Методы машинного обучения. Обучение без учителя: байесовские латентные модели

Воронцов Константин Вячеславович
www.MachineLearning.ru/wiki?title=User:Vokov
вопросы к лектору: k.vorontsov@iai.msu.ru

материалы курса:
github.com/MSU-ML-COURSE/ML-COURSE-24-25
орг.вопросы по курсу: ml.cmc@mail.ru

ВМК МГУ • 11 марта 2025

Содержание

Напоминание. Вероятностные модели порождения данных

Дано:

$X = (x_1, \ldots, x_\ell)$ — исходные данные, *наблюдаемые переменные*

Найти:

$p(X|\Omega)$ — модель порождения данных, с параметром Ω

$p(X|\Omega) = \prod_{i=1}^{\ell} p(x_i|\Omega)$ — в случае простой (i.i.d.) выборки

Критерии максимизации:

— *правдоподобия* (Maximum Likelihood Estimate, MLE):

$$\ln p(X|\Omega) = \sum_{i=1}^{\ell} \ln p(x_i|\Omega) \;\to\; \max_{\Omega}$$

— *апостериорной вероятности* (Maximum a Posteriori, MAP):

$$\ln p(X,\Omega) = \ln p(X|\Omega)p(\Omega|\gamma) = \sum_{i=1}^{\ell} \ln p(x_i|\Omega) + \underbrace{\ln p(\Omega|\gamma)}_{R(\Omega)} \;\to\; \max_{\Omega}$$

где $p(\Omega|\gamma)$ — априорное распределение с гиперпараметром γ

Порождающая модель с латентными переменными

Дано:
$X = (x_1, \dots, x_\ell)$ — исходные данные, *наблюдаемые переменные*
$Z = (z_1, \dots, z_m)$ — *латентные (скрытые) переменные*

Найти:
$p(X, Z|\Omega)$ — вероятностная модель совместного порождения наблюдаемых данных и скрытых переменных, с параметром Ω

Критерий:
максимум правдоподобия (Maximum Likelihood Estimate, MLE):

$$\ln p(X|\Omega) = \ln \int_Z p(X, Z|\Omega)\, dZ \;\to\; \max_\Omega$$

Для дискретных переменных Z вместо интеграла $\int_Z$ сумма $\sum_Z$
Договоримся далее dZ опускать

Bishop C. M. Pattern recognition and machine learning. Springer, 2006.

1. Разделение смеси распределений — мягкая кластеризация

$X = (x_1, \dots, x_\ell)$ — исходные данные, *наблюдаемые переменные*
$Z = (j_1, \dots, j_\ell)$ — компонента смеси j_i порождает объект x_i

Порождающая модель смеси k вероятностных распределений:

$$p(x) = \sum_{j=1}^{k} p(x,j) = \sum_{j=1}^{k} p(j)p(x|j) = \sum_{j=1}^{k} w_j \varphi(x, \theta_j)$$

$\Omega = \bigl(w_j, \theta_j\bigr)_{j=1}^{k}$ — параметры модели: $w_j = p(j)$, $\varphi(x,\theta_j) = p(x|j)$

Порождающая модель наблюдаемых и скрытых i.i.d. данных:

$$p(X, Z|\Omega) = \prod_{i=1}^{\ell} p(x_i, j_i) = \prod_{i=1}^{\ell} p(j_i)p(x_i|\theta_{j_i}) = \prod_{i=1}^{\ell} w_{j_i}\varphi(x_i, \theta_{j_i})$$

Задача разделения смеси — максимизация log правдоподобия:

$$\ln p(X|\Omega) = \ln \prod_{i=1}^{\ell} p(x_i) = \sum_{i=1}^{\ell} \ln \sum_{j=1}^{k} w_j \varphi(x_i, \theta_j) \;\to\; \max_{\Omega}$$

2. Тематическое моделирование — мягкая би-кластеризация

$X = (w_1|d_1, \ldots, w_n|d_n)$ — слова $w_i \in W$ в документах $d_i \in D$
$Z = (t_1, \ldots, t_n)$, скрытая тема $t_i \in T$ порождает слово $w_i|d_i$

