Language model

мама мыла ?

$p(x_i \mid x_1 \ldots x_{i-1}) \approx ?$
$\approx NN(x_{i-1}, \bar{h})$

$$p(x_1, x_2, \ldots, x_N) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_1, x_2) \ldots p(x_N \mid x_1, \ldots, x_{N-1})$$

$$x_1 \to x_2 \to \ldots \to x_{k-1} \to x_k \to \ldots \to x_N$$

$p(x_i \mid x_1 \ldots x_{i-1}) = p(x_i \mid x_{i-1})$ unigrams

$p(x_i \mid x_1 \ldots x_{i-1}) = \underline{p(x_i \mid x_{i-1}, \ldots, x_{i-n})}$ n-grams

$$x_1 \to x_2 \to x_3 \to x_4 \to x_5 \to \ldots \to x_N$$

$$y_1 \to y_2 \to \ldots \to y_N$$

$p(w_i, \underbrace{w_{i-1}, w_{i-2}} \mid \underbrace{w_{i-1}, w_{i-2}, w_{i-3}})$

$$p(w \mid w_1, \ldots, w_n) = \frac{\#\{w_1, \ldots, w_n, w\} + 1}{\#\{w_1, \ldots, w_n, \ldots\} + |V|}$$

мама мыла Раму утром рано

backoff

Kneser-Ney smoothing

мама_мыла мыла_Раму Раму_утром утром_рано

Евгений_Онегин Гуин_Понсеgy

$$\left(\frac{p(\text{Евг\_Онегин})}{\frac{\#[\text{Евг\_Он}]}{N-1}} \Big/ \frac{p(\text{Евг}) \cdot p(\text{Онегин})}{\frac{\#\text{Евг}}{N}} \right)$$

Bag of words

$\{w_1, \ldots, w_N\}$

$(\ldots \#w_i \ldots)$

$tf(w,d) \cdot idf(w,D)$

$$tf(w,d) = \frac{\#\{w \in d\}}{|d|} \qquad idf(w,D) = \log \frac{|D|}{|\{d \mid w \in d\}|}$$

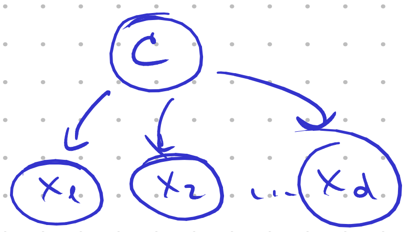
term freq.

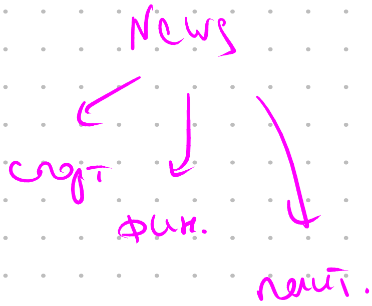
inverse doc. freq.

Naive Bayes / Idiot's Bayes

$D = \{(\bar{x}, \bar{y})\}$

$(x_1, \dots, x_d)$ "$(0 \dots 1 \dots 0)$"



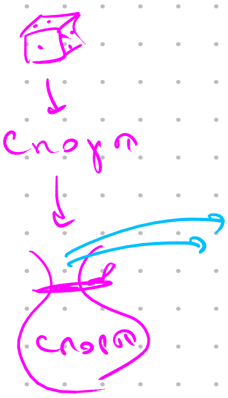


$$p((\bar{x}, c) \mid \theta) = p(c) \cdot \prod_{i=1}^{d} p(x_i \mid c)$$

θ

Generative model $\quad p((\bar{x}, y) \mid \theta) = p(y \mid \theta) \cdot \prod_i p(x_i \mid y, \theta)$

Discriminative model $\quad p(y = k \mid \bar{x}, \theta) \propto e^{x^T \theta_k}$ — log. regr.



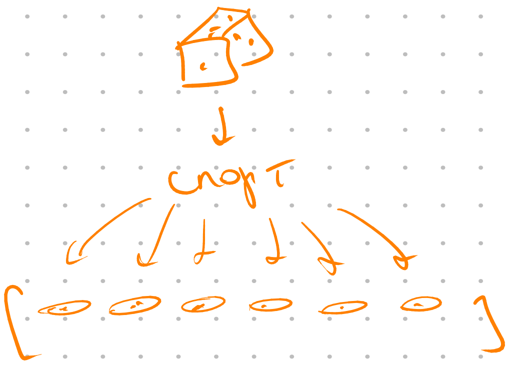
Multinomial NB

$$p(d, c) = p(c) \cdot \prod_{i=1}^{|d|} p(w_i \mid c)$$

$$\theta_k = p(c = k) \approx \frac{\#\{d \in C_k\} + 1}{|D| + |C|}$$

$$\theta_{mk} = p(w = m \mid c = k) \approx \frac{\#\{m \in d,\ d \in C_k\} + 1}{\#\{\text{слов в } C_k\} + |V|}$$

$$p(c = k \mid d) \propto p(d, c = k) = \theta_k \cdot \prod_{i=1}^{|d|} \theta_{w_i, k}$$



