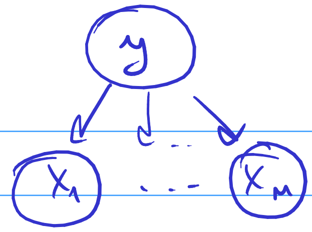
Naive Bayes classifier (Idiot's Bayes)

порождающая / generative

$\bar{x} = (x_1, x_2, \dots, x_M) \longrightarrow y \in \{C_1, \dots, C_K\}$

$p(y|\bar{x}) \propto p(\bar{x}, y)$



$$p(\bar{x}) \qquad p(\bar{x}, y) = p(y)\,p(\bar{x}|y) = p(y)\prod_{i=1}^{M} p(x_i|y)$$

лог. регр.

$$p(y = C_k|\bar{x}) \propto e^{\bar{w}_k^T \bar{x}} = e^{w_{k1}x_1} e^{w_{k2}x_2} \dots e^{w_{kM}x_M}$$

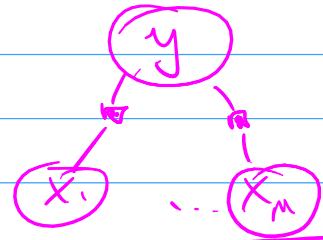
$$= \frac{e^{\bar{w}_k^T \bar{x}}}{\sum_\ell e^{\bar{w}_\ell^T \bar{x}}}$$

$\cancel{p(\bar{x})} \qquad \cancel{p(\bar{x}, y)}$

дискриминативная

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

generative-discriminative pair



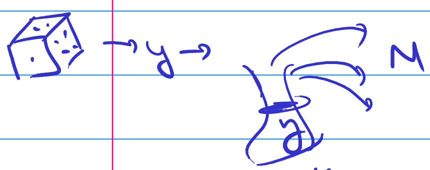
$$p(\bar{x}, y) = p(y)\prod_{i=1}^{M} p(x_i|y)$$

x_i - tokens

bag-of-words / n-grams / tokens

$y \in \{$спорт, фин., полит.$\}$

Multinomial NB

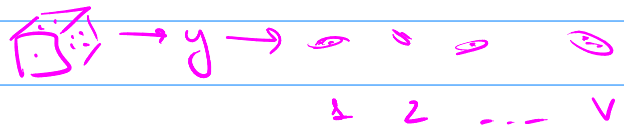


$$p(y, d) = p(y) \cdot \prod_{j=1}^{N_d} p(w_j|y)$$

$$p(y = C_k) = \pi_k$$

$$p(w_j|y = C_k) = \varphi_{w_j,k}$$

Multivariate NB



$$p(y, d) = p(y) \cdot \prod_{i=1}^{V} p(w_i \in d\,|\,y)$$

$$p(w|y = C_k) = \frac{1 + \sum_d [y = C_k][w \in d]}{2 + \sum_d [y = C_k]}$$

$$\pi_k = \frac{\sum_d [y = C_k]}{D}; \qquad \varphi_{wk} = \frac{1 + \sum_d \sum_{j=1}^{N_d} [y = C_k][w_j = w]}{|V| + \sum_d \sum_{j=1}^{N_d} [y = C_k]}$$

Кластеризация через NB $\qquad D = \{d_1, \dots, d_N\} \qquad t$ - topic

EM-алгоритм:

$$Z = \{\bar z_1, \dots, \bar z_N\}$$

$$\bar z_d = (\dots z_{td} \dots)$$

$$z_{td} = [d \in t]$$

$$p(D, Z \mid \pi_t, \varphi_{wt}) = \prod_{d \in D} \prod_{t=1}^{T} \Big(\pi_t \cdot \prod_{w \in d} \varphi_{wt}\Big)^{z_{td}}$$

$$p(D \mid \underbrace{\pi, \Phi}_{=\lambda}) \to \max$$

$$Q(\lambda, \lambda^{(m)}) = \mathbb{E}_{Z \mid \lambda^{(m)}} \left[\ln p(D, Z \mid \lambda)\right] \xrightarrow[\lambda]{} \max$$

$$Q(\lambda, \lambda^{(m)}) = \mathbb{E}_Z \Big[\sum_d \sum_t z_{td} \Big(\ln \pi_t + \sum_w \ln \varphi_{wt}\Big)\Big] =$$

$$= \sum_d \sum_t \mathbb{E}_{Z \mid \lambda^{(m)}}[z_{td}] \cdot \Big(\ln \pi_t + \sum_{j=1}^{N_d} \ln \varphi_{w_j, t}\Big) \xrightarrow[\pi, \Phi]{} \max$$

