

Assignment 15

李云志 2015K8009929014

2017 年 12 月 8 日

1 Topic Modeling over Short Text

原始目标函数：

$$\prod_{i=1}^N \prod_{j=1}^M P(w_i, w_j)^{n(w_i, w_j)}$$

取对数后得到需要优化的目标函数为：

$$\begin{aligned} & \sum_{i=1}^N \sum_{j=1}^M n(w_i, w_j) \log P(w_i, w_j) \\ (\text{from the PGM}) &= \sum_{i=1}^N \sum_{j=1}^M n(w_i, w_j) \log \sum_{k=1}^K P(w_i, w_j, z_k) \\ &= \sum_{i=1}^N \sum_{j=1}^M n(w_i, w_j) \log \sum_{k=1}^K P(z_k) P(w_i | z_k) P(w_j | z_k) \end{aligned}$$

由于 $\sum_{k=1}^K P(z_k) = 1$ 并且 \log 是凹函数，所以我们得到目标函数的下界：

$$L = \sum_{i=1}^N \sum_{j=1}^M n(w_i, w_j) \sum_{k=1}^K P(z_k) \log P(w_i | z_k) P(w_j | z_k)$$

此外，我们还有两个限制条件：

$$\begin{aligned} \sum_{i=1}^N P(w_i | z_k) &= 1 \\ \sum_{j=1}^M P(w_j | z_k) &= 1 \end{aligned}$$

由拉格朗日乘数法得到拉格朗日函数：

$$L + \sum_{k=1}^K \alpha_k (1 - \sum_{j=1}^M P(w_j | z_k)) + \sum_{k=1}^K \beta_k (1 - \sum_{i=1}^N P(w_i | z_k))$$

对各个自变量求导，并使其偏导为0得到

等式组1:

$$\sum_{i=1}^N n(w_i, w_j) P(z_k) - \alpha_k P(w_j | z_k) = 0, \text{ for } 1 \leq j \leq M, 1 \leq k \leq K$$

等式组2:

$$\sum_{j=1}^M n(w_i, w_j) P(z_k) - \beta_k P(w_i | z_k) = 0, \text{ for } 1 \leq i \leq N, 1 \leq k \leq K$$

解之得到迭代式:

$$P(w_j | z_k) = \frac{\sum_{i=1}^N n(w_i, w_j) P(z_k)}{\sum_{i=1}^N \sum_{j=1}^M n(w_i, w_j) P(z_k)}$$

$$P(w_i | z_k) = \frac{\sum_{j=1}^M n(w_i, w_j) P(z_k)}{\sum_{i=1}^N \sum_{j=1}^M n(w_i, w_j) P(z_k)}$$