A Appendix

Let $f(y_i; \mu_i, \phi)$ be the density function for a negative binomial distribution with mean μ_i and dispersion parameter ϕ:

$$f(y_i; \mu_i, \phi) = \frac{\Gamma(y_i + 1/\phi)}{y_i!\Gamma(1/\phi)} \left(\frac{1}{1+\phi\mu_i}\right)^{1/\phi} \left(\frac{\phi\mu_i}{1+\phi\mu_i}\right)^{y_i}. \qquad (9)$$

As $\phi \to 0$, $f(y_i; \mu_i, \phi)$ converges to a Poisson distribution with mean μ_i. The log likelihood is

$$l(\mathbf{y}, \boldsymbol{\mu}, \phi) \;=\; \sum_{i=1}^{n} \left[\log\left(\frac{\Gamma(y_i + 1/\phi)}{y_i!\Gamma(1/\phi)}\right) - \left(\frac{1}{\phi} + y_i\right)\log(1+\phi\mu_i) + y_i \log(\phi\mu_i)\right].$$

where $\mathbf{y} = (x_1, ..., x_n)^\mathsf{T}$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)^\mathsf{T}$, and n indicates sample size. We further assume

$$\log(\mu_i) = b_0 + \sum_{j=1}^{J} x_{ij} b_j$$

and maximize the penalized log likelihood

$$l(\mathbf{y}, \boldsymbol{\mu}, \phi) - \sum_{j=1}^{J} (1+\delta)\log(|b_j| + \tau).$$

The Iteratively Reweighted Least Squares (IRLS) approach is often used to maximize the log likelihood of a generalized linear model by solving a series of lease squares problem. Given the current parameter estimates $\tilde{b}_j$ $(0 \le j \le J)$ and $\tilde{\phi}$, the negative (unpenalized) log likelihood can be written as a quadratic form

$$l_Q(\mathbf{y}, \boldsymbol{\mu}, \phi) = -\sum_{i=1}^{n} w_i \left[z_i - b_0 - \sum_{j=1}^{J} x_{ij} b_j\right]^2,$$

where $z_i = \tilde{\mu}_i + (y_i - \tilde{\mu}_i)/v(y_i)$, $\tilde{\mu}_i = \tilde{b}_0 + \sum_{j=1}^{J} x_{ij}\tilde{b}_j$, $v(y_i) = \tilde{\mu}_i + \tilde{\mu}_i^2\tilde{\phi}$, and $w_i = v(y_i)$. Next, given new estimates of b_j $(0 \le j \le J)$, ϕ can be re-estimated by a maximize conditional likelihood of ϕ.

Similar to the approach proposed by Friedman et al. [Friedman et al., 2010], We employed an ECM algorithm [Sun et al., 2010] named iterative adaptive Lasso (IAL) to solve this problem of penalized generalized linear model, which include four levels of loops:

- Loop 1 (outer loop): Iterate all combinations of tuning parameter δ and τ.
- Loop 2: Alternately updates regression coefficients b_j $(0 \leq j \leq J)$ and over dispersion parameter ϕ.
- Loop 3: Update the quadratic approximation l_Q using current estimate of b_j $(0 \leq j \leq J)$ and ϕ.
- Loop 4 (inner loop): Run the IAL to solve the penalized weighted least squares problem.