Скрытые марковские модели: Алгоритм Баума-Велча

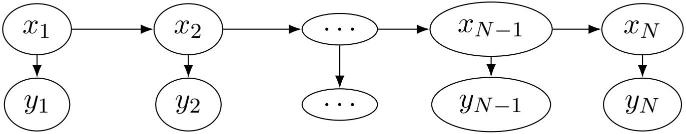
Александр Адуенко

2е апреля 2024

Содержание предыдущих лекций

- EM-алгоритм. Использование EM-алгоритма для отбора признаков в байесовской линейной регрессии.
- Вариационный EM-алгоритм и его использование для вывода в смеси моделей линейной регрессии.
- Ориентированные графические модели и их представление plate notation. Критерий условной независимости d-separation.
- Неориентированные графические модели и их связь с ориентированными.
- Факторные графы и алгоритм Sum-Product для вывода в ациклических графических моделях.
- Скрытые марковские модели и алгоритм Витерби. Алгоритм Max-Sum как обобщение алгоритма Витерби.

Скрытые марковские модели



$$p(\mathbf{x},\, \mathbf{y}) = p(x_1)\prod_{i=2}^{N} p(x_i|x_{i-1})\prod_{i=1}^{N} p(y_i|x_i).$$

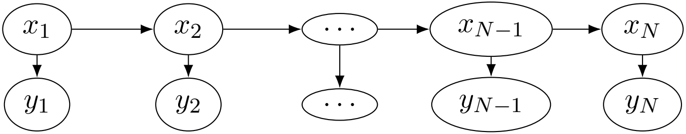
Пусть $x_i \in [K]$, $\mathbf{A} = \|a_{ij}\| = \|\mathsf{P}(x_l = j|x_{l-1} = i)\|$, $\pi_k = \mathsf{P}(x_1 = k)$.

$$p(\mathbf{x},\, \mathbf{y}|\mathbf{A},\, \boldsymbol{\pi},\, \mathbf{B}) = p(x_1|\boldsymbol{\pi})\prod_{i=2}^{N} p(x_i|x_{i-1},\, \mathbf{A})\prod_{i=1}^{N} p(y_i|x_i,\, \mathbf{B}).$$

Задачи:

- $p(x_i|\mathbf{y},\, \mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi})$ – алгоритм Sum-Product;
- $p(x_i,\, x_{i+1}|\mathbf{y},\, \mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi})$ – алгоритм Sum-Product;
- $p(\mathbf{x}|\mathbf{y},\, \mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}) \to \max\limits_{\mathbf{x}}$ – алгоритм Витерби / Max-Sum;
- $p(\mathbf{x}|\mathbf{y},\, \mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi})$ – последовательное сэмплирование;
- $p(\mathbf{y}|\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}) \to \max\limits_{\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}}$.

Сэмплирование состояний СММ (HMM)



Задача: Найти $p(\mathbf{x}|\mathbf{y},\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi})$.

$$p(\mathbf{x},\ \mathbf{y}|\mathbf{A},\ \boldsymbol{\pi},\ \mathbf{B}) = p(x_1|\boldsymbol{\pi})\prod_{i=2}^{N} p(x_i|x_{i-1},\ \mathbf{A})\prod_{i=1}^{N} p(y_i|x_i,\ \mathbf{B}).$$

$$p(\mathbf{x}|\mathbf{y},\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) \propto \underbrace{p(x_1|\boldsymbol{\pi})p(y_1|x_1,\ \mathbf{B})}_{\tilde{g}(x_1)}\underbrace{p(x_2|x_1,\ \mathbf{A})p(y_2|x_2,\ \mathbf{B})}_{\tilde{g}(x_2|x_1)}\cdot\ldots\cdot$$

$$\underbrace{p(x_N|x_{N-1},\ \mathbf{A})p(y_N|x_N,\ \mathbf{B})}_{\tilde{g}(x_N|x_{N-1})}.$$

Идея:

- $x_1 \sim g(x_1)$;
- $x_2 \sim g(x_2|x_1)$;
- $\vdots$
- $x_N \sim g(x_N|x_{N-1})$.