Вероятностная языковая модель порождения слов документа, при *гипотезе условной независимости* $p(w|t,d) = p(w|t)$:

$$p(w|d) = \sum_{t\in T} p(w,t|d) = \sum_{t\in T} p(w|t)p(t|d) = \sum_{t\in T} \phi_{wt}\theta_{td}$$

$\Omega = (\Phi, \Theta)$ — параметры модели: $\phi_{wt} = p(w|t)$, $\theta_{td} = p(t|d)$

Порождающая модель наблюдаемых и скрытых i.i.d. данных:

$$p(X, Z|\Omega) = \prod_{i=1}^{n} p(w_i, t_i|d_i) = \prod_{i=1}^{n} \phi_{w_i t_i}\theta_{t_i d_i}$$

Задача разделения смеси тем в каждом документе:

$$\ln p(X|\Omega) = \ln \prod_{i=1}^{n} p(w_i|d_i) = \sum_{i=1}^{n} \ln \sum_{t\in T} \phi_{w_i t}\theta_{td_i} \to \max_{\Omega}$$

3. Сегментация временного ряда

$X = (x_i, y_i)_{i=1}^{\ell}$ — наблюдаемый временной ряд, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$
$Z = (s_i)_{i=1}^{\ell}$ — сегменты, $s_i \in \{1, \dots, S\}$, $s_i \leqslant s_{i+1}$, $S \ll \ell$

Порождающая модель сегментированного ряда (x_i, y_i):

$$p(y|x) = \sum_{s=1}^{S} p(y, s|x) = \sum_{s=1}^{S} p(s)p(y|s, x) = \sum_{s=1}^{S} w_s \mathcal{N}(y|f(x, \alpha_s), \sigma_s)$$

$\Omega = (w_s, \alpha_s, \sigma_s)_{s=1}^{S}$ — параметры модели, $\mathsf{E}(y|s, x) = f(x, \alpha_s)$

Порождающая модель наблюдаемых и скрытых i.i.d. данных:

$$p(X, Z|\Omega) = \prod_{i=1}^{\ell} p(y_i, s_i|x_i) = \prod_{i=1}^{\ell} w_{s_i} \mathcal{N}(y|f(x_i, \alpha_{s_i}), \sigma_{s_i})$$

Задача сегментации ряда — максимизация log правдоподобия:

$$\ln p(X|\Omega) = \ln \prod_{i=1}^{\ell} p(y_i|x_i) = \sum_{i=1}^{\ell} \ln \sum_{s=1}^{S} w_s \mathcal{N}(y_i|f(x_i, \alpha_s), \sigma_s) \to \max_{\Omega}$$

Байесовская порождающая модель с латентными переменными

$X = (x_1, \dots, x_\ell)$ — исходные данные, *наблюдаемые переменные*
$Z = (z_1, \dots, z_m)$ — *латентные (скрытые) переменные*
$p(X, Z|\Omega)$ — модель наблюдаемых и скрытых переменных
$p(\Omega|\gamma)$ — *априорное распределение* с гиперпараметрами γ

Найти: распределение $p(\Omega|X)$, затем точечную оценку Ω.

Апостериорное распределение, по формуле Байеса:

$$p(\Omega|X,\gamma) \;=\; \frac{1}{p(X)} p(X|\Omega)\, p(\Omega|\gamma) \;\propto\; \int_Z p(X,Z|\Omega)\, p(\Omega|\gamma)$$

Критерий максимума апостериорной вероятности:

$$\ln p(\Omega|X,\gamma) = \ln \int_Z p(X,Z|\Omega) + \underbrace{\ln p(\Omega|\gamma)}_{R(\Omega)} \;\to\; \max_{\Omega}$$

$R(\Omega) = \ln p(\Omega|\gamma)$ — байесовский регуляризатор, хотя
$R(\Omega)$ может и не иметь вероятностной интерпретации.

