Notes on Probabilistic Latent Semantic Analysis

https://hustwj.github.io/notes/
https://github.com/hustwj/notes

Notations

We first list related notations as follows.

- M is the number of documents. K is the number of topics. V is the number of distinct words.
- N_m is the length of the m-th document.
- $w_{m,n}(1 \leq m \leq M; 1 \leq n \leq N_m)$ is the n-th word in the m-th document.
- $z_{m,n}(1 \leq m \leq M; 1 \leq n \leq N_m)$ is the topic assigned to the n-th word in the m-th document, which is a latent (or hidden) variable.
- $\theta_m(1 \leq m \leq M)$ is topic proportion of the m-th document, which is a K-dimension vector. $\theta_{m,k}$ is the k-th element in θ_m, which is corresponding the proportion of the topic k in the m-th document, and $\sum_{k=1}^K \theta_{m,k} = 1$.
- $\phi_k(1 \leq k \leq K)$ is word proportion of the k-th topic, which is a V-dimension vector. $\phi_{k,v}$ is the v-th element in ϕ_k, which is corresponding to the proportion of the word v in the k-th topic. Each v is an element of the dictionary with V distinct words, and $\sum_{v=1}^V \phi_{k,v} = 1$.

We can represent $\theta_1 \cdots \theta_m \cdots \theta_M$ as a $M \times K$ matrix Θ.

$$\mathbf{\Theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_m \\ \vdots \\ \theta_M \end{bmatrix} = \begin{bmatrix} \theta_{1,1} & \dots & \theta_{1,k} & \dots & \theta_{1,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{m,1} & \dots & \theta_{m,k} & \dots & \theta_{m,K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{M,1} & \dots & \theta_{M,k} & \dots & \theta_{M,K} \end{bmatrix}$$

Notes:
In the pLSA model, Θ, θ_m and $\theta_{m,k}$ are parameters instead of random variables.

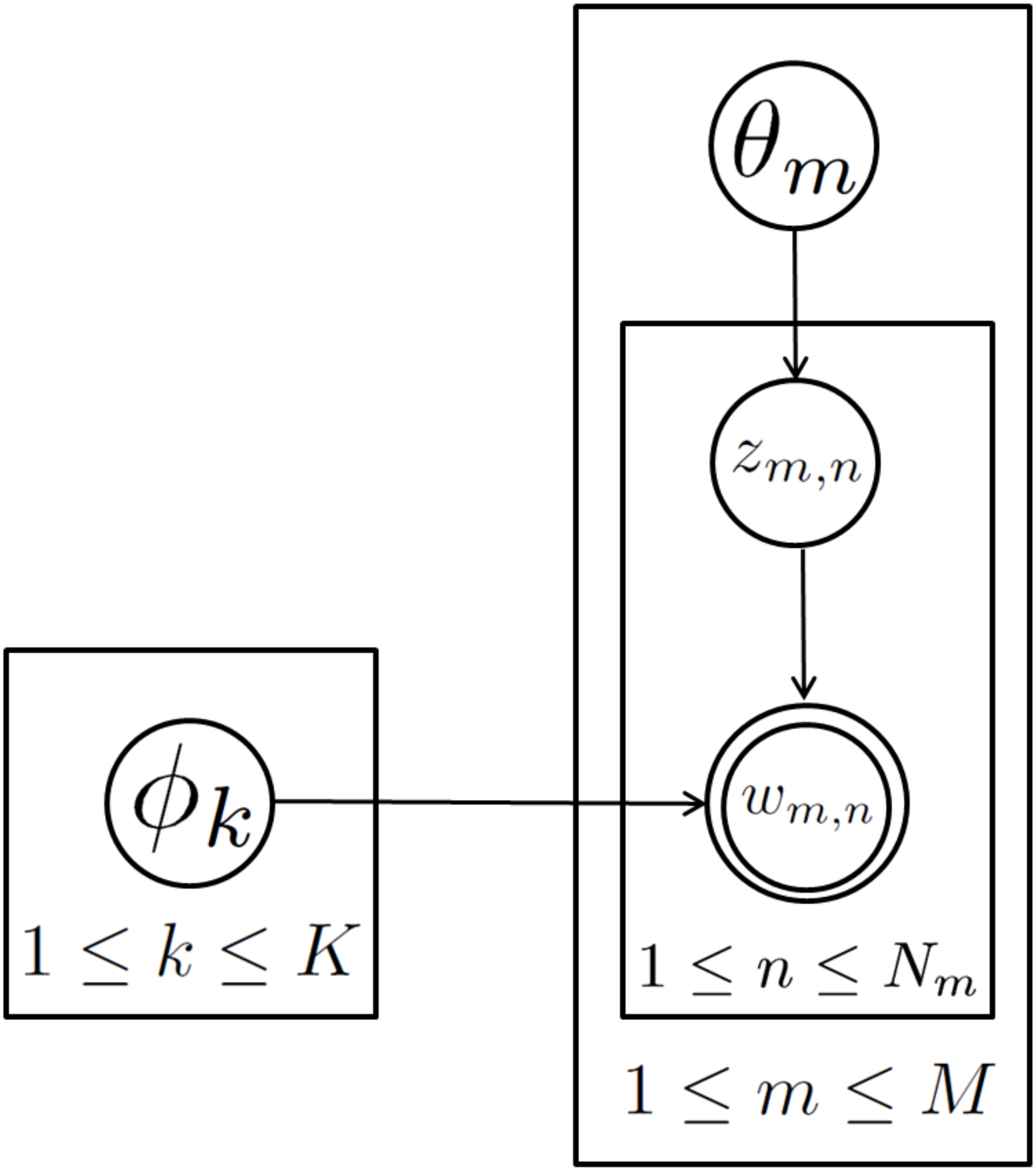
We can represent $\phi_1 \cdots \phi_k \cdots \phi_K$ as a $K \times V$ matrix Φ.