Пусть $\mathbf{D} = (\mathbf{X},\ \mathbf{y})$ – наблюдаемые переменные, $\mathbf{Z}$ – скрытые переменные.
$p(\mathbf{D},\ \mathbf{Z}|\boldsymbol{\Theta}) = p(\mathbf{D}|\mathbf{Z},\ \boldsymbol{\Theta})p(\mathbf{Z}|\boldsymbol{\Theta})$.

Вопрос 1: как решить задачу $p(\mathbf{D}|\boldsymbol{\Theta}) = \int p(\mathbf{D},\ \mathbf{Z}|\boldsymbol{\Theta})d\mathbf{Z} \to \max\limits_{\boldsymbol{\Theta}}$?

Пример 1. $\mathbf{y} = \mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon},\ \mathbf{w} \sim \mathcal{N}(\mathbf{w}|\mathbf{0},\ \mathbf{A}^{-1}),\ \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\ \beta^{-1}\mathbf{I})$
$p(\mathbf{y},\ \mathbf{w}|\mathbf{X},\ \mathbf{A},\ \beta) = p(\mathbf{y}|\mathbf{X},\ \mathbf{w},\ \beta)p(\mathbf{w}|\mathbf{A})$.
$\log p(\mathbf{y}|\mathbf{X},\ \underbrace{\mathbf{A},\ \beta^{-1}}_{\boldsymbol{\Theta}}) \propto -\frac{1}{2}\log\det(\beta^{-1}\mathbf{I}+\mathbf{X}\mathbf{A}^{-1}\mathbf{X}^{\mathsf{T}}) - \frac{1}{2}\mathbf{y}^{\mathsf{T}}(\beta^{-1}\mathbf{I}+\mathbf{X}\mathbf{A}^{-1}\mathbf{X}^{\mathsf{T}})^{-1}\mathbf{y}$.

ЕМ-алгоритм

Введем $F(q,\ \boldsymbol{\Theta}) = -\int q(\mathbf{Z})\log q(\mathbf{Z})d\mathbf{Z} + \int q(\mathbf{Z})\log p(\mathbf{D},\ \mathbf{Z}|\boldsymbol{\Theta})d\mathbf{Z} =$

$$-\int q(\mathbf{Z})\log q(\mathbf{Z})d\mathbf{Z} + \int q(\mathbf{Z})\log p(\mathbf{Z}|\mathbf{D},\ \boldsymbol{\Theta})d\mathbf{Z} + \int \log p(\mathbf{D}|\boldsymbol{\Theta})q(\mathbf{Z})d\mathbf{Z} =$$

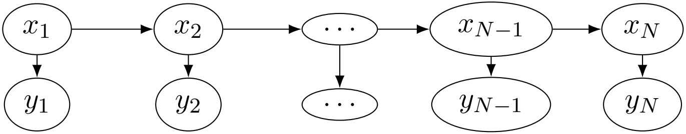
$$\log p(\mathbf{D}|\boldsymbol{\Theta}) - \int q(\mathbf{Z})\log\frac{q(\mathbf{Z})}{p(\mathbf{Z}|\mathbf{D},\boldsymbol{\Theta})}d\mathbf{Z} = \log p(\mathbf{D}|\boldsymbol{\Theta}) - D_{\mathrm{KL}}(q\|p(\mathbf{Z}|\mathbf{D},\ \boldsymbol{\Theta})).$$

Идея 1: $p(\mathbf{D}|\boldsymbol{\Theta}) \to \max\limits_{\boldsymbol{\Theta}}$ заменим на $F(q,\ \boldsymbol{\Theta}) \to \max\limits_{q,\ \boldsymbol{\Theta}}$.

Идея 2: Пошагово оптимизируем по $\boldsymbol{\Theta}$ и q, то есть

1. Е-шаг: $q^s = F(q,\ \boldsymbol{\Theta}^{s-1}) \to \max\limits_{q}$;
2. М-шаг: $\boldsymbol{\Theta}^s = F(q^s,\ \boldsymbol{\Theta}) \to \max\limits_{\boldsymbol{\Theta}}$.

