AHMED MAZARI

HOMEWORK 2

① PLSA

On pose $C = \{corpus\}$, ~~…~~ $T = \{thème\}$

$$P(C,T) = \prod_{d=1}^{n_D} P(D=d) \left[\prod_{d=1}^{\lambda_d} \alpha_{d t_0} \beta_{T_d w_0} \right]$$

sachant que $\alpha_{dt} = P(T=t \mid D=d)$

$\beta_{tw} = P(W=w \mid T=t$

$$P(C,T) = \prod_{d=1}^{n_d} P(D=d) \prod_{w=1}^{n_w} \left[\sum_{t=1}^{n_T} \alpha_{dt} \beta_{tw} \right)^{c_{dw}}$$

$$\log P(C) = \sum_{d=1}^{n_D} \left(\log P(D=d) + \sum_{w=1}^{n_w} c_{wd} \log \left(\sum_{t=1}^{n_T} \alpha_{dt} \beta_{tw} \right) \right)$$

On a $\log\left(\sum_{t=1}^{n_T} \alpha_{dt} \beta_{tw} \right) = \log \left(\sum_{t=1}^{n_T} \frac{P(t \mid d, \theta')(\alpha_{dt} \beta_{tw})}{P(t \mid d, \theta')} \right)$ ④

$$\log\left(\sum_{t=1}^{n_T} \alpha_{dt} \beta_{tw} \right) \geq \sum_{t=1}^{n_T} P(t \mid D) \log \frac{\alpha_{dt} \beta_{tw}}{P(t \mid d, \theta')}$$

$$\log P(C) \geq \sum_{d=1}^{n_D} \left(\log P(D=d) + \sum_{w=1}^{n_w} c_{w,d} \left(\sum_{t=1}^{n_T} P(t,D) \log \frac{(\alpha_{dt} \beta_{tw})}{P(t \mid d, \theta')} \right.\right.$$

$$\underset{\theta}{\operatorname{argmax}}\, P(C) = \underset{\theta}{\operatorname{argmax}} \sum_{d=1}^{n_d} \Big(\log p(D=d) + \sum_{w=1}^{n_w} c_{w,d} \Big(\sum_{t=1}^{n} p(t \mid D, \theta') \log \alpha_{dt} \beta_{tw} \Big) \Big]$$

sachant que $\sum_{t=1}^{n_T} p(t \mid D, \theta') \log \frac{1}{p(t \mid D, \theta')}$ est une constante

d'où l'équation auxiliaire est comme suit :

$$\underset{\theta}{\operatorname{argmax}}\, p(c) = \underset{\theta}{\operatorname{argmax}} \Big(\sum_{d=1}^{n} \Big(\log p(D=d) + \sum_{w=1}^{n_w} c_{wd} \Big(\sum p(t \mid d, \theta) \log \alpha_{dt} \beta(w) \Big) \Big) \Big)$$

- ajouter l'apriori pour α et β revient à faire du LDA

Comme vu en cours on a

$$P(W, Z, \theta, \varphi, \alpha, \beta) = \prod_{i=1}^{k} p(\varphi_i, \beta) \prod_{j=1}^{m} P(\theta_j, \alpha) \prod_{t=1}^{N} p(z_{j,t} \mid \theta_j)\, p(w_{j,t} \mid \varphi_{z_{j,t}})$$

II EM

On pose $X = x$ variables observées,

Y : variables non observées

a) $\ell(\theta) \geq F(q,\theta)\ \forall q$

$$\ell(\theta) = p(X=x|\theta) = \log \sum_y p(X=x, Y=y|\theta)$$

$$= \log \sum_y p(y) \frac{p(X=x, Y=y|\theta)}{p(y)}$$

d'après la concavité du log on a

$$\log \ell(\theta) \geq \sum_y p(y) \log p(X=x, Y=y|\theta) -$$

$$\sum_y p(y) \log p(y) = F(q,\theta)$$

b) d'après l'inégalité de Jensen

si $x_i = x_j\ \forall i,j$ si $\log(\sum \lambda_i x_i) \geq \sum_i \lambda_i \log x_i$

d'où $\ell(\theta) = \sum_y p(y) \log \frac{p(Y=x, Y=y|\theta)}{q(y)}$

si $\frac{p(X=x, Y=y|\theta)}{p(y)} \Rightarrow$ constante

$\Rightarrow q(y) \propto p(X=x, Y=y|\theta)$

$\Rightarrow q(y) = p(Y=y|x,\theta)$