Text Mining — Assignment #2

Roger Cuscó, Matthew-Sudmann-Day and Miquel Torrens

Exercise 1

Answer.

Exercise 2

Part (a)

The components of the model are the following:

- Parameters:
  - $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_k, \ldots, \rho_K) \in \mathbb{R}^K$
  - $\mathbf{B}^{(1)} = \left(\beta^{(1)}_{z_1}, \ldots, \beta^{(1)}_{z_i}, \ldots, \beta^{(1)}_{z_N}\right)$, where $\beta^{(1)}_{z_i} \in \Delta^{V_1 - 1}, \forall i$
  - $\mathbf{B}^{(2)} = \left(\beta^{(2)}_{z_1}, \ldots, \beta^{(2)}_{z_i}, \ldots, \beta^{(2)}_{z_N}\right)$, where $\beta^{(2)}_{z_i} \in \Delta^{V_2 - 1}, \forall i$

  For notational purposes, let $\mathbf{B} = \{\mathbf{B}^{(1)}, \mathbf{B}^{(2)}\}$.

- Latent variables:
  - $\mathbf{z} = (z_1, \ldots, z_i, \ldots, z_N) \in \mathbb{R}^N$, for $i \in \{1, \ldots, N\}$ observations

- The observed data:
  - $\mathbf{x}^{(1)}_i = (x^1_{i1}, \ldots, x^1_{iv_1}, \ldots, x^1_{iV_1}) \in \mathbb{R}^{V_1}, \forall i \in \{1, \ldots, N\}$
  - $\mathbf{x}^{(2)}_i = (x^2_{i1}, \ldots, x^2_{iv_2}, \ldots, x^2_{iV_2}) \in \mathbb{R}^{V_2}, \forall i \in \{1, \ldots, N\}$

  For notational purposes, let $\mathbf{x}_i = \left\{\mathbf{x}^{(1)}_i, \mathbf{x}^{(2)}_i\right\} \forall i$ and $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N\}$.

Part (b)

The density function is:

$$
\begin{aligned}
\mathbb{P}(\mathbf{x}_i | \boldsymbol{\rho}, \mathbf{B}) &= \mathbb{P}(\mathbf{x}^1_i, \mathbf{x}^2_i | \boldsymbol{\rho}, \mathbf{B}) \\
&= \mathbb{P}(\mathbf{x}^1_i | \boldsymbol{\rho}, \mathbf{B}) \mathbb{P}(\mathbf{x}^2_i | \boldsymbol{\rho}, \mathbf{B}) \quad \text{(Assume independence of the two vectors)} \\
&= \sum_{k=1}^{K} \mathbb{P}(z_k | \boldsymbol{\rho}, \mathbf{B}) \mathbb{P}(\mathbf{x}^1_i | z_k, \boldsymbol{\rho}, \mathbf{B}) \mathbb{P}(\mathbf{x}^2_i | z_k, \boldsymbol{\rho}, \mathbf{B}) \\
&= \sum_{k=1}^{K} \rho_k \mathbb{P}(\mathbf{x}^1_i | z_k, \boldsymbol{\rho}, \mathbf{B}) \mathbb{P}(\mathbf{x}^2_i | z_k, \boldsymbol{\rho}, \mathbf{B}) \\
&= \sum_{k=1}^{K} \rho_k \prod_{v_1}^{V_1} \beta_{v_1}^{x^1_{iv_1}} \prod_{v_2}^{V_2} \beta_{v_2}^{x^2_{iv_2}}.
\end{aligned}
$$

The likelihood function is:

$$
\ell(\mathbf{x} | \boldsymbol{\rho}, \mathbf{B}) = \prod_{i=1}^{D} \sum_{k=1}^{K} \rho_k \prod_{v_1=1}^{V_1} \beta_{v_1}^{x^1_{iv_1}} \prod_{v_2=1}^{V_2} \beta_{v_2}^{x^2_{iv_2}}.
$$

The complete likelihood writes as follows:

$$
\ell_{\text{comp}}(\mathbf{x}, \mathbf{z} | \boldsymbol{\rho}, \mathbf{B}) = \prod_{i=1}^{D} \prod_{k=1}^{K} \left[\rho_k \prod_{v_1=1}^{V_1} \beta_{v_1}^{x^1_{iv_1}} \prod_{v_2=1}^{V_2} \beta_{v_2}^{x^2_{iv_2}} \right]^{\mathbb{1}_{\{z_i = k\}}}.
$$