Вывод параметров скрытой марковской модели



Задача: $p(\mathbf{y}|\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}) \to \max\limits_{\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}}.$

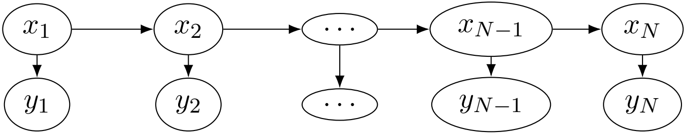
$$p(\mathbf{x},\, \mathbf{y}|\mathbf{A},\, \boldsymbol{\pi},\, \mathbf{B}) = p(x_1|\boldsymbol{\pi}) \prod_{i=2}^{N} p(x_i|x_{i-1},\, \mathbf{A}) \prod_{i=1}^{N} p(y_i|x_i,\, \mathbf{B}).$$

Введем $\mathbf{Z} = \|z_{ik}\|,\ z_{ik} \in \{0,\, 1\}$ и пусть $y_i|x_i = k = \mathcal{N}(y_i|m_k,\, \sigma_k^2)$.

$$p(\mathbf{y},\, \mathbf{z}|\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{1k}} \prod_{i=2}^{N} \prod_{k=1}^{K} \prod_{l=1}^{K} a_{kl}^{z_{i-1,\,k} z_{il}} \prod_{i=1}^{N} \prod_{k=1}^{K} \mathcal{N}(y_i|m_k,\, \sigma_k^2)^{z_{ik}}.$$

Вопрос: Как решить $p(\mathbf{y}|\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}) \to \max\limits_{\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}}$, где $\mathbf{B} = (\mathbf{m},\, \boldsymbol{\sigma}^2)$?

ЕМ-алгоритм для вывода параметров СММ



Задача: $p(\mathbf{y}|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) \to \max\limits_{\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi}}$, где $\mathbf{B} = (\mathbf{m},\ \boldsymbol{\sigma}^2)$.

$$p(\mathbf{y},\ \mathbf{z}|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{1k}} \prod_{i=2}^{N} \prod_{k=1}^{K} \prod_{l=1}^{K} a_{kl}^{z_{i-1,\,k} z_{il}} \prod_{i=1}^{N} \prod_{k=1}^{K} \mathcal{N}(y_i|m_k,\ \sigma_k^2)^{z_{ik}}.$$

$$\log p(\mathbf{y},\ \mathbf{z}|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) = \sum_{k=1}^{K} z_{1k} \log \pi_k + \sum_{i=2}^{N} \sum_{k=1}^{K} \sum_{l=1}^{K} z_{i-1,\,k} z_{il} \log a_{kl} +$$

$$\sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left(-\frac{1}{2} \log \sigma_k^2 - \frac{1}{2\sigma_k^2} (y_i - m_k)^2 \right).$$

Е-шаг: $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{y},\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi})$.

М-шаг: $\mathsf{E}_q \log p(\mathbf{y},\ \mathbf{z}|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) \to \max\limits_{\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi}}$.

М-шаг

$$\mathsf{E}_q \log p(\mathbf{y},\ \mathbf{z}|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) = \sum_{k=1}^{K} \mathsf{E}z_{1k} \log \pi_k + \sum_{i=2}^{N}\sum_{k=1}^{K}\sum_{l=1}^{K} \mathsf{E}z_{i-1,\,k}z_{il} \log a_{kl} +$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \mathsf{E}z_{ik} \left(-\frac{1}{2}\log \sigma_k^2 - \frac{1}{2\sigma_k^2}(y_i - m_k)^2\right).$$

$$\mathsf{E}_q \log p(\mathbf{y},\ \mathbf{z}|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) \to \max_{\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi}}.$$

$$\pi_k = \mathsf{E}z_{1k},\ a_{kl} \propto \sum_{i=2}^{N} \mathsf{E}z_{i-1,\,k}z_{il};$$

$$m_k = \frac{\sum_{i=1}^{N} \mathsf{E}z_{ik}y_i}{\sum_{i=1}^{N} \mathsf{E}z_{ik}},\ \sigma_k^2 = \frac{\sum_{i=1}^{N} \mathsf{E}z_{ik}(y_i - m_k)^2}{\sum_{i=1}^{N} \mathsf{E}z_{ik}}.$$

Вопрос: Что требуется знать про $q(\mathbf{Z})$, чтобы осуществить М-шаг?