$$\mathbf{\Phi} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_k \\ \vdots \\ \phi_K \end{bmatrix} = \begin{bmatrix} \phi_{1,1} & \dots & \phi_{1,v} & \dots & \phi_{1,V} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{k,1} & \dots & \phi_{k,v} & \dots & \phi_{k,V} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{K,1} & \dots & \phi_{K,v} & \dots & \phi_{K,V} \end{bmatrix}$$

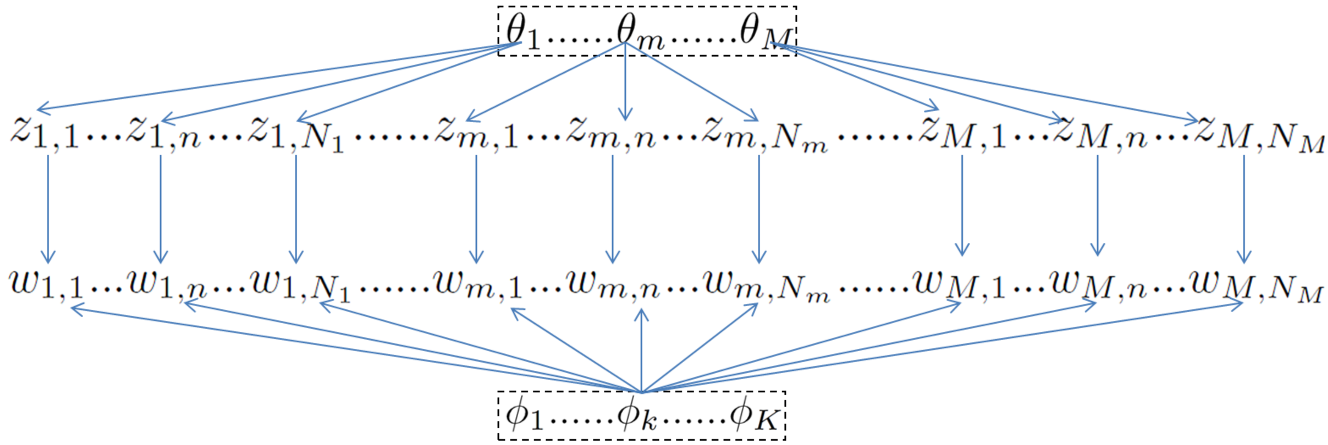
Notes:
In the pLSA model, Φ, ϕ_k and $\phi_{k,v}$ are parameters instead of random variables.

Graphical model for pLSA

Based on the generative process of pLSA, we can represent the pLSA model using “collapsed”’ plate notation.



For an easy understanding, the corresponding ``expanded'' model is also shown.



