

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 entries $x_{ij} \geq 0$, the aim is to fit a Poisson model of the counts,

$$(1.1) \quad \begin{aligned} p(x) &= \prod_{i=1}^n \prod_{j=1}^p p(x_{ij}) \\ &= \prod_{i=1}^n \prod_{j=1}^p \text{Poisson}(x_{ij}; \lambda_{ij}), \end{aligned}$$

in which the Poisson rates are given by $\lambda_{ij} = \sum_{k=1}^K l_{ik} f_{jk}$. The model is determined 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

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

where the constant of proportionality is obtained from factorial terms in the Poisson densities. Our specific aim is to find a F and L that maximizes the log-likelihood (1.2); that is, we would like to solve

$$(1.3) \quad \begin{aligned} &\text{minimize} && -\log p(x | F, L) \\ &\text{subject to} && F \geq 0, L \geq 0. \end{aligned}$$

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

Our strategy for solving (1.3) 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:

$$(1.4) \quad \begin{aligned} &\text{minimize} && \phi(y; B, w) \\ &\text{subject to} && y_k \geq 0 \text{ for all } k = 1, \dots, K, \end{aligned}$$

in which the objective function is defined as

$$(1.5) \quad \phi(y; B, w) = - \sum_{i=1}^n w_i \log(\sum_{k=1}^K b_{ik} y_k) + \sum_{i=1}^n \sum_{k=1}^K b_{ik} y_k.$$

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

To see the connection between subproblem (1.4) and the original optimization problem, observe that the negative log-likelihood can be written as

$$(1.6) \quad -\log p(x | F, L) = \sum_{i=1}^n \phi(l_i; F, x_i),$$

where x_i is the i th row of X and l_i is the i th row of L , and it can additionally be written as

$$(1.7) \quad -\log p(x | F, L) = \sum_{j=1}^p \phi(f_j; L, x_j),$$

in which x_j is the j th column of X , and f_j is the j th row of F . Therefore, when F is unchanging, each row of L can be optimized separately by solving a problem of the form (1.4), and when L is fixed, each row of F can be optimized separately by solving a problem of the form (1.4).

Instead, we will solve the following optimization problem,

$$(1.8) \quad \begin{array}{ll} \text{minimize} & f(t; P, u) \\ \text{subject to} & t_k \geq 0 \text{ for all } k = 1, \dots, K, \end{array}$$

in which the objective function is defined as

$$(1.9) \quad f(t; P, u) = -\sum_{i=1}^n u_i \log \left(\sum_{k=1}^K p_{ik} t_k \right) + \sum_{k=1}^K t_k.$$

This is equivalent to the above when

$$\begin{aligned} u_k &= w_k / \left(\sum_{k'} w_{k'} \right) \\ p_{ik} &= b_{ik} \times \sum_i w_i / \sum_i b_{ik} \end{aligned}$$

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.