_{n=2}^N \sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \left(\sum_{n=1}^N \sum_{k=1}^K t_{nk} \log p(\mathbf{x}_n | \phi_k) \right) \end{aligned}$$

- Первое слагаемое определяет «пункт отбытия», второе слагаемое — стоимость переезда из города страны $n - 1$ в город в стране n , а третье слагаемое отражает «стоимость ночлега» в выбранном городе страны n

Алгоритм Витерби

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

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

Функция Беллмана $V(t_{1j}) = \log \pi_j$

$$V(t_{nj}) = \max_i \left[V(t_{n-1,i}) + \left(\sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \log p(\mathbf{x}_n | \phi_j) \right]$$

Функция $S(t_{nj})$ определяется аналогично: $S(t_{1j}) = \emptyset$

$$S(t_{nj}) = \arg \max_i \left[V(t_{n-1,i}) + \left(\sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \log p(\mathbf{x}_n | \phi_j) \right]$$

Алгоритм Витерби

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

$$V(t_{nj}) = \max_i \left[V(t_{n-1,i}) + \left(\sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \log p(\mathbf{x}_n | \phi_j) \right]$$
$$S(t_{nj}) = \arg \max_i \left[V(t_{n-1,i}) + \left(\sum_{i=1}^K \sum_{j=1}^K t_{n-1,i} t_{nj} \log A_{ij} \right) + \log p(\mathbf{x}_n | \phi_j) \right]$$

Выполнив прямой проход по сигналу, мы оцениваем $V(t_{nj})$ и $S(t_{nj})$, а выполнив обратный проход, мы получаем оптимальные номера оптимальных состояний $(i^*(1), \dots, i^*(N))$: $i^*(N) = \arg \max_i V(t_{Ni})$

$$i^*(n) = S(t_{n+1, i^*(n+1)})$$

Легко видеть, что значения переменных t_n определяются так:
 $t_{n, i^*(n)} = 1, t_{ni} = 0, \forall i \neq i^*(n)$

- Алгоритм Витерби позволяет быстро проводить сегментацию очень длинных сигналов
- Существует версия алгоритма Витерби, позволяющая осуществлять сегментацию в реальном времени (точнее, с небольшой задержкой)

Пример использования

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

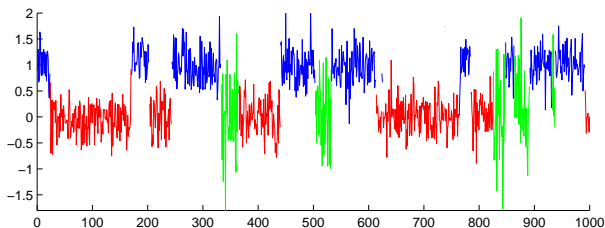
Определение
СММ

Обучение СММ
с учителем

Алгоритм
Витерби

EM-алгоритм

Пример разметки сигнала на три состояния с помощью алгоритма Витерби



План

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

1 Ликбез

Метод динамического программирования

2 Основы применения СММ

Определение СММ

Обучение СММ с учителем

Алгоритм Витерби

3 EM-алгоритм

Графические модели с неполными данными

Разделение гауссовской смеси

Максимум неполного правдоподобия

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

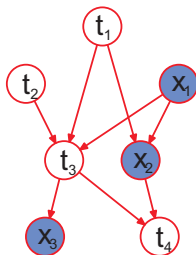
EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

- Предположим имеется графическая модель, в которой известна только часть значений переменных
- Атомарные распределения известны с точностью до вектора параметров θ
- Требуется оценить параметры по наблюдаемым величинам с помощью метода максимального правдоподобия, т.е. найти

$$\theta_{ML} = \arg \max p(X|\theta)$$



Трудности оптимизации неполного правдоподобия

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

- По правилу суммирования вероятностей неполное правдоподобие может быть получено в виде суммирования по скрытым переменным полного правдоподобия

$$p(X|\theta) = \sum_T p(X, T|\theta)$$

- Во многих случаях (в частности, в байесовских сетях) подсчет полного правдоподобия тривиален
- При оптимизации правдоподобия удобно переходить к логарифму, в частности выше мы получили явные формулы для $\arg \max_{\theta} p(X, T|\theta) = \arg \max_{\theta} \log p(X, T|\theta)$ в СММ
- Прямая оптимизация логарифма неполного правдоподобия очень затруднительна даже в итерационной форме, т.к. функционал имеет вид «логарифм суммы», в то время как удобно оптимизировать «сумму логарифмов»

$$\log p(X|\theta) = \log \sum_T p(X, T|\theta)$$

Схема EM-алгоритма

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

- На входе: выборка X , зависящая от набора параметров θ
- Инициализируем θ некоторыми начальным приближением
- **E-шаг:** Оцениваем распределение скрытой компоненты при фиксированном значении параметров θ_{old}

$$p(T|X, \theta_{old}) = \frac{p(X, T|\theta_{old})}{\sum_T p(X, T|\theta_{old})}$$