E-шаг

Общий шаг: $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{y},\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi})$.
Достаточно: $\mathsf{E}z_{ik}$, $\mathsf{E}z_{i-1,\,k}z_{il}$.
Вопрос: Можно ли воспользоваться алгоритмом Sum-Product для получения $\mathsf{E}z_{ik}$, $\mathsf{E}z_{i-1,\,k}z_{il}$?

Введем $\alpha_k(t) = p(y_1,\ \ldots,\ y_t,\ x_t = k|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi})$ и $\beta_k(t) = p(y_{t+1},\ \ldots,\ y_N|x_t = k,\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi})$.

$$\mathsf{E}z_{tk} = \mathsf{P}(x_t = k|\mathbf{y},\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) \propto \mathsf{P}(x_t = k,\ \mathbf{y}|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) = \underbrace{p(y_1,\ \ldots,\ y_t,\ x_t = k|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi})}_{\alpha_k(t)}\underbrace{p(y_{t+1},\ \ldots,\ y_N|x_t = k,\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi})}_{\beta_k(t)}.$$

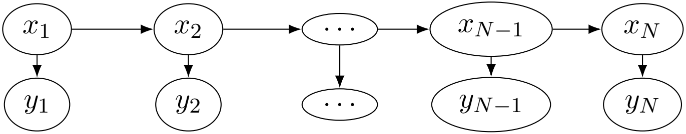
Вопрос 1: Какое свойство СММ было использовано при выводе выше?

$$\mathsf{E}z_{t-1,\,k}z_{tl} = \mathsf{P}(x_{t-1} = k,\ x_t = l|\mathbf{y},\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) \propto p(x_{t-1} = k,\ x_t = l,\ \mathbf{y}|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) = p(y_1,\ \ldots,\ y_{t-1},\ x_{t-1} = k|\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi})p(x_t = l|x_{t-1} = k)p(y_t|x_t = l)p(y_{t+1},\ \ldots,\ y_N|x_t = l,\ \mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}) \Longrightarrow$$

$$\mathsf{E}z_{t-1,\,k}z_{tl} \propto \alpha_k(t-1)a_{kl}p(y_t|x_t = l,\ \mathbf{B})\beta_l(t).$$

Вопрос 2: Какие свойства СММ были использованы при выводе выше?

Е-шаг 2: Получение $\alpha_k(t)$ и $\beta_k(t)$



Сосчитаем $\alpha_k(t) = p(y_1, \ldots, y_t, x_t = k | \mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ пошагово:

- $\alpha_k(1) = \pi_k p(y_1 | x_1 = k, \mathbf{B})$;
- $\alpha_k(t+1) = \sum\limits_{j=1}^{K} \alpha_j(t) a_{jk} p(y_{t+1} | x_{t+1} = k, \mathbf{B})$.

Сосчитаем $\beta_k(t) = p(y_{t+1}, \ldots, y_N | x_t = k, \mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ пошагово:

- $\beta_k(N) = 1$;
- $\beta_k(t) = \sum\limits_{j=1}^{K} \beta_j(t+1) a_{kj} p(y_{t+1} | x_{t+1} = j, \mathbf{B})$.

Вопрос 1: Какие численные проблемы стоит ожидать при вычислениях по описанной схеме?

Вопрос 2: Как разрешить численные проблемы при угасании значений $\alpha_k(t)$, $\beta_k(N-t)$, $t \gg 1$?

EM-алгоритм: Итоговая схема

$p(\mathbf{y}|\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}) \to \max\limits_{\mathbf{A},\, \mathbf{B},\, \boldsymbol{\pi}}$, где $\mathbf{B} = (\mathbf{m},\, \boldsymbol{\sigma}^2)$.