$$\Phi = \begin{array}{c} \\ \left(\; \varphi_{wt} \;\right) w \\ t \end{array}$$

$$\mathbb{E}_{Z \mid \lambda^{(m)}}[z_{td}] = p(t \mid d, \lambda^{(m)}) = \frac{p(d, t \mid \lambda^{(m)})}{p(d \mid \lambda^{(m)})} = \frac{\pi_t^{(m)} \cdot \prod_{w \in d} \varphi_{wt}^{(m)}}{\sum_s \pi_s^{(m)} \prod_{w \in d} \varphi_{ws}^{(m)}}$$

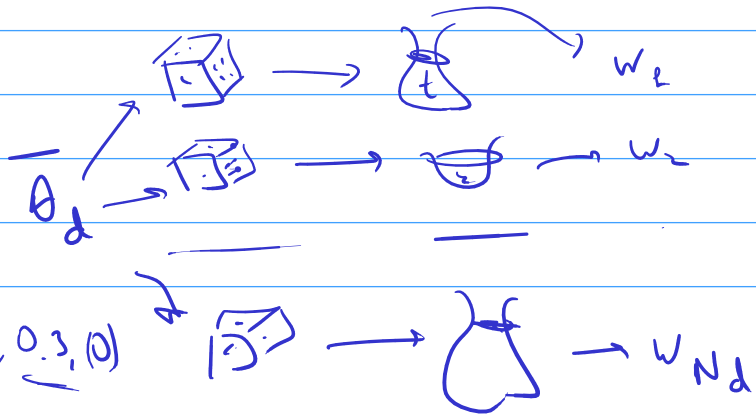
$$Q(\lambda, \lambda^{(m)}) = \sum_t \pi_t \cdot \Big(\sum_d \mathbb{E}[z_{td}]\Big) +$$

$$+ \sum_t \sum_{w \in V} \ln \varphi_{wt} \cdot \Big(\sum_d \sum_{j : w_j = w} \mathbb{E}[z_{td}]\Big)$$

$$\pi_t = \frac{\sum_d \mathbb{E}[z_{td}]}{D}, \qquad \varphi_{wt} = \frac{\sum_d \#\{w \in d\} \cdot \mathbb{E}[z_{td}]}{\sum_d N_d \cdot \mathbb{E}[z_{td}]}$$

Topic modeling

t	$1 \dots T$
w	$1 \dots V$
d	$1 \dots N$

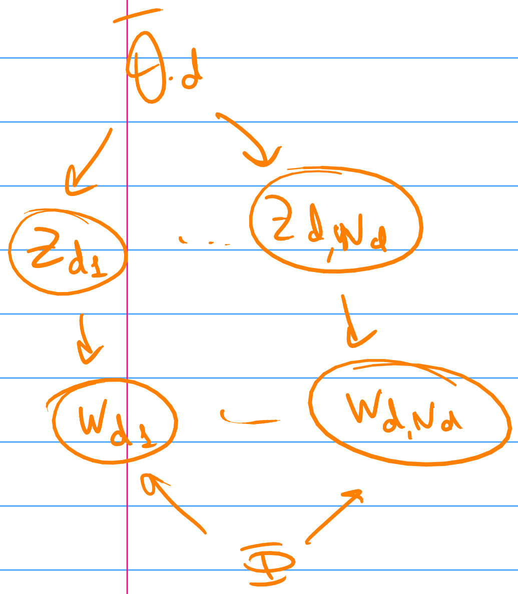


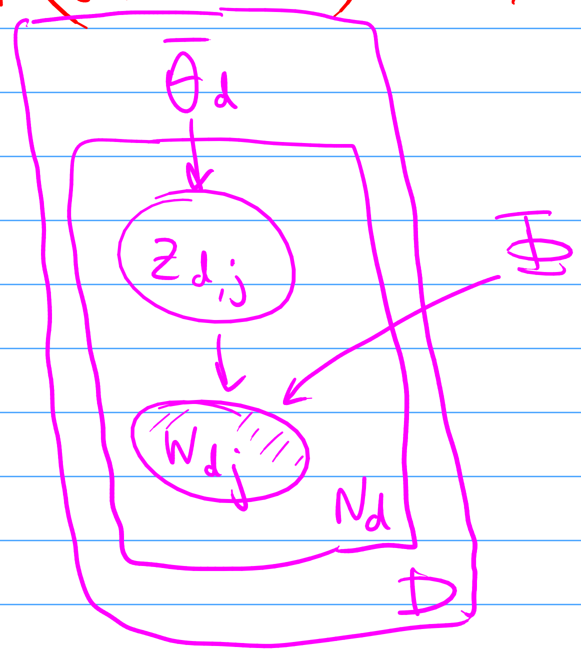
$$\underset{V\times T}{\Phi} = \left(\varphi_{wt}\right) \qquad \underset{T\times N}{\Theta} = \left(\theta_{td}\right)$$

