Tsinghua-Berkeley Shenzhen Institute
LEARNING FROM DATA
Fall 2019

Homework 3

TIAN Chenyu October 17, 2019

- **Acknowledgments:** This template takes some materials from course CSE 547/Stat 548 of Washington University: `https://courses.cs.washington.edu/courses/cse547/17sp/index.html`.
- **Collaborators:** I finish this homework by myself.

2.1. The log-likelihood of the naive Bayes model can be writen as

$$\begin{aligned} l(\phi_y, \phi_j(x \mid y)) &= \sum_{i=1}^{m} \log p(x^{(i)}, y^{(i)}) \\ &= \sum_{i=1}^{m} (\log p(x^{(i)} \mid y^{(i)}) + \log p(y^{(i)})) \\ &= \sum_{i=1}^{m} (\sum_{j=1}^{d} \log p(x_j^{(i)} \mid y^{(i)}) + \log p(y^{(i)})) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{d} \log \phi_j(x_j^{(i)} \mid y^{(i)}) + \sum_{i=1}^{m} \log \phi_{y^{(i)}} \end{aligned}$$

Also, there exist constrains for $\phi_y, \phi_j(x \mid y)$:

$$\sum_{y=1}^{k} \phi_y = 1$$

$$\sum_{x \in \{0,1\}} \phi_j(x \mid y) = 1 \quad y \in \{1, ..., K\}$$

Then to find the maximum value of l with bringing these constrains into equation using lagrange multiplier

$$\begin{aligned} g(l, \lambda, \lambda_{jk}) = \sum_{i=1}^{m} \sum_{j=1}^{d} \log \phi_j(x_j^{(i)} \mid y^{(i)}) + \sum_{i=1}^{m} \log \phi_{y^{(i)}} \\ + \lambda(\sum_{y=1}^{k} \phi_y - 1) + \lambda_{jk}(\sum_{x \in \{0,1\}} \phi_j(x \mid k) - 1) \end{aligned}$$

Using lagrange multiplier method to find the parameters of the maximum value, there exists:

$$\frac{dg}{d\phi_y} = \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = y) \frac{1}{\phi_y} + \lambda = 0$$

$$\Rightarrow \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 1) \frac{1}{\phi_1} = \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = 2) \frac{1}{\phi_2} = ... = \sum_{i=1}^{m} \mathbb{1}(y^{(i)} = K) \frac{1}{\phi_K}$$

Then here has

$$\phi_y = \frac{\sum_{i=1}^m \mathbb{1}(y^{(i)} = y)}{m}$$

Also,

$$\frac{dg}{d\phi_j(x \mid y)} = \sum_{i=1}^m \mathbb{1}(y^{(i)} = y)\,\mathbb{1}(x_j^{(i)} = x)\frac{1}{\phi_j(x \mid y)} + \lambda_j(x \mid y) = 0$$

$$\Rightarrow \sum_{i=1}^m \mathbb{1}(y^{(i)} = y)\,\mathbb{1}(x_j^{(i)} = 0)\frac{1}{\phi_j(x = 0 \mid y)} = \sum_{i=1}^m \mathbb{1}(y^{(i)} = y)\,\mathbb{1}(x_j^{(i)} = 1)\frac{1}{\phi_j(x = 1 \mid y)}$$

Thus,

$$\phi_j(x \mid y) = \frac{\sum_{i=1}^m \mathbb{1}(y^{(i)} = y)\,\mathbb{1}(x_j^{(i)} = x)}{\sum_{i=1}^m \mathbb{1}(y^{(i)} = y)}$$