

## 统计学习：第十七章

1 试将图 17.1 的例子进行潜在语义分析，并对结果进行观察。

答：

$$U = \begin{bmatrix} -0.14 & 0.52 & 0.39 & 0.55 \\ -0.15 & 0.22 & -0.69 & 0.43 \\ -0.28 & 0.50 & 0.41 & -0.26 \\ -0.68 & -0.57 & 0.25 & 0.10 \\ -0.13 & -0.18 & 0.07 & 0.57 \\ -0.63 & 0.27 & -0.38 & -0.33 \end{bmatrix},$$

$$S = \begin{bmatrix} 4.57 & 2.99 & 2.50 & 0.96 \end{bmatrix},$$

$$V^T = \begin{bmatrix} -0.32 & -0.34 & -0.63 & -0.61 \\ 0.77 & 0.32 & -0.04 & -0.55 \\ 0.48 & -0.86 & 0.05 & 0.17 \\ 0.26 & 0.21 & -0.77 & 0.54 \end{bmatrix}$$

可以看出 U 矩阵有 4 个列向量，第 1 列值均为负数，剩余三列有正有负。中间对角矩阵有 4 个奇异值，从大到小排列。V 矩阵也有 4 个列向量，第一列均为负值，其余三列有正有负。

2 给出损失函数是散度损失时的非负矩阵分解（潜在语义分析）的算法。

答：

$$\begin{aligned} \frac{\partial D(X||WH)}{\partial W_{il}} &= \sum_j X_{ij} \frac{(WH)_{ij}}{X_{ij}} \left( -\frac{X_{ij}}{(WH)_{ij}^2} \right) H_{lj} + H_{lj} \\ &= \sum_j H_{lj} \left( 1 - \frac{X_{ij}}{(WH)_{ij}^2} \right) \\ \frac{\partial D(X||WH)}{\partial H_{lj}} &= \sum_i W_{il} \left( 1 - \frac{X_{ij}^2}{(WH)_{ij}^2} \right) \end{aligned}$$

梯度下降法的更新规则为：

$$W_{il} = W_{il} + \lambda_{il} \left( \sum_j H_{lj} \left( \frac{X_{ij}}{(WH)_{ij}^2} \right) - 1 \right)$$

$$H_{lj} = H_{lj} + \mu_{lj} \left( \sum_i W_{il} \left( \frac{X_{ij}}{(WH)_{ij}^2} \right) - 1 \right)$$

令

$$\lambda_{il} = \frac{W_{il}}{\sum_j H_{lj}}, \mu_{lj} = \frac{H_{lj}}{\sum_i W_{il}}$$

则有定理 17.2 对应的乘法更新规则。

**3** 给出潜在语义分析的两种算法的计算复杂度，包括奇异值分解法和非负矩阵分解法。

答：奇异值分解法取决于奇异值分解计算为  $o(\min(m^2n, mn^2))$ ，非负矩阵分解法取决于梯度下降计算为  $o(mnk)$ 。

**4** 列出潜在语义分析与主成分分析的异同

答：二者均为矩阵降维特征提取的非监督学习算法，依赖奇异值分解实现。PCA 为通用邻域算法，LSA 为语义邻域算法；PCA 需要做数据标准化，LSA 需要统计词语和文本的频数或者权重。