Общий EM-алгоритм для задачи со скрытыми переменными

Теорема. Точка Ω локального максимума регуляризованного *маргинализованного правдоподобия* (Marginal log-Likelihood)

$$\ln \int_Z p(X, Z|\Omega) + R(\Omega) \;\to\; \max_{\Omega} \qquad \text{(RML)}$$

удовлетворяет системе уравнений, решение которой методом простых итераций сводится к чередованию двух шагов:

$$\text{E-шаг:}\quad q(Z) = p(Z|X, \Omega);$$

$$\text{M-шаг:}\quad \int_Z q(Z) \ln p(X, Z|\Omega) + R(\Omega) \;\to\; \max_{\Omega}.$$

Общий EM-алгоритм используется не только для разделения смесей, но и в анализе сигналов, изображений, текстов и др.

A.P.Dempster, N.M.Laird, D.B.Rubin. Maximum likelihood from incomplete data via the EM algorithm. 1977.

Откуда название Expectation–Maximization?

Запишем p-ю итерацию в одну строчку:

$$\Omega_{p+1} := \arg\max_{\Omega} \underbrace{\int_Z p(Z|X,\Omega_p) \ln p(X,Z|\Omega)}_{Q(\Omega,\Omega_p)} + R(\Omega)$$

и снова разобьём на два шага, но по-другому:

шаг Expectation: $Q(\Omega,\Omega_p) := \mathsf{E}_Z\bigl(\ln p(X,Z|\Omega) \bigm| X,\Omega_p\bigr)$

шаг Maximization: $\Omega_{p+1} := \arg\max_{\Omega} Q(\Omega,\Omega_p)$

A.P.Dempster, N.M.Laird, D.B.Rubin. Maximum likelihood from incomplete data via the EM algorithm. 1977.

Шлезингер М. И. О самопроизвольном различении образов // Читающие автоматы. Киев: Наукова думка, 1965. С. 38–45.

Шлезингер М. И., Главач В. Десять лекций по статистическому и структурному распознаванию. Киев: Наукова думка, 2004.

Доказательство теоремы

Необходимые условия локального экстремума:

$$\frac{\partial}{\partial\Omega}\left(\ln\int_Z p(X,Z|\Omega)+R(\Omega)\right)=\frac{1}{p(X|\Omega)}\int_Z\frac{\partial p(X,Z|\Omega)}{\partial\Omega}+\frac{\partial R(\Omega)}{\partial\Omega}=0$$

По формуле условной вероятности $p(X|\Omega)=\frac{p(X,Z|\Omega)}{p(Z|X,\Omega)}$, подставляем:

$$\int_Z\frac{p(Z|X,\Omega)}{p(X,Z|\Omega)}\frac{\partial p(X,Z|\Omega)}{\partial\Omega}+\frac{\partial R(\Omega)}{\partial\Omega}=0$$

$$\int_Z\underbrace{p(Z|X,\Omega)}_{q(Z)}\frac{\partial}{\partial\Omega}\ln p(X,Z|\Omega)+\frac{\partial R(\Omega)}{\partial\Omega}=0$$

Это уравнение совпадает с необходимым условием локального экстремума для задачи M-шага, образуя систему уравнений вместе с равенством $q(Z)=p(Z|X,\Omega)$.

■

Ещё более общий EM-алгоритм и его сходимость

Теорема. Значение маргинализованного правдоподобия

$$\ln \int_Z p(X,Z|\Omega) + R(\Omega) \;\to\; \max_{\Omega} \qquad \text{(RML)}$$

не убывает на каждом шаге итерационного процесса

$$\text{E-шаг:} \quad \mathrm{KL}\big(q(Z) \,\|\, p(Z|X,\Omega)\big) \;\to\; \min_{q};$$

$$\text{M-шаг:} \quad \int_Z q(Z) \ln p(X,Z|\Omega) + R(\Omega) \;\to\; \max_{\Omega}.$$

$q(Z) = p(Z|X,\Omega)$ является точным решением задачи E-шага.

Минимизация KL-дивергенции на E-шаге используется в случаях, когда не удаётся вычислить $p(Z|X,\Omega)$ в явном виде.

Сходимость *в слабом смысле*: глобальный max не гарантируется.