$$p(w|d) = \sum_{t=1}^{T} p(w,t|d) = \sum_{t=1}^{T} p(t|d)\,p(w|t)$$

$$p(w|d) = \sum_{t=1}^{T} \theta_{td}\varphi_{wt}$$

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





$$\bar z_{dj} = (\ldots z_{djt} \ldots)$$

$$z_{djt} = [w_{dj} \in t]$$

$$p(W,Z|\Theta,\Phi) = \prod_{d\in D}\prod_{j=1}^{N_d}\prod_{t=1}^{T}\left(\theta_{td}\varphi_{wt}\right)^{z_{djt}} \xrightarrow[\Theta,\Phi]{} \max$$

$$Q(\lambda,\lambda^{(m)}) = \mathbb{E}_{Z|\lambda^{(m)}}\left[\ln p(W,Z|\lambda)\right] = \sum_{d}\sum_{j=1}^{N_d}\sum_{t=1}^{T}\mathbb{E}[z_{djt}]\left(\ln\theta_{td} + \ln\varphi_{w_{dj},t}\right)$$

E-шаг: $\mathbb{E}[z_{djt}] = p(t|w_j,d) = \dfrac{p(d,t,w_j)}{p(d,w_j)} = \dfrac{\theta^{(m)}_{td}\varphi^{(m)}_{wt}}{\sum_s \theta^{(m)}_{sd}\varphi^{(m)}_{ws}}$

$$Q(\lambda, \lambda^{(m)}) = \sum_d \sum_t \left[\ln \theta_{td} \cdot \overbrace{\left(\sum_{j=1}^{N_d} \mathbb{E}[z_{djt}] \right)}^{\sum_w n_{dwt}} + \sum_{w=1}^{V} \sum_{t=1}^{T} \ln \varphi_{wt} \overbrace{\left(\sum_d \sum_{j=1}^{N_d} [w_{dj} = w] \cdot \mathbb{E}[z_{djt}] \right)}^{n_{dwt}} \right]$$

$$n_{dwt} = \mathbb{E}\left[\#\{ w \in t \text{ в } d \} \right] = \sum_{j=1}^{N_d} \left([w_{dj} = w]\, \mathbb{E}[z_{djt}] \right)$$

$$\theta_{td} = \frac{\sum_{w=1}^{V} n_{dwt}}{N_d} \qquad \varphi_{wt} = \frac{\sum_d n_{dwt}}{\sum_d \sum_w n_{dwt}}$$

LSA — latent semantic analysis

pLSI — prob. lat sem. indexing

$$V \underbrace{\begin{pmatrix} & | & \\ - & \varphi_{wt} & - \\ & | & \end{pmatrix}}_{T} \cdot \underbrace{\begin{pmatrix} & | & \\ - & \theta_{td} & - \\ & | & \end{pmatrix}^{T}}_{D} \approx \overset{W}{\underbrace{\begin{pmatrix} & & \\ & \circ & \\ & & \end{pmatrix}}_{D}} V$$

$$\boxed{\sum_t \varphi_{wt} \theta_{td}}$$

ARTM — К. Воронцов

$$p(D \mid \Theta, \Phi) \cdot p(\Theta, \Phi)$$

$$\ln p(D \mid \Theta, \Phi) + \overbrace{\sum_i \lambda_i R_i(\Theta, \Phi)}^{R} \to \max$$