- **M-шаг:** Оптимизируем

$$\mathbb{E}_{T|X, \theta_{old}} \log p(X, T|\theta) = \sum_T p(T|X, \theta_{old}) \log p(X, T|\theta) \rightarrow \max_{\theta}$$

Если бы мы **точно знали значение $T = T_0$** , то вместо мат. ожидания по всевозможным (с учетом наблюдаемых данных) $T|X, \theta_{old}$, мы бы оптимизировали $\log p(X, T_0|\theta)$

- Переход к E-шагу, пока процесс не сойдется

Замечания

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

- Оптимизация проводится итерационно методом покоординатного спуска: на каждой итерации последовательно уточняются возможные значения T (E-шаг), а потом пересчитываются значения θ (M-шаг)
- Во многих случаях на M-шаге можно получить явные формулы, т.к. там происходит оптимизация выпуклой комбинации логарифмов полных правдоподобий

$$\sum_T p(T|X, \theta_{old}) \log p(X, T|\theta) \rightarrow \max_{\theta},$$

имеющей вид взвешенной «суммы логарифмов»

План

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

1 Ликбез

Метод динамического программирования

2 Основы применения СММ

Определение СММ

Обучение СММ с учителем

Алгоритм Витерби

3 EM-алгоритм

Графические модели с неполными данными

Разделение гауссовской смеси

Смесь гауссовских распределений

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

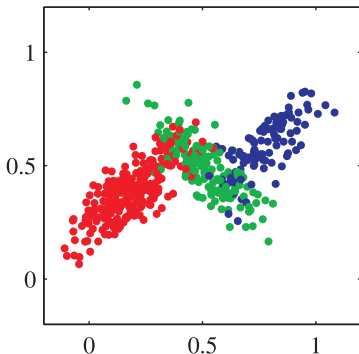
Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

- Имеется выборка $X \sim \sum_{j=1}^l w_j \mathcal{N}(\mathbf{x} | \mu_j, \Sigma_j) \in \mathbb{R}^d$, $w_j \geq 0$,
 $\sum_{j=1}^l w_j = 1$
- Требуется восстановить плотность генеральной совокупности



EM-алгоритм

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

- Выбираем начальное приближение μ_j, w_j, Σ_j
- E-шаг: Вычисляем распределение скрытых переменных $z_i \in \{0, 1\}$, $\sum_j z_{ij} = 1$, которые определяют, к какой компоненте смеси принадлежит объект x_i

$$\gamma(z_{ij}) = \frac{w_j \mathcal{N}(x_i | \mu_j, \Sigma_j)}{\sum_{k=1}^l w_k \mathcal{N}(x_i | \mu_k, \Sigma_k)}$$

- M-шаг: С учетом новых вероятностей на z_i , пересчитываем параметры смеси

$$\mu_j^{new} = \frac{1}{N_j} \sum_{i=1}^n \gamma(z_{ij}) x_i \quad w_j^{new} = \frac{N_j}{n} \quad N_j = \sum_{i=1}^n \gamma(z_{ij})$$

$$\Sigma_j^{new} = \frac{1}{N_j} \sum_{i=1}^n \gamma(z_{ij}) (x_i - \mu_j^{new})(x_i - \mu_j^{new})^T$$

- Переход к E-шагу, пока не будет достигнута сходимость

Вывод формул на M-шаге

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

Выражение для мат. ожидания логарифма правдоподобия по апостериорному распределению имеет вид

$$\mathbb{E}_{Z|X, \theta} \log p(X, Z | \theta) = \sum_{i=1}^n \sum_{j=1}^l \gamma(z_{ij}) (\log w_j + \log \mathcal{N}(x_i | \mu_j, \Sigma_j))$$

Дифференцирование по μ_j и Σ_j (точнее по Σ_j^{-1}) выполняется аналогично случаю одной многомерной гауссианы, только объекты теперь берутся с весом $\gamma(z_{ij})$.

