Phillip Nicol HSPH Cancer WG

Estimation in log-bilinear models

Date: March 6, 2024 *Boston, MA*

1 Model setup

- Single-cell count matrix $Y \in \mathbb{R}^{I \times J}$ (I genes, J cells).
- Outcome distribution $Y_{ij} \sim \text{Pois}(\mu_{ij})$
- Log link $\log(\mu_{ij}) = \alpha_i + \beta_j + \sum_{m=1}^{M} \sigma_m u_{im} v_{jm}$

In matrix form,

$$\log(\mu) = \alpha 1_J^\top + 1_I \beta^\top + U\Sigma V^\top \tag{1}$$

where $U^\top U = V^\top V = I_M$ and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_M)$.

$V \in \mathbb{R}^{J \times M}$ provides a low-dimensional embedding of cells.

2 Existence of MLE

Define $X = U\Sigma V^\top$ so the log-likelihood can be written

$$\ell(\alpha, \beta, X) = \text{const} + \sum_{ij} Y_{ij}(\alpha_i + \beta_j + X_{ij}) - \exp(\alpha_i + \beta_j + X_{ij}) \tag{2}$$

The MLE is a solution to the constrained optimization problem

$$\text{argmax}_{X:\text{rk}(X)=M} \ell(\alpha, \beta, X) \tag{3}$$

However, it is unclear if a solution to (3) exists. For motivation consider a non-existence result for GLMs:

Theorem 1 (Correia et al. (2019))**.** *For a Poisson GLM with log link and full rank model matrix $X \in \mathbb{R}^{n \times p}$. The MLE does not exist if and only if there is a non-zero $\gamma \in \mathbb{R}^p$ such that $y_i > 0 \Rightarrow (X\gamma)_i = 0$ and $y_i = 0 \Rightarrow (X\gamma)_i \leq 0$.*

The bilinear model is essentially a Poisson GLM where both the covariates and coefficients are simultaneously estimated.

Contrived example. $Y_{ij} = 1$ for all i, j except $Y_{11} = 0$. Note that $-IJ + 1$ is an upper bound to likelihood. Now take $\alpha_i = \beta_j = 0$, $M = 1$, $u = e_1 \in \mathbb{R}^I$, $v = -e_1 \in \mathbb{R}^J$. Then

$$\lim_{\sigma_1 \to \infty} \ell(\alpha, \beta, \sigma_1 u v^\top) = -IJ + 1 \tag{4}$$

So the MLE is undefined.

A slight generalization of this is that the MLE will not exist when Y is binary and

$$M \geq \text{rank}(1 - Y) \tag{5}$$

We would like to know if there are more general conditions where MLE does not exist.

Failed proof. Starting from arbitrary $\alpha, \beta, X = U\Sigma V^\top$, we would like to show that perturbing X in a particular direction always increases likelihood. In particular, let $\tilde{X} \in \mathbb{R}^{I \times J}$ such that $Y_{ij} > 0 \Rightarrow \tilde{X}_{ij} = 0$ and $Y_{ij} = 0 \Rightarrow \tilde{X}_{ij} \leq 0$. Now it is straightforward to see that

$$\ell(\alpha, \beta, X + \lambda\tilde{X}) - \ell(\alpha, \beta, X) > 0 \tag{6}$$

for any $\lambda > 0$.
Unfortunately, $X + \lambda\tilde{X}$ will in general not satisfy the rank M constraint. One sufficient condition for rank at most M is $\mathrm{row}(\tilde{X}) \subset \mathrm{row}(X)$ or $\mathrm{col}(\tilde{X}) \subset \mathrm{col}(X)$. However, it is not clear (and seems unlikely) that for arbitrary X we can find such a $\tilde{X}$.

3 Stabilizing singular values

We consider a Bayesian approach by placing independent exponential priors on σ_m:

$$p(\sigma_m) = \lambda \exp(-\lambda\sigma_m) \tag{7}$$

Now the MAP estimate is

$$\mathrm{argmin}_{X:\mathrm{rk}(X)=M} - \ell(\alpha, \beta, X) + \lambda \sum_{m=1}^{M} \sigma_m(X) \tag{8}$$

4 Estimation with proximal gradient descent

General optimization problem[1]

$$\mathrm{argmin}_x f(x) + h(x) \tag{9}$$

where f is convex and smooth and h is convex but not necessarily differentiable. Many problems can be written in this form.

Example 1. $f(\beta) = ||y - X\beta||_2^2$ and $h(\beta) = \lambda||\beta||_1$ is the LASSO.

Example 2. For Poisson log-bilinear model $f(X) = -\ell(\alpha, \beta, X)$, $h(X) = 0$ if $\mathrm{rank}(X) \leq M$ and $h(X) = \infty$ if $\mathrm{rk}(X) > M$

Idea: Replace f with quadratic approximation

$$f(x) = f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x}) + \frac{1}{2\gamma}||x - \hat{x}||_2^2 \tag{10}$$

Note that this avoids calculating the Hessian of f as this could be prohibitively large. A quick calculation shows the solution is

$$\mathrm{prox}\left(\hat{x} - \gamma\nabla f(\hat{x})\right) \tag{11}$$

where

$$\mathrm{prox}_\gamma(x) := \mathrm{argmin}_z \frac{1}{2\gamma}||x - z||_2^2 + h(z) \tag{12}$$

Example 1. For LASSO,

$$\mathrm{prox}(\beta) = S_{\lambda\gamma}(\beta) \tag{13}$$

where $S_{\lambda t}$ is the soft-threshold operator:

$$S_{\lambda\gamma}(\beta_j) = \begin{cases} \beta_j - \lambda\gamma & \text{if } \beta_j > \lambda\gamma \\ 0 & \text{if } |\beta_j| < \lambda\gamma \\ \beta_j + \lambda\gamma & \text{if } \beta_j < -\lambda\gamma \end{cases} \tag{14}$$

[1] This section is based on these lecture notes.

Moreover, $\nabla f(\beta) = X^\top(y - X\beta)$ so a simple algorithm to fit LASSO is to iteratively soft-threshold the estimated β:

$$\hat{\beta}^{(t+1)} = S_{\lambda\gamma}\left(\hat{\beta}^{(t)} - \gamma X^\top(y - X\hat{\beta}^{(t)})\right) \tag{15}$$

Example 2. For the bilinear model,

$$\text{prox}(X) = \text{argmin}_{Z:\text{rk}(Z)=M}||X - Z||_F^2 = \text{SVD}_M(X) := U_X \Sigma_X V_X^\top \tag{16}$$

and

$$\nabla \ell(\alpha, \beta, X) = Y - \mu \tag{17}$$

which inspires the following iteratively reweighted SVD algorithm

$$\hat{X}^{(t+1)} = \text{SVD}_M\left(\hat{X}^{(t)} + \gamma(Y - \hat{\mu})\right) \tag{18}$$

It can be shown that when the exponential prior is added to the singular values, the proximal step becomes

$$\text{Prox}(X) = U_X \text{diag}((\sigma_{1X} - \lambda)_+, \ldots, (\sigma_{mX} - \lambda)_+)V_X^\top \tag{19}$$

So the estimation algorithm is the same except for the additional step of soft-thresholding the singular values.

References

S. Correia, P. Guimarães, and T. Zylkin. Verifying the existence of maximum likelihood estimates for generalized linear models. *arXiv preprint arXiv:1903.01633*, 2019.