EM for pLSA

If the Maximum Likelihood Estimation (MLE) is used to estimate the parameters for the pLSA model, we need to fit the parameters which maximize probability observed data of all words W in documents, or likelihood of W with respect to parameters Θ and Φ.

$$\begin{aligned} W = (&w_{1,1} \cdots w_{1,n} \cdots w_{1,N_1} \\ &\cdots w_{m,1} \cdots w_{m,n} \cdots w_{m,N_m} \\ &\cdots w_{M,1} \cdots w_{M,n} \cdots w_{M,N_M}) \end{aligned}$$

We can get log-likelihood

$$\ell(\Theta, \Phi) = \log p(W; \Theta, \Phi) = \log(\prod_{m=1}^{M} \prod_{n=1}^{N_m} p(w_{m,n}; \Theta, \Phi))$$

$w_{m,n}$ is the n-th word in the m-th document, and m, n represents the position of the word in the documents. Words in different positions (corresponding to different values of m and n) can be instances of the same v in the dictionary. In the m-th document, we can set the total number of words which is equal to v as $c_{m,v}$, and $\sum_{v=1}^{V} c_{m,v} = N_m$. So we can store the value of $c_{m,v}$ into a $M \times V$ matrix C. The word v is observed with the probability of $p(v; \Theta, \Phi)$, and the corresponding topic of v is z.

$$\begin{aligned}\ell(\Theta, \Phi) &= \log(\prod_{m=1}^{M}\prod_{v=1}^{V} p(v;\Theta,\Phi)^{c_{m,v}}) \\ &= \sum_{m=1}^{M}\sum_{v=1}^{V}(c_{m,v} \times \log p(v;\Theta,\Phi)) \\ &= \sum_{m=1}^{M}\sum_{v=1}^{V}(c_{m,v} \times \log \sum_{z} p(v,z;\Theta,\Phi)) \\ &= \sum_{m=1}^{M}\sum_{v=1}^{V}(c_{m,v} \times \log \sum_{k=1}^{K} p(v,z=k;\Theta,\Phi))\end{aligned}$$

Based on the generative process of pLSA, in the m-th document, we can get

$$p(v,z;\Theta,\Phi) = p(z;\theta_m)p(v|z;\Phi) = p(z;\theta_m)p(v|z;\phi_z)$$

Based on the notations we can get

$$\begin{aligned}p(z=k;\theta_m) &= \theta_{m,k} \\ p(v|z=k;\phi_k) &= \phi_{k,v} \\ p(v,z=k;\Theta,\Phi) &= \theta_{m,k} \times \phi_{k,v}\end{aligned}$$

Based on Jensen's inequality, we have

$$
\begin{aligned}
\ell(\Theta, \Phi) &= \sum_{m=1}^{M} \sum_{v=1}^{V} (c_{m,v} \times \log \sum_{k=1}^{K} p(v, z=k; \Theta, \Phi)) \\
&= \sum_{m=1}^{M} \sum_{v=1}^{V} (c_{m,v} \times \log \sum_{k=1}^{K} \theta_{m,k} \times \phi_{k,v}) \\
&= \sum_{m=1}^{M} \sum_{v=1}^{V} (c_{m,v} \times \log \sum_{k=1}^{K} (Q_{m,v}(z=k) \times \frac{\theta_{m,k} \times \phi_{k,v}}{Q_{m,v}(z=k)})) \\
&\geq \sum_{m=1}^{M} \sum_{v=1}^{V} (c_{m,v} \times \sum_{k=1}^{K} (Q_{m,v}(z=k) \times \log \frac{\theta_{m,k} \times \phi_{k,v}}{Q_{m,v}(z=k)})) \\
&= L(\Theta, \Phi)
\end{aligned}
$$

E-step

In the E-step in EM method, we calculate the value of $Q_{m,v}(z=k)$, with which the lower bound $L(\Theta, \Phi)$ is equal to $\ell(\Theta, \Phi)$, as follows:

$$
\begin{aligned}
Q_{m,v}(z=k) &= \frac{p(v, z=k; \Theta, \Phi)}{\sum_{k'=1}^{K} p(v, z=k'; \Theta, \Phi)} \\
&= \frac{\theta_{m,k} \times \phi_{k,v}}{\sum_{k'=1}^{K} \theta_{m,k'} \times \phi_{k',v}}
\end{aligned}
$$

Notes:

$$
\frac{p(v, z=k; \Theta, \Phi)}{\sum_{k'=1}^{K} p(v, z=k'; \Theta, \Phi)} = \frac{p(v, z=k; \Theta, \Phi)}{p(v; \Theta, \Phi)} = p(z=k|v; \Theta, \Phi)
$$

So, actually we have

$$
Q_{m,v}(z=k) = p(z=k|v; \Theta, \Phi)
$$

M-step

In the M-step of the EM method, we assign the value of $Q_{m,v}(z=k)$ calculated from the E-step, and try to maximize the lower bound $L(\Theta, \Phi)$ with respect to Θ and Φ.