Оптимизация по w_j проводится с учетом ограничения $\sum_{j=1}^l w_j = 1$ по правилу множителей Лагранжа

$$\sum_{i=1}^n \sum_{j=1}^l \gamma(z_{ij}) \log w_j + \lambda \left(\sum_{j=1}^l w_j - 1 \right)$$

Дифференцируя лагранжиан по w_j получаем уравнения

$$\frac{\sum_{i=1}^n \gamma(z_{ij})}{w_j} + \lambda = 0, \quad w_j = -\frac{\sum_{i=1}^n \gamma(z_{ij})}{\lambda}$$

Учитывая, что $\sum_{j=1}^l w_j = 1$ окончательно получаем

$$\lambda = -n, \quad w_j = \frac{\sum_{i=1}^n \gamma(z_{ij})}{n} = \frac{N_j}{n}$$

Пример работы EM-алгоритма

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

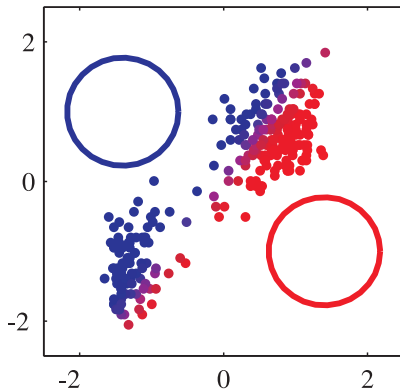
Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

Применение EM-алгоритма для разделения смеси двух гауссиан. Итерация 0.



Пример работы EM-алгоритма

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

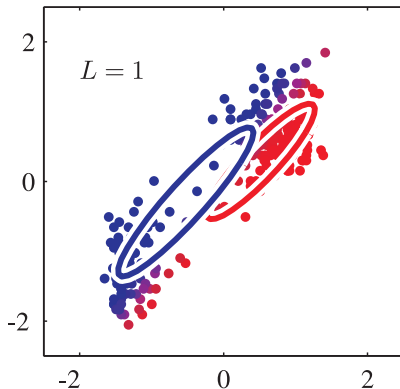
Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

Применение EM-алгоритма для разделения смеси двух гауссиан. Итерация 1.



Пример работы EM-алгоритма

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

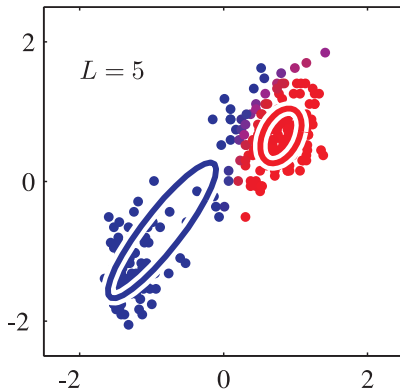
Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

Применение EM-алгоритма для разделения смеси двух гауссиан. Итерация 5.



Пример работы EM-алгоритма

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

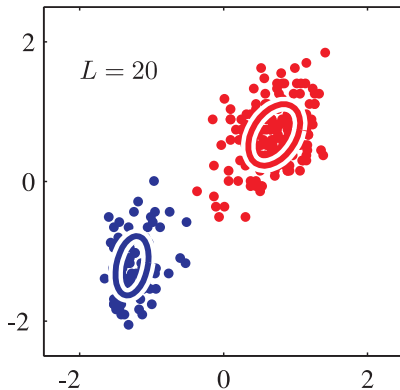
Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

Применение EM-алгоритма для разделения смеси двух гауссиан. Итерация 20.



Недостатки EM-алгоритма

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

- В зависимости от выбора начального приближения может сходиться к разным точкам
- EM-алгоритм находит локальный экстремум, в котором значение правдоподобия может оказаться намного ниже, чем в глобальном максимуме
- EM-алгоритм **не позволяет определить количество компонентов смеси l**
- **!! Величина l является структурным параметром!!**

Применение EM-алгоритма для описания сложных плотностей

Лекция 3.
Скрытые
марковские
модели.

Ветров

Ликбез

Основы
применения
СММ

EM-алгоритм

Графические
модели с
неполными
данными

Разделение
гауссовской
смеси

Смесью гауссиан можно эффективно приближать сложные, плохо-параметризуемые плотности. При этом получаются аналитически заданные приближения