E-шаг:

- Вычисляем $\alpha_k(t),\ \beta_k(t),\ t \in [N],\ k \in [K]$ при фиксированных $\mathbf{A},\ \mathbf{B},\ \boldsymbol{\pi}$;
- $\mathsf{E} z_{tk} \propto \alpha_k(t)\beta_k(t)$ и нормируем;
- $\mathsf{E} z_{t-1,\,k} z_{tl} \propto \alpha_k(t-1) a_{kl} p(y_t|x_t = l,\, \mathbf{B})\beta_l(t)$ и нормируем.

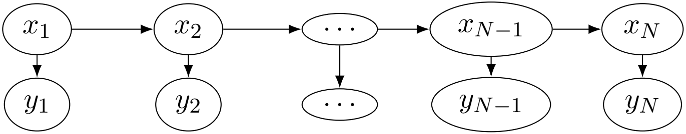
M-шаг:

- $\pi_k = \mathsf{E} z_{1k},\ a_{kl} \propto \sum\limits_{i=2}^{N} \mathsf{E} z_{i-1,\,k} z_{il}$;
- $m_k = \dfrac{\sum_{i=1}^N \mathsf{E} z_{ik} y_i}{\sum_{i=1}^N \mathsf{E} z_{ik}},\ \sigma_k^2 = \dfrac{\sum_{i=1}^N \mathsf{E} z_{ik}(y_i - m_k)^2}{\sum_{i=1}^N \mathsf{E} z_{ik}}$.

Вопрос 1: Как учесть ненаблюдаемость части $\mathbf{y}$?

Вопрос 2: Как предсказать $y_{N+\Delta}$?

Appendix: Вывод формулы пересчета $\alpha_k(t+1)$



$$\color{blue}{\alpha_k(t+1)} = p(y_1,\,\ldots,\,y_{t+1},\,x_{t+1}=k|\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi}) =$$

$$\sum_{j=1}^{K} p(y_1,\,\ldots,\,y_{t+1},\,x_{t+1}=k,\,x_t=j|\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi}) =$$

$$\sum_{j=1}^{K} p(y_1,\,\ldots,\,y_t,\,x_t=j|\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi})p(y_{t+1},\,x_{t+1} =$$

$$k|\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi},\,y_1,\,\ldots,\,y_t,\,\mathbf{x}_t=j) =$$

$$\sum_{j=1}^{K} p(y_1,\,\ldots,\,x_t=j|\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi})p(y_{t+1},\,x_{t+1}=k|\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi},\,\mathbf{x}_t=j) =$$

$$\alpha_j(t)p(x_{t+1}=k|\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi},\,\mathbf{x}_t=j)p(y_{t+1}|x_{t+1}=k,\,\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi},\,\mathbf{x}_t=j) =$$

$$\sum_{j=1}^{K} \alpha_j(t)a_{jk}p(y_{t+1}|x_{t+1}=k,\,\mathbf{A},\,\mathbf{B},\,\boldsymbol{\pi}).$$

Литература

1. Bishop, Christopher M. "Pattern recognition and machine learning". Springer, New York (2006). Pp. 394-418.
2. Material on HMMs: https://web.stanford.edu/ jurafsky/slp3/A.pdf
3. Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.
4. McCallum, Andrew. "Hidden Markov models Baum-Welch algorithm."Introduction to Natural Language Processing CS 585 (2004).
5. Mor, Bhavya, Sunita Garhwal, and Ajay Kumar. "A systematic review of hidden markov models and their applications."Archives of computational methods in engineering 28.3 (2021): 1429-1448.
6. Old Sklearn Class Gaussian HMMs: https://ogrisel.github.io/scikit-learn.org/sklearn-tutorial/modules/generated/sklearn.hmm.GaussianHMM.html
7. New HMM Learn package: https://github.com/hmmlearn/hmmlearn
8. Python library on HMMs: https://pypi.org/project/hmms/#files
9. Python Library on Graphical Models: https://github.com/pgmpy/pgmpy