$$\sum_{t,s} KL(\bar{\varphi}_t \| \bar{\varphi}_s)$$

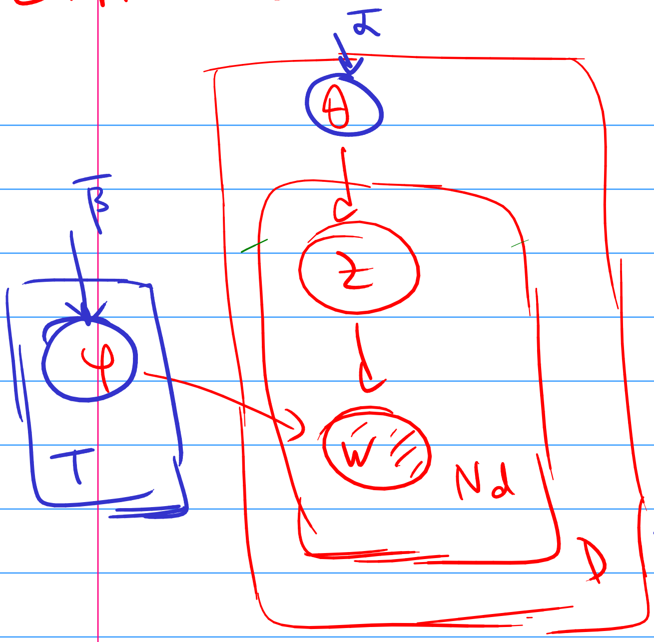
$$n_{*wt} = \left[\sum_d n_{dwt} + \varphi_{wt} \frac{\partial R}{\partial \varphi_{wt}} \right]_+$$

$$KL(\bar{\varphi}_t \| Unif) \to \max$$

$$n_{d*t} = \left[\sum_w n_{dwt} + \theta_{td} \frac{\partial R}{\partial \theta_{td}} \right]_+$$

$$KL(\bar{\varphi}_t \| Unif) \to \min$$

LDA Latent Dirichlet Allocation



$$p(W, Z, \Theta, \Phi) =$$

$$= p(\Phi) \prod_{j=1}^{D} p(\theta_d) \cdot \prod_{n=1}^{N_d} p(z_{d_j} \mid \theta_d)\, p(w_{d_j} \mid z_{d_j})$$

$$= \left[\prod_{t=1}^{T} p(\bar{\varphi}_t)\right] \cdot \left[\prod_{j=1}^{D} p(\bar{\theta}_j)\right] \cdot \prod_{j=1}^{D} \prod_{n=1}^{N_d} \left[p(z_{jn} \mid \bar{\theta}_j)\, p(w_{jn} \mid z_{jn}, \Phi)\right]$$

$$p(\bar{\theta}_j \mid \bar{\alpha}) = \frac{1}{B(\bar{\alpha})} \prod_{t=1}^{T} \theta_{jt}^{\alpha_t - 1}$$

$$p(\bar{\varphi}_t \mid \bar{\beta}) = \frac{1}{B(\bar{\beta})} \prod_{v=1}^{V} \varphi_{tv}^{\beta_v - 1}$$

$$p(\Theta, \Phi \mid W, \alpha, \beta) \xrightarrow[\Theta, \Phi]{} \max$$

$$\propto p(\Theta, \Phi, W \mid \alpha, \beta) = \sum_{Z} p(\Theta, \Phi, W, Z \mid \alpha, \beta)$$

$$\log p(X) = \log p(X, Z) - \log p(Z \mid X)$$

$$\log p(X) = E_{q(z)}\left[\log p(X, Z) - \log p(Z \mid X)\right] \pm E_q\left[\log q(Z)\right]$$

$$\log p(X) = \underbrace{E_q\left[\log \frac{p(X, Z)}{q(Z)}\right]}_{\mathcal{L}(q)} + \underbrace{E_q\left[\log \frac{q(Z)}{p(Z \mid X)}\right]}_{KL(q \,\|\, p(Z \mid X))}$$

q max

fix Φ - параметры

$$p(W, Z, \Theta) = \prod_{j=1}^{D} p(\bar\theta_j)\left(\cdot \prod_{j=1}^{D}\prod_{n=1}^{N_d}\prod_{t=1}^{T}[z_{jn}=t]\cdot\theta_{jt}\varphi_{w_{jn}t}\right)$$

$$q(Z, \Theta \mid \ldots) = \prod_{j=1}^{D}\left(q(\theta_j)\prod_{n=1}^{N_d} q(z_{jn})\right)$$

$$L(q) = \mathbb{E}_q\left[\frac{\log p(W, Z, \Theta)}{q(Z,\Theta)}\right] = \int \log p(\_)\, q\, dZ\, d\Theta - $$
$$- \int \log q(\_)\, q(\_)\, dZ\, d\Theta$$

$$L(q_j) = \iint \left[\log p(\bar w_j, \bar z_j, \bar\theta_j \mid \Phi, \alpha, \beta) - \log q_j(\bar z_j, \bar\theta_j)\right] q_j(\bar z_j, \bar\theta_j)\, dz_j\, d\theta_j$$

$$L(q_j) = \mathbb{E}_q\left[\log p(\bar\theta_j) + \log p(\bar z_j \mid \bar\theta_j) + \log p(\bar w_j \mid \bar z_j) - \log q_j(\bar z_j) - \log q_j(\bar\theta_j)\right]$$