Доказательство теоремы

По формуле условной вероятности $p(X|\Omega) = \frac{p(X,Z|\Omega)}{p(Z|X,\Omega)}$.
Для произвольного распределения $q(Z)$

$$\ln p(X|\Omega) = \int_Z q(Z) \ln p(X|\Omega) = \int_Z q(Z) \ln \frac{p(X,Z|\Omega)}{p(Z|X,\Omega)} =$$
$$= \underbrace{\int_Z q(Z) \ln \frac{p(X,Z|\Omega)}{q(Z)}}_{L(q,\Omega)} + \underbrace{\int_Z q(Z) \ln \frac{q(Z)}{p(Z|X,\Omega)}}_{\mathrm{KL}\left(q(Z)\,\middle\|\,p(Z|X,\Omega)\right)\geqslant 0}$$

Максимизируем достижимую нижнюю оценку RML то по q, то по Ω:

Е-шаг: $L(q,\Omega) + \cancel{R(\Omega)} \to \max\limits_q \;\Leftrightarrow\; \mathrm{KL}\big(q(Z) \,\|\, p(Z|X,\Omega)\big) \to \min\limits_q$
М-шаг: $L(q,\Omega) + R(\Omega) \to \max\limits_\Omega \;\Leftrightarrow\; \int_Z q(Z) \ln p(X,Z|\Omega) + R(\Omega) \to \max\limits_\Omega$

На каждом шаге значение функционала может только возрастать, откуда и следует сходимость в слабом смысле. ■

Задача разделения смеси: применение общего EM-алгоритма

$X = (x_1, \ldots, x_\ell)$ — исходные данные, *наблюдаемые переменные*
$Z = (j_1, \ldots, j_\ell)$ — компоненты смеси j_i, порождающие объекты x_i

$$p(X, Z|\Omega) = \prod_{i=1}^{\ell} p(x_i, j_i|\Omega) = \prod_{i=1}^{\ell} p(j_i)p(x_i|\theta_{j_i}) = \prod_{i=1}^{\ell} w_{j_i}\varphi(x_i, \theta_{j_i})$$

E-шаг: в силу независимости элементов выборки и формулы Байеса

$$q(Z) = p(Z|X, \Omega) = \prod_{i=1}^{\ell} p(j_i|x_i, \Omega); \qquad p(j|x, \Omega) = \frac{p(x,j|\Omega)}{p(x|\Omega)} = \frac{w_j\varphi(x,\theta_j)}{\sum_t w_t\varphi(x,\theta_t)}$$

M-шаг: подставим $q(Z)$ и $p(X, Z|\Omega)$ в общую формулу M-шага:

$$\sum_Z q(Z) \ln p(X, Z|\Omega) + R(\Omega) =$$

$$= \sum_{j_1=1}^{k} \cdots \sum_{j_\ell=1}^{k} \prod_{t=1}^{\ell} p(j_t|x_t, \Omega) \sum_{i=1}^{\ell} \ln p(x_i, j_i|\Omega) + R(\Omega) =$$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k} \underbrace{p(j|x_i, \Omega)}_{g_{ij}} \ln \underbrace{p(x_i, j|\Omega)}_{w_j\varphi(x_i,\theta_j)} + R(\Omega) \;\to\; \max_{\Omega}, \qquad \Omega = (w_j, \theta_j)_{j=1}^{k}$$

EM-алгоритм: вывод формул M-шага из условий ККТ

M-шаг распадается на $2k$ подзадач по w_j и θ_j, $j = 1, \dots, k$:

$$\sum_{j=1}^{k} \sum_{i=1}^{\ell} \Bigl(g_{ij} \ln w_j + g_{ij} \ln \varphi(x_i, \theta_j) \Bigr) + R(\Omega) \;\to\; \max_{\Omega}$$