Finally, the complete log-likelihood gives:

$$\log \ell_{\text{comp}}(\mathbf{x}, \mathbf{z}|\boldsymbol{\rho}, \mathbf{B}) = \sum_{i=1}^{D} \sum_{k=1}^{K} \mathbb{1}_{\{z_i=k\}} \left[\log \rho_k + \sum_{v_1=1}^{V_1} x_{iv_1}^1 \log \beta_{v_1} + \sum_{v_2=1}^{V_2} x_{iv_2}^2 \log \beta_{v_2} \right].$$

Part (c)

Denote the function Q as:

$$Q(\boldsymbol{\rho}, \mathbf{B}) = \sum_{i=1}^{D} \sum_{k=1}^{K} \hat{z}_{i,k} \left[\log \rho_k + \sum_{v_1=1}^{V_1} x_{iv_1}^1 \log \beta_{v_1} + \sum_{v_2=1}^{V_2} x_{iv_2}^2 \log \beta_{v_2} \right],$$

where using the Bayes' rule as in the slides:

$$\hat{z}_{i,k} \propto \rho_k \prod_{v_1=1}^{V_1} \beta_{v_1}^{x_{iv_1}^1} \prod_{v_2=1}^{V_2} \beta_{v_2}^{x_{iv_2}^2}.$$

Part (d)

The Lagrangian is defined as:

$$\mathcal{L}(\boldsymbol{\rho}, \mathbf{B}, \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)}, \nu) = Q(\boldsymbol{\rho}, \mathbf{B}) + \nu \left(1 - \sum_{k=1}^{K} \rho_k \right) + \sum_{k=1}^{K} \lambda_k^{(1)} \left(1 - \sum_{v_1=1}^{V_1} \beta_{k,v_1} \right) + \sum_{k=1}^{K} \lambda_k^{(2)} \left(1 - \sum_{v_2=1}^{V_2} \beta_{k,v_2} \right).$$

We take derivatives with respect to the parameters:

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \rho_k} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{D} \frac{\hat{z}_{i,k}}{\rho_k} - \nu = 0 \tag{1}$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \nu} = 0 \quad \Leftrightarrow \quad \sum_{k=1}^{K} \rho_k = 1 \tag{2}$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \beta_{k,v_1}} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{D} \hat{z}_{i,k} x_{v_1,d} \frac{1}{\beta_{k,v_1}} - \lambda_k^{(1)} = 0 \tag{3}$$

$$\frac{\partial \mathcal{L}(\cdot)}{\partial \lambda_k^{(1)}} = 0 \quad \Leftrightarrow \quad \sum_{v_1=1}^{V_1} \beta_{v_1} = 1 \tag{4}$$

Combining (1) and (2) we see:

$$\nu = \frac{1}{\rho_k} \sum_{i=1}^{D} \hat{z}_{i,k} \Leftrightarrow \nu \underbrace{\sum_{k=1}^{K} \rho_k}_{=1} = \sum_{k=1}^{K} \sum_{i=1}^{D} \hat{z}_{i,k} \underbrace{\Leftrightarrow}_{\text{Plug } \nu \text{ into (1)}} \rho_k = \frac{\sum_{i=1}^{D} \hat{z}_{i,k}}{\sum_{k=1}^{K} \sum_{i=1}^{D} \hat{z}_{i,k}}$$

Analogously, combining (3) and (4) we get:

$$\beta_{k,v_1} \lambda_k^{(1)} = \sum_{i=1}^{D} \hat{z}_{i,k} x_{v_1,d} \Leftrightarrow \underbrace{\sum_{v_1=1}^{V_1} \beta_{k,v_1}}_{=1} \lambda_k^{(1)} = \sum_{i=1}^{D} \hat{z}_{i,k} \sum_{v_1=1}^{V_1} x_{v_1,d} \underbrace{\Leftrightarrow}_{\text{Plug } \lambda_k^{(1)} \text{ into (3)}} \beta_{k,v_1} = \frac{\sum_{i=1}^{D} \hat{z}_{i,k} x_{v_1,d}^1}{\sum_{i=1}^{D} \hat{z}_{i,k} \sum_{v_1=1}^{V_1} x_{v_1,d}^1}.$$

The exact same procedure can be applied (simply changing subscripts) to obtain:

$$\beta_{k,v_2} = \frac{\sum_{i=1}^{D} \hat{z}_{i,k} x_{v_2,d}^2}{\sum_{i=1}^{D} \hat{z}_{i,k} \sum_{v_2=1}^{V_2} x_{v_2,d}^2}.$$