$$\mathbb{E}_q\left[\log p(\bar\theta_j \mid \alpha)\right] = \mathbb{E}_q\left[\log \mathrm{Dir}(\bar\theta_j \mid \bar\alpha)\right] =$$

$$q_j(\bar\theta_j) = \mathrm{Dir}(\bar\theta_j \mid \bar\gamma_j)$$

$$= \mathbb{E}_{\mathrm{Dir}(\theta_j \mid \gamma_j)}\left[\log \prod \theta_{jt}^{\alpha_t - 1} + \mathrm{const}\right]$$

$$= \log\Gamma\left(\sum \alpha_t\right) - \sum \log\Gamma(\alpha_t)$$
$$+ \sum_t (\alpha_t - 1)\left(\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right)\right)$$

$$q_j(\bar z_j) = \prod_n q_j(z_{jn}) =$$
$$= \prod_n \mathrm{Mult}(z_{jn} \mid \bar\pi_j)$$
$$= \prod_n \prod_j \pi_{jt}^{[z_{jn}=t]}$$

digamma $\psi(x) = \dfrac{\partial \log\Gamma(x)}{\partial x}$

$$\mathbb{E}_q\left[\log p(\bar z_j \mid \bar\theta_j)\right] = \mathbb{E}_q\left[\sum_n \sum_t [z_{jn}=t] \log\theta_{jt}\right] =$$

$$= \sum_n \sum_t \underline{\mathbb{E}_q\left[[z_{jn}=t]\right]} \cdot \mathbb{E}_q\left[\log\theta_{jt}\right] =$$

$$= \sum_n \sum_t \pi_{jt} \left(\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right) \right)$$

$$L(\bar\gamma_j, \bar\pi_j) = (1)+(2)+(3)-(4)-(5) \xrightarrow[\bar\gamma,\, \bar\pi]{} \max$$

$$\pi_{jt}(\ldots) \qquad \gamma_{jt} \ldots$$

LDA : EM по Φ, Z

- E-шаг $KL(q_j \| p_j) \xrightarrow[\gamma_j,\, \pi_j]{} \min$ для всех j, фикс. Φ

- M-шаг: $\varphi_{ts} \propto \ldots$

  $z_{jnt} \propto \ldots$

$$q(Z, \Theta, \Phi) = q(Z, \Theta) \cdot \prod_t q_t(\bar\varphi_t)$$

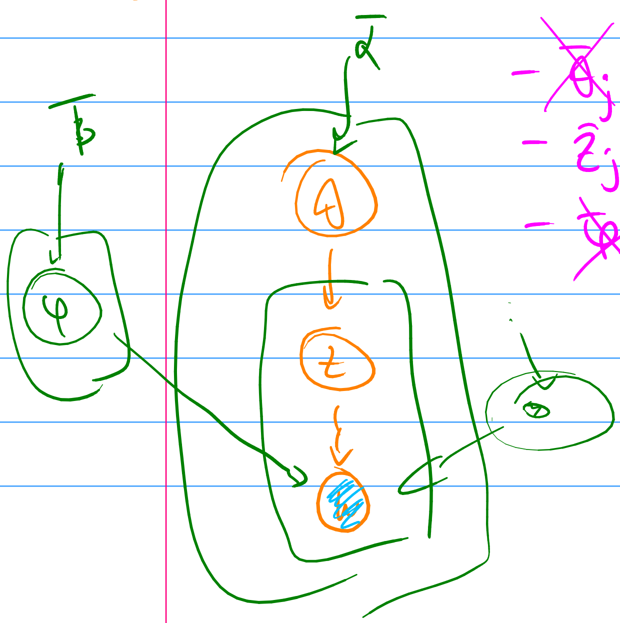
Gibbs sampling $\Phi, \Theta \sim p(\Phi, \Theta \mid W, \alpha, \beta)$

- $\cancel{\theta_j}$
- $\bar z_{jn}$
- $\cancel{\varphi_t}$

Collapsed Gibbs sampling

$$\bar z_{jn} \sim p(\bar z_{jn} \mid Z_{-jn}, W, \alpha, \beta) =$$

$$= \int p(\bar z_{jn}, \Phi, \Theta \mid Z_{-jn}, W, \alpha, \beta)\, d\Phi\, d\Theta$$



$$\theta_{td} = \frac{n_{td}}{n_d}, \quad \varphi_{st} = \frac{n_{st}}{n_t}$$