Необходимые условия локального экстремума:

$$\begin{cases} \dfrac{\partial}{\partial \theta_j} \left(\sum\limits_{i=1}^{\ell} g_{ij} \ln \varphi(x_i, \theta_j) + R(\Omega) \right) = 0; \\ \dfrac{\partial}{\partial w_j} \left(\sum\limits_{i=1}^{\ell} g_{ij} \ln w_j + R(\Omega) + \Bigl(1 - \sum\limits_{j=1}^{k} w_j \Bigr) \lambda_j - w_j \mu_j \right) = 0; \\ w_j \geqslant 0, \quad \sum_{j=1}^{k} w_j = 1, \quad w_j \mu_j = 0, \quad \mu_j \geqslant 0 \end{cases}$$

Относительно w_j решение аналитическое, из условий ККТ:

$$w_j = \operatorname*{norm}_j \left(\sum_{i=1}^{\ell} g_{ij} + w_j \frac{\partial R}{\partial w_j} \right), \quad \text{где} \;\; \operatorname*{norm}_j (v_j) \stackrel{def}{=} \frac{\max(v_j, 0)}{\sum_k \max(v_k, 0)}$$

EM-алгоритм для разделения смеси распределений

Задача разделения смеси: $(X, Z) = (x_i, j_i)_{i=1}^{\ell}$, $\Omega = (w_j, \theta_j)_{j=1}^{k}$

Теорема. Точка Ω локального максимума (RML)

$$\ln \sum_Z p(X, Z|\Omega) + R(\Omega) = \sum_{i=1}^{\ell} \ln \sum_{j=1}^{k} w_j \varphi(x_i, \theta_j) + R(\Omega)$$

удовлетворяет системе уравнений, решение которой методом простых итераций сводится к чередованию двух шагов:

E-шаг: $g_{ij} \equiv p(j|x_i) = \operatorname*{norm}_j \bigl(w_j \varphi(x_i, \theta_j)\bigr)$, $i = 1..\ell$, $j = 1..k$;

M-шаг: $\theta_j = \arg\max_{\theta} \left(\sum_{i=1}^{\ell} g_{ij} \ln \varphi(x_i, \theta) + R(\Omega)\right)$, $j = 1..k$;

$$w_j = \operatorname*{norm}_j \left(\sum_{i=1}^{\ell} g_{ij} + w_j \frac{\partial R}{\partial w_j}\right), \; j = 1..k.$$

Регуляризация в EM-алгоритме для разделения смеси

- При $R(\Omega) = 0$ это уже знакомый нам EM-алгоритм для разделения смеси вероятностных распределений
- При разделении смеси n-мерных гауссиан $\mathcal{N}(x; \mu_j, \Sigma_j)$ регуляризация ковариационных матриц $\Sigma_j + \tau I_n$
- $R(w) = \tau \sum_j p_j \ln w_j$ — регуляризатор сглаживания, приближает веса w_j к априорному распределению p_j:

$$w_j = \operatorname*{norm}_j \left(\sum_{i=1}^{\ell} g_{ij} + \tau p_j \right)$$

- $R(w) = -\tau \sum_j \ln w_j$ — регуляризатор разреживания, удаляет компоненты с весами $w_j < \tau$:

$$w_j = \operatorname*{norm}_j \left(\sum_{i=1}^{\ell} g_{ij} - \tau \right)$$

Задача вероятностного тематического моделирования (ВТМ)

W — конечное множество (словарь) термов (слов)
D — конечное множество документов
T — конечное множество тем

$X = (d_i, w_i)_{i=1}^n$ — исходные данные, *наблюдаемые переменные*
$Z = (t_i)_{i=1}^n$ — *латентные (скрытые) переменные*, $t_i \in T$
n_{dw} — частота терма $w \in W$ в документе $d \in D$

$\Omega = (\Phi, \Theta)$ — параметры порождающей модели $p(X|\Omega)$
$\Phi = \bigl(\phi_{wt} = p(w|t)\bigr)_{W \times T}$ — распределения слов в темах
$\Theta = \bigl(\theta_{td} = p(t|d)\bigr)_{T \times D}$ — распределения тем в документах

Порождающая модель наблюдаемых и скрытых i.i.d. данных:

$$p(X, Z|\Omega) = \prod_{i=1}^n p(d_i, w_i, t_i) = \\ = \prod_{i=1}^n p(w_i|t_i)p(t_i|d_i)p(d_i) = \prod_{i=1}^n \phi_{w_i t_i} \theta_{t_i d_i} p(d_i)$$