Notes:
Each $Q_{m,v}(z=k)$ in the M-step is a fixed value instead of variable.

Because $\sum_{k=1}^{K} \theta_{m,k} = 1$ and $\sum_{v=1}^{V} \phi_{k,v} = 1$, we can use the method of Lagrange multiplier.

$$
\begin{aligned}
L(\Theta, \Phi) = & \sum_{m=1}^{M} \sum_{v=1}^{V} (c_{m,v} \times \sum_{k=1}^{K} (Q_{m,v}(z=k) \times \log \frac{\theta_{m,k} \times \phi_{k,v}}{Q_{m,v}(z=k)})) \\
& + \sum_{k=1}^{K} \lambda_k (1 - \sum_{v=1}^{V} \phi_{k,v}) \\
& + \sum_{m=1}^{M} \rho_m (1 - \sum_{k=1}^{K} \theta_{m,k})
\end{aligned}
$$

First, we get the partial derivative with respect to $\phi_{k,v}$.

$$\frac{\partial L}{\partial \phi_{k,v}} = \sum_{m=1}^{M} (c_{m,v} \times Q_{m,v}(z=k)) - \lambda_k \times \phi_{k,v} = 0$$

And

$$\phi_{k,v} = \frac{1}{\lambda_k} \times \sum_{m=1}^{M} (c_{m,v} \times Q_{m,v}(z=k))$$

Because $\sum_{v=1}^{V} \phi_{k,v} = 1$, we can get

$$\sum_{v=1}^{V} \phi_{k,v} = \frac{1}{\lambda_k} \times \sum_{v=1}^{V} \sum_{m=1}^{M} (c_{m,v} \times Q_{m,v}(z=k)) = 1$$

So we can get

$$\lambda_k = \sum_{v=1}^{V}\sum_{m=1}^{M}(c_{m,v} \times Q_{m,v}(z=k))$$

$$\phi_{k,v} = \frac{\sum_{m=1}^{M}(c_{m,v} \times Q_{m,v}(z=k))}{\sum_{v=1}^{V}\sum_{m=1}^{M}(c_{m,v} \times Q_{m,v}(z=k))}$$

Similarly, we get the partial derivative with respect to $\theta_{m,k}$.

$$\frac{\partial L}{\partial \theta_{m,k}} = \sum_{v=1}^{V}(c_{m,v} \times Q_{m,v}(z=k)) - \rho_m \times \theta_{m,k} = 0$$

And

$$\theta_{m,k} = \frac{1}{\rho_m} \times \sum_{v=1}^{V}(c_{m,v} \times Q_{m,v}(z=k))$$

Because $\sum_{k=1}^{K}\theta_{m,k} = 1$, we can get

$$\sum_{k=1}^{K}\theta_{m,k} = \frac{1}{\rho_m} \times \sum_{k=1}^{K}\sum_{v=1}^{V}(c_{m,v} \times Q_{m,v}(z=k)) = 1$$

Because $\sum_{k=1}^{K} Q_{m,v}(z=k) = 1$ and $\sum_{v=1}^{V} c_{m,v} = N_m$, we can get

$$\begin{aligned}\rho_m &= \sum_{k=1}^{K}\sum_{v=1}^{V}(c_{m,v} \times Q_{m,v}(z=k)) \\ &= \sum_{v=1}^{V}\sum_{k=1}^{K}(c_{m,v} \times Q_{m,v}(z=k)) \\ &= \sum_{v=1}^{V}(c_{m,v} \times \sum_{k=1}^{K} Q_{m,v}(z=k)) \\ &= \sum_{v=1}^{V}(c_{m,v} \times 1) \\ &= N_m\end{aligned}$$

We can further get

$$\theta_{m,k} = \frac{\sum_{v=1}^{V}(c_{m,v} \times Q_{m,v}(z=k))}{N_m}$$