ST758, Homework 6

Due Nov 26, 2013

In Homework 3, you implemented the PageRank algorithm to rank webpages according to their link structure. In this homework, we rank sports teams. Consider a league of p teams. Each team has a strength parameter $\gamma_i > 0$.Team i beats team j with probability $\gamma_i/(\gamma_i + \gamma_j)$. For simplicity, we assume ties are impossible. Let y_{ij} be the number of times team i beats team j. Assuming independence between the games