Задача ВТМ: применение общего EM-алгоритма

E-шаг: в силу независимости элементов выборки и формулы Байеса

$$q(Z) = p(Z|X,\Omega) = \prod_{i=1}^{n} p(t_i|d_i,w_i) = \prod_{i=1}^{n} \frac{p(w_i,t_i|d_i)}{p(w_i|d_i)} = \prod_{i=1}^{n} \underset{t_i\in T}{\mathrm{norm}}\bigl(\phi_{w_i t_i}\theta_{t_i d_i}\bigr)$$

M-шаг: подставим $q(Z)$ и $p(X,Z|\Omega)$ в общую формулу M-шага:

$$\sum_{Z\in T^n} q(Z)\ln p(X,Z|\Omega) + R(\Omega) =$$

$$\sum_{t_1\in T}\cdots\sum_{t_n\in T}\prod_{k=1}^{n} p(t_k|d_k,w_k)\sum_{i=1}^{n}\ln p(d_i,w_i,t_i|\Omega) + R(\Omega) =$$

$$\sum_{i=1}^{n}\sum_{t_1\in T}\cdots\sum_{t_n\in T}\prod_{k=1}^{n} p(t_k|d_k,w_k)\ln p(d_i,w_i,t_i|\Omega) + R(\Omega) =$$

$$\sum_{i=1}^{n}\sum_{t\in T} p(t|d_i,w_i)\ln p(d_i,w_i,t|\Omega) + R(\Omega) \;\to\; \max_{\Omega}$$

$$\sum_{d\in D}\sum_{w\in W}\sum_{t\in T} n_{dw}\,p(t|d,w)\ln\bigl(\phi_{wt}\theta_{td}\bigr) + R(\Phi,\Theta) \;\to\; \max_{\Phi,\Theta}$$

Регуляризованный EM-алгоритм для задачи BTM

Переменные $X=(d_i, w_i)_{i=1}^n$ наблюдаемые, $Z=(t_i)_{i=1}^n$ латентные

Лемма. Точка $\Omega = (\Phi, \Theta)$ локального максимума RML (регуляризованного маргинализованного log-правдоподобия)

$$\ln \sum_Z p(X, Z|\Omega) + R(\Omega) = \sum_{d,w} n_{dw} \ln \sum_t \phi_{wt}\theta_{td} + R(\Phi, \Theta)$$

удовлетворяет системе уравнений, решение которой методом простых итераций сводится к чередованию двух шагов:

$$\text{E-шаг:}\quad p(t|d,w) = \operatorname*{norm}_{t\in T}\bigl(\phi_{wt}\theta_{td}\bigr),\quad \forall(d\in D, w\in d, t\in T)$$

$$\text{M-шаг:}\quad \sum_{d,w,t} n_{dw}\, p(t|d,w) \ln\bigl(\phi_{wt}\theta_{td}\bigr) + R(\Phi,\Theta) \to \max_{\Phi,\Theta}$$

EM-алгоритм: вывод формул M-шага из условий ККТ

Задача M-шага декомпозируется на независимые подзадачи

$$\sum_{w,t} \ln \phi_{wt} \underbrace{\sum_d n_{dw} p_{tdw}}_{n_{wt}} + \sum_{d,t} \ln \theta_{td} \underbrace{\sum_w n_{dw} p_{tdw}}_{n_{td}} + R(\Phi, \Theta) \to \max_{\Phi,\Theta}$$

при ограничениях $\phi_{wt} \geqslant 0$, $\sum_w \phi_{wt} = 1$, $\theta_{td} \geqslant 0$, $\sum_t \theta_{td} = 1$

