$(10^{-3}\ 10^{-7}\ 10^{-6})\ (10^{-3}\ 10^{-7}\ 10^{-6}) \rightsquigarrow (10^{-6}, 10^{-14}\ 10^{-12})$

$p(c_1)$ Спорт

$p(c_2)$ Финансы

$p(c_3)$ Политика

Кр. Рон. сломал ногу в душе на карантине

$\bar{x} = (w_1, w_2, \dots, w_p)$

$w_1\ w_2 \dots$

generative model

$p(y \mid \bar{x})$

$y = (1, 0, 0)$

Naive Bayes: $p(\bar{x} \mid y) = \prod_i p(w_i \mid y)$

$$p(y \mid \bar{x}) = \frac{p(\bar{x}, y)}{p(\bar{x})} \propto p(y) \cdot \prod_i p(w_i \mid y)$$

Multinomial naive Bayes

$$p(w_i \mid y) = \frac{\#\{w_i \in y\} + 1}{\sum_k \#\{w_k \in y\} + |V|}$$

Multivariate NB

$[\dots\dots 1 \dots 0 \dots 1]$

$$p(\bar{x} \mid y) = \prod_{v \in V} p(x_v \mid y) = \prod_{w \in \bar{x}} p(w \mid y) \cdot \prod_{w \notin \bar{x}} (1 - p(w \mid y))$$

$$p(y \mid \bar{x}) \propto p(y) \cdot \prod_{w \in \bar{x}} p(w \mid y) \prod_{w \notin \bar{x}} (1 - p(w \mid y))$$

$$p(w \mid y) = \frac{\#\{d \in y \mid w \in d\} + 1}{\#\{d \in y\} + 2}$$

1) Bag of words
2) Naive assumption
3) Labeled dataset
4) $\forall d \in C_k$ полностью

EM-clustering on NB

$z_{jk} = [d_j \in C_k]$

$D = \{d_j = (w_1 \dots w_{N_j})\}_{j=1}^{M}$

$\pi_k = p(C_k)$

$\varphi_{kw} = p(w \mid C_k)$

$\sum_w \varphi_{kw} = 1$

$$p(D, Z \mid \pi, \Phi) = \prod_{j=1}^{M} \prod_{k=1}^{K} \Big(\pi_k \cdot \prod_{n=1}^{N_j} \varphi_{k, w_{jn}}\Big)^{z_{jk}}$$

$p(D \mid \pi, \Phi) = \sum_Z p(D, Z \mid \pi, \Phi)$

E-шаг: $\mathbb{E}[z_{jk}] = p(d_j \in C_k) = \dfrac{\pi_k \prod_n \varphi_{k, w_{jn}}}{\sum_\ell \pi_\ell \prod_n \varphi_{\ell, w_{jn}}}$

$p(D, Z, \pi, \Phi \mid \alpha, \beta)$

$p(\pi \mid \alpha)$

$p(\Phi \mid \beta)$

M-шаг: $\pi^{(t+1)}, \Phi^{(t+1)} = \arg\max_{\pi, \Phi} \sum_{j=1}^{M} \sum_{k=1}^{K} (\mathbb{E} z_{jk}) \cdot \Big(\log \pi_k + \sum_n \log \varphi_{k, w_{jn}}\Big)$

$$\pi_k^{(t+1)} = \frac{\sum_j \mathbb{E} z_{jk}}{M}; \quad \varphi_{kw}^{(t+1)} = \frac{\sum_j n_{jw} \cdot (\mathbb{E} z_{jk})}{\sum_j N_j (\mathbb{E} z_{jk})}$$

$p(w \mid C_k)$

$\sum_j \sum_k \mathbb{E} z_{jk} \log \pi_k \to \max$

$\sum \pi_k = 1$

$L = (\ ) + \lambda(\sum \pi_k - 1)$

$\partial / \partial \pi_k = \sum_j \mathbb{E} z_{jk} / \pi_k + \lambda = 0$

Topic Modeling

Generative models

$w \quad 1, \dots, V$

$j \quad 1, \dots, M$

$t \quad 1, \dots, T$

$$p(d \mid \lambda) = \prod_{w \in d} \sum_{t=1}^{T} p(t \mid d) \cdot p(w \mid t)$$

