SKETCH OF THE "ALTERNATING SQP" METHOD FOR FITTING POISSON TOPIC MODELS

PETER CARBONETTO*

1. Some derivations. Given an $n \times p$ matrix of counts X with individual entries $x_{ij} \geq 0$, our aim is to fit a Poisson model of the counts,

$$p(x) = \prod_{i=1}^{n} \prod_{j=1}^{p} p(x_{ij})$$

(1.1)
$$= \prod_{i=1}^{n} \prod_{j=1}^{p} \text{Poisson}(x_{ij}; \lambda_{ij}),$$

in which the Poisson rates are given by the mixture $\lambda_{ij} = \sum_{k=1}^{K} l_{ik} f_{jk}$. Therefore, the Poisson model is specified by a $p \times K$ matrix F with entries $f_{ik} \geq 0$ (the "factors"), and an $n \times K$ matrix L with entries $l_{ik} \geq 0$ (the "loadings"). Fitting F and L is equivalent to non-negative matrix factorization with the "beta-divergence" cost function [2]. It can also be used to recover a maximum-likelihood estimate for the latent Dirichlet allocation (LDA) model [1]. So fitting this model is useful for a wide range of applications.

The log-likelihood for the Poisson model is

(1.2)
$$\log p(x \,|\, F, L) \propto \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij} \log(\textstyle\sum_{k=1}^{K} l_{ik} f_{jk}) - \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{K} l_{ik} f_{jk},$$

where the constant of proportionality is obtained from factorial terms in the Poisson densities. We propose a slight modification to this log-likelihood to allow for missing counts, which may be useful for some applications:

(1.3)
$$\log p(x \,|\, F, L) \propto \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij} \log(\textstyle\sum_{k=1}^{K} a_{ij} l_{ik} f_{jk}) - \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{K} a_{ij} l_{ik} f_{jk},$$

in which we introduce an $n \times p$ matrix A such that $a_{ij} = 0$ and $x_{ij} = 0$ whenever the count is missing (we also define $x \log x$ to be zero whenever x is zero). Our specific aim is to find a F and L that maximizes the log-likelihood (1.3); that is, we would like to solve

(1.4)
$$\begin{aligned} \text{minimize} \quad & -\log p(x \,|\, F, L) \\ \text{subject to} \quad & F \geq 0, L \geq 0. \end{aligned}$$

In the remainder, we derive an efficient approach to doing this.

Our strategy for solving (1.4) is to alternate between solving for F with L fixed, and solving for L with F fixed. When solving for F with L fixed, and vice versa, the problem naturally decomposes into a collection of much smaller subproblems that are more tractable to solve. All the subproblems are of the following form:

(1.5)
$$\begin{aligned} \text{minimize} \quad & \phi(x; B, w) \\ \text{subject to} \quad & x_k \geq 0 \text{ for all } k = 1, \ldots, K, \end{aligned}$$

*Dept. of Human Genetics and the Research Computing Center, University of Chicago, Chicago, IL

1

in which the objective function is defined as

$$(1.6) \qquad \phi(x; B, w) = -\sum_{j=1}^{n} w_j \log \left(\sum_{k=1}^{K} B_{jk} x_k \right) + \sum_{j=1}^{n} \sum_{k=1}^{K} B_{jk} x_k.$$

To see the connection between subproblem (1.5) and the original optimization problem, observe that the negative log-likelihood can be written as

$$(1.7) \qquad -\log p(x \mid F, L) =$$

and it can be equally written as

$$(1.8) \qquad -\log p(x \mid F, L) =$$

REFERENCES

[1] D. M. BLEI, A. Y. NG, AND M. I. JORDAN, *Latent Dirichlet allocation*, Journal of Machine Learning Research, 3 (2003), pp. 993–1022.
[2] D. D. LEE AND H. S. SEUNG, *Algorithms for non-negative matrix factorization*, in Advances in Neural Information Processing Systems 13, 2001, pp. 556–562.