EM-алгоритм: метод простой итерации для системы уравнений

$$\begin{array}{ll} \text{E-шаг:} & \left\{ \begin{array}{l} p_{tdw} \equiv p(t|d,w) = \mathop{\mathrm{norm}}\limits_{t\in T} \big(\phi_{wt}\theta_{td}\big) \\[2ex] \phi_{wt} = \mathop{\mathrm{norm}}\limits_{w\in W} \Big(n_{wt} + \phi_{wt} \frac{\partial R}{\partial \phi_{wt}} \Big), \quad n_{wt} = \sum\limits_{d\in D} n_{dw} p_{tdw} \\[2ex] \theta_{td} = \mathop{\mathrm{norm}}\limits_{t\in T} \Big(n_{td} + \theta_{td} \frac{\partial R}{\partial \theta_{td}} \Big), \quad n_{td} = \sum\limits_{w\in d} n_{dw} p_{tdw} \end{array} \right. \\ \text{M-шаг:} & \end{array}$$

Воронцов К. В. Аддитивная регуляризация тематических моделей коллекций текстовых документов. Доклады РАН, 2014.

Модель вероятностного латентного семантического анализа

PLSA — Probabilistic Latent Semantic Analysis [Хофманн, 1999]:

- $R(\Phi, \Theta) = 0$ — нет никакой регуляризации.

ЕМ-алгоритм: метод простой итерации для системы уравнений

$$\begin{array}{rl} \text{Е-шаг:} & \left\{ \begin{array}{l} p_{tdw} = \underset{t \in T}{\mathrm{norm}} \big(\phi_{wt} \theta_{td} \big) \\ \phi_{wt} = \underset{w \in W}{\mathrm{norm}} \Big(\sum\limits_{d \in D} n_{dw} p_{tdw} \Big) \\ \theta_{td} = \underset{t \in T}{\mathrm{norm}} \Big(\sum\limits_{w \in d} n_{dw} p_{tdw} \Big) \end{array} \right. \\ \text{М-шаг:} & \end{array}$$

Е-шаг — формула Байеса
М-шаг — частотные оценки условных вероятностей

Hofmann T. Probabilistic latent semantic indexing. SIGIR 1999.

Модель латентного размещения Дирихле

LDA — Latent Dirichlet Allocation [Блэй, Ын, Джордан, 2001]:

- $R(\Phi, \Theta) = \sum\limits_{t\in T} \sum\limits_{w\in W} \beta_w \ln \phi_{wt} + \sum\limits_{d\in D} \sum\limits_{t\in T} \alpha_t \ln \theta_{td}$

  распределения ϕ_t близки к заданному распределению β
  распределения θ_d близки к заданному распределению α

ЕМ-алгоритм: метод простой итерации для системы уравнений

$$
\begin{array}{rl}
\text{Е-шаг:} & \left\{ \begin{array}{l}
p_{tdw} = \operatorname*{norm}\limits_{t\in T} \bigl(\phi_{wt}\theta_{td}\bigr) \\[2ex]
\phi_{wt} = \operatorname*{norm}\limits_{w\in W} \Bigl(\sum\limits_{d\in D} n_{dw} p_{tdw} + \beta_w \Bigr) \\[2ex]
\theta_{td} = \operatorname*{norm}\limits_{t\in T} \Bigl(\sum\limits_{w\in d} n_{dw} p_{tdw} + \alpha_t \Bigr)
\end{array} \right. \\
\text{М-шаг:} &
\end{array}
$$

Blei D., Ng A., Jordan M. Latent Dirichlet Allocation. NIPS-2001. JMLR 2003.

Резюме

- Широкий класс задач связан с выявлением латентных структур, порождающих наблюдаемые данные:
  — разделение смеси вероятностных распределений
  — вероятностное тематическое моделирование
  — «мягкая» и «жёсткая» кластеризация
  — сегментация временных рядов, сигналов, изображений
  — восстановление пропущенных данных
  — выявление аномальных наблюдений в данных
- Имеется универсальный рецепт вывода вычислительных формул E и M шагов по порождающей модели $p(X, Z|\Omega)$
- Можно добавлять какие угодно регуляризаторы $R(\Omega)$, причём не обязательно вероятностные вида $\ln p(\Omega|\gamma)$
- Гарантируется сходимость в слабом смысле
- Это обучение без учителя, не требующее разметки