$\prod_{w \in d} p(w \mid d)$

$p(w, t \mid d)$

$j \rightsquigarrow$ $1 \dots T$ $\theta_{td} = p(t \mid d)$

$\forall w \rightsquigarrow t, w$

$\varphi_{wt} = p(w \mid t)$

$p(\Phi,\Theta|D) \propto p(D|\Phi,\Theta)\,p(\Phi)p(\Theta)$

$$p(D|\Phi,\Theta) = \prod_{d\in D}\prod_{w\in d}\left(\sum_{t=1}^{T}\theta_{td}\cdot\varphi_{wt}\right) \xrightarrow[\Theta,\Phi]{} \max$$

$z_{tdw} = [w,d \in t]$

$$p(D,Z|\Phi,\Theta) = \prod_{d\in D}\prod_{w\in d}\prod_{t=1}^{T}(\theta_{td}\varphi_{wt})^{z_{tdw}} \to \max$$

(Hoffmann, 2000)

pLSI

LSA

$$\log p = \sum_d\sum_w\sum_t\left(z_{tdw}\log\theta_{td} + z_{tdw}\log\varphi_{wt}\right) \to \max$$

M-шаг

$$\theta^*_{td} = \frac{\sum_w \hat z_{tdw}}{\sum_w\sum_t \hat z_{tdw}} \qquad \varphi^*_{wt} = \frac{\sum_d \hat z_{tdw}}{\sum_d\sum_w \hat z_{tdw}}$$

$N\times V \sim (N\times T)\cdot(T\times V)$

Θ	$N\times T$
Φ	$T\times V$
Z	$T\times N\times V$

E-шаг:

$$E z_{tdw} = p(t|w,d) = \frac{\theta_{td}\cdot\varphi_{wt}}{\sum_s\theta_{sd}\varphi_{ws}}$$

$= \hat z_{tdw}$

ARTM - additive reg. of topic model

Константин Воронцов

$$\log p(D,Z|\Phi,\Theta) = \sum_{d,w,t} z(\ldots) + R(\Phi,\Theta)$$

$$L(\Phi,\Theta) = \sum_{d,w,t}\left(z_{tdw}\log\theta_{td} + z_{tdw}\log\varphi_{wt}\right) + R \xrightarrow[\theta,\varphi]{} \max$$

$$L|_{\theta_{*d}} = \sum_{w\in d}\sum_t z_{tdw}\log\theta_{td} - R \xrightarrow[\theta_{td}]{} \max \qquad \theta_{td}\geq 0,\ \sum\theta_{td} = 1$$

$$\sum_w\sum_t z_{tdw}\log\theta_{td} - R + \lambda\left(\sum_t\theta_{td} - 1\right) \to \max \qquad \theta_{td}\geq 0$$

$$\partial/\partial\theta_{td} = \frac{\sum_w z_{tdw}}{\theta_{td}} - \frac{\partial R}{\partial\theta_{td}} - \lambda = 0$$

$$\theta_{td} = \frac{\sum_w z_{tdw}}{\lambda + \partial R/\partial\theta_{td}}$$

$\sum z_k\log\pi_k \to \max$

$-\lambda(\sum\pi_k - 1) \qquad \sum\pi_k = 1$

$\frac{\partial}{\partial\pi_k} = \frac{z_k}{\pi_k} - \lambda = 0$

$1 = \frac{\sum z_k}{\lambda}$

$\pi_k = \frac{z_k}{\lambda} \qquad \lambda = \sum z_k$

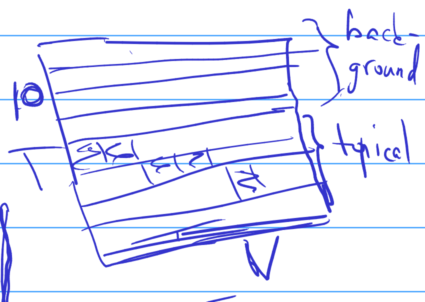
$$\theta_{td} \propto \max\left(\sum_w z_{td} - \theta_{td}\frac{\partial R}{\partial\theta_{td}},\ 0\right)$$

$$\varphi_{wt} \propto \max\left(\sum_d z_{tdw} - \varphi_{wt}\frac{\partial R}{\partial\varphi_{wt}},\ 0\right)$$

Sparsity:

$$R(\Phi) = \sum_{t\neq t'}\sum_w \varphi_{wt}\cdot\varphi_{wt'}$$

$$R(\varphi_{*t}) = -\sum_w \varphi_{wt}\log\varphi_{wt} \to \min$$



Smoothness:

$$R(\varphi) = \sum_w \varphi_{wt}\log\varphi_{wt}$$

Coherence: $R(\Phi) = -\sum\varphi\cdot\log\varphi'$