清华大学实验报告

系别＿＿＿＿ 班号＿＿＿＿ 姓名 ZHANG NAIFU (同组姓名:＿＿＿＿＿＿)

作实验日期 年 月 日 2018280351 教师评定:

Q1.1

$$\cancel{p(x)}\ l(\mu) = \ln P(x|\mu)$$

$$= \ln\left(\frac{n!}{\prod_{i=1} x_i!}\prod_{i}^{d}\mu_i^{x_i}\right)$$

$$= \ln n! - \sum_{i=1}^{d}\ln x_i! + \sum_{i=1}^{d} x_i \ln \mu_i$$

Constrained optimisation => Lagrange method

$$\mathcal{L}(\mu, \lambda) = \ln n! - \sum_{i=1}^{d}\ln x_i! + \sum_{i=1}^{d} x_i \ln \mu_i$$

$$- \lambda_1\left(\sum_{i=1}^{d} x_i - n\right) - \lambda_2\left(\sum_{i=1}^{d}\mu_i - 1\right)$$

$$\frac{\partial}{\partial \mu_i}\mathcal{L}(\mu,\lambda) = \frac{x_i}{\mu_i} - \lambda_2 = 0$$

$$\mu_i = \frac{x_i}{\lambda_2}$$

$$\sum \mu_i = \sum \frac{x_i}{\lambda_2}$$

$$1 = \frac{n}{\lambda_2}$$

$$\lambda_2 = n$$

$$\mu_i = \frac{x_i}{\lambda_2}$$

$$\therefore \mu_i = \frac{x_i}{n}$$

$$\vec{\mu} = n^{-1}\vec{x}$$

//

Q 1.2 $\quad p(d) = \dfrac{n_d!}{\prod_w T_{dw}!} \sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}}$

$$\ell(\pi, \mu) = \sum \ln p(d)$$

$$= \sum \left[\ln \frac{n_d!}{\prod_w T_{dw}!} + \ln\left(\sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}} \right) \right]$$

$$\mathcal{L}(\pi, \mu, \lambda) = \ell(\pi, \mu) - \lambda_1 \left(\sum_k^K \pi_k - 1 \right) - \lambda_2 \left(\sum_w \mu_{wk} - 1 \right)$$

$$\frac{\partial \mathcal{L}}{\partial \mu_{wk}} = \sum_d \frac{\pi_k \prod_w \mu_{wk}^{T_{dw}}}{\sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}}} \frac{T_{dw}}{\mu_{wk}} - \lambda_2$$

Let $\gamma_{dk} = \dfrac{\pi_k \prod_w \mu_{wk}^{T_{dw}}}{\sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}}}$ be responsibility

$$\frac{\partial \mathcal{L}}{\partial \mu_{wk}} = \sum_d \gamma_{dk} \frac{T_{dw}}{\mu_{wk}} - \lambda_2$$

$$\sum_d \gamma_{dk} T_{dw} = \lambda_2 \mu_{wk}$$

$$\sum_w \sum_d \gamma_{dk} T_{dw} = \lambda_2 \sum_w \mu_{wk}$$

$$\lambda_2 = \sum_w \sum_d \gamma_{dk} T_{dw}$$

$$\mu_{wk} = \frac{\sum_d \gamma_{dk} T_{dw}}{\sum_w \sum_d \gamma_{dk} T_{dw}}$$

//

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = \sum_d \frac{\pi_k \prod_w \mu_{wk}^{T_{dw}}}{\sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}}} \frac{1}{\pi_k} - \lambda_1 \cdots$$

$$\sum_d \gamma_{dk} \cancel{\frac{1}{\pi_k}} = \lambda_1 \pi_k$$

$$\sum_k \sum_d \gamma_{dk} = \lambda_1 \sum_k \pi_k$$

$$\lambda_1 = \sum_k \sum_d \gamma_{dk}$$

$$\pi_k = \frac{\sum_d \gamma_{dk}}{\sum_k \sum_d \gamma_{dk}}$$

$$\pi_k = \frac{\sum_d \gamma_{dk}}{D} \quad //$$

E step: For each document d, update the responsibility

$$\gamma_{dk}^{(t+1)} = \frac{\pi_k \prod_w \mu_{wk}^{T_{dw}}}{\sum_{k=1}^{K} \pi_k \prod_w \mu_{wk}^{T_{dw}}}$$

This corresponds to the ~~responsibility of~~ posterior probability of topic $C_d = k$ for generating the document.

M step: For each $C_d = k$, update parameter estimates

$$\pi_k^{(t+1)} = \frac{\sum_d \gamma_{dk}}{D}$$

$$\mu_{wk}^{(t+1)} = \frac{\sum_d \gamma_{dk} T_{dw}}{\sum_w \sum_d \gamma_{dk} T_{dw}}$$

$\pi_k^{(t+1)}$: Numerator is no. of documents in topic $C_d = k$

Denominator is total number of documents D

π_k represents the proportion of documents in ~~e~~ each topic.

$\mu_{wk}^{(t+1)}$: Numerator is expected times of ~~words~~ word w in topic $C_d = k$

Denominator is total number of words in topic $C_d = k$.

μ_{wk} represents proportion of word w in topic k.

清华大学实验报告

系别________ 班号________ 姓名________ (同组姓名:________)

作实验日期　年　月　日 2018280351 教师评定:

Q2 $J = \frac{1}{N}\sum_{n=1}^{N} \| x_n - \tilde{x}_n \|^2$

$$J = \frac{1}{N}\sum_{n=1}^{N} \Big\| x_n - \sum_{i=1}^{d} z_{ni} u_i - \sum_{i=d+1}^{p} b_i u_i \Big\|^2$$

$$\frac{\partial J}{\partial z_{ni}} = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial}{\partial z_{ni}} \Big\| x_n - \sum_{i=1}^{d} z_{ni} u_i - \sum_{i=d+1}^{p} b_i u_i \Big\|^2$$

$$0 = \frac{2}{N}\sum_{n=1}^{N} \Big(x_n - \sum_{i=1}^{d} z_{ni} u_i - \sum_{i=d+1}^{p} b_i u_i \Big)(-u_i) \quad i = 1 \dots d$$

$$0 = \sum_{n=1}^{N} \left(x_n^T u_i - z_{ni} \right)$$

As z_{ni} is unique for each data point n

~~so~~ we can just look at each term in the summation

$$z_{ni} = x_n^T u_i \quad \forall i = 1 \dots d$$

$$\frac{\partial J}{\partial b_i} = \frac{1}{N}\sum_{n=1}^{N} \frac{\partial}{\partial b_i} \Big\| x_n - \sum_{i=1}^{d} z_{ni} u_i - \sum_{i=d+1}^{p} b_i u_i \Big\|^2$$

$$0 = \frac{2}{N}\sum_{n=1}^{N} \Big(x_n - \sum_{i=1}^{d} z_{ni} u_i - \sum_{i=d+1}^{p} b_i u_i \Big)(-u_i) \quad i = d+1 \dots p$$

$$0 = \sum_{n=1}^{N} \left(x_n^T u_i - b_i \right)$$

$$\sum_{n=1}^{N} b_i = \sum_{n=1}^{N} x_n^T u_i$$

$$b_i = \bar{x}_n^T u_i \quad \forall i = d+1 \dots p$$