Multivariate NB

$d \sim [0, 1, \dots 0, 1, 1, \dots 0]$ $|V|$

$$p(d, c) = p(c) \cdot \prod_{j=1}^{|V|} p(w_j \in c)^{[w_j \in d]} \cdot (1 - p(w_j \in c))^{[w_j \notin d]}$$

$$\theta_k = p(c = k)$$

$$\theta_{jk} = p(w_j \in C_k) \approx \frac{\#\{d \in C_k \mid w_j \in d\} + 1}{\#\{d \in C_k\} + 2}$$

$\bar{x} = (x_1, x_2, \dots, x_d) \qquad y \in \{1, \dots, K\}$

$$p(D|\bar\theta) = \prod_{d\in D} p(d|\bar\theta) = \prod_{(\bar x,y)\in D} p(y) \cdot \prod_{i=1}^{d} p(x_i|y) \xrightarrow{\bar\theta} \max$$

$\theta_k = p(y=k) \qquad \theta_{kim} = p(x_i = m \mid y=k)$

$$p(D|\bar\theta) = \prod_{(\bar x,y)\in D} \left(\prod_{k=1}^{K} \theta_k^{[y=k]}\right)\left(\prod_{i=1}^{d}\prod_{k=1}^{K}\prod_{m=1}^{M} \theta_{kim}^{[x_i=m,\, y=k]}\right)$$

$$\ln p(D|\bar\theta) = \sum_{(\bar x,y)} \sum_{k=1}^{K} \left([y=k] \underbrace{\ln\theta_k}_{w_k} + \sum_{i=1}^{d}\sum_{m=1}^{M} [x_i=m,\, y=k]\underbrace{\ln\theta_{kim}}_{w_{kim}}\right)$$

$$\ln p(D|\theta) = \left(\sum_{(\bar x,y)} \bar f(\bar x, y)\right)^T \bar w$$

NB и LR образуют generative - discriminative pair

LR:

$$p(y=k|\bar x) = \frac{e^{\bar g(\bar x)^T \bar w_k}}{\underbrace{\sum_{l} e^{\bar g(\bar x)^T \bar w_l}}_{= Z(\bar x)}}$$

$$\ln p(y=k|\bar x, \bar w) = \bar g(\bar x)^T \bar w_k \boxed{- \ln Z(\bar x)}$$

$$p(\bar x, y) = p(y|\bar x)\, p(\bar x) = \frac{e^{\bar g(\bar x)^T \bar w_k}}{Z'(\bar x)}$$

Naive Bayes:
- bag of words
- supervised learning
- $\forall$doc has only one topic

$D = \{d\} \qquad t = 1, \dots, T$

$$p(D|\theta) = \prod_{d\in D} p(d|\theta) = \prod_{d\in D} \sum_{t=1}^{T} p(d,t|\theta) = \prod_{d\in D} \sum_{t=1}^{T} \overbrace{p(t)}^{\pi_t = p(d\in t)} \cdot \prod_{w\in d} \overbrace{p(w|t)}^{\varphi_{wt} = p(w|t)}$$

$$= \prod_{d\in D} \sum_{t=1}^{T} \pi_t \prod_{w\in d} \varphi_{wt} \xrightarrow[\bar\pi,\ \bar\varphi]{} \max$$

$$\bar{z}_d = (\ldots z_{dt} \ldots)_{t=1}^{T}$$

$$p(D, Z \mid \Theta) = \prod_d \prod_t \left(\pi_t \prod_{w \in d} \varphi_{wt} \right)^{z_{dt}}$$

<u>EM-алгоритм:</u>

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

$$= \sum_d \sum_t \left(\ln \pi_t + \sum_{w \in d} \ln \varphi_{wt} \right) \cdot \mathbb{E}[z_{dt}]$$

E-шаг

$$\left[\begin{array}{l} p(t \mid d, \Theta^{(m)}) = \dfrac{p(t, d \mid \Theta^{(m)})}{p(d \mid \Theta^{(m)})} = \\[2ex] = \dfrac{\pi_t^{(m)} \cdot \prod_{w \in d} \varphi_{wt}^{(m)}}{\sum_{s=1}^{T} \pi_s^{(m)} \cdot \prod_{w \in d} \varphi_{ws}^{(m)}} \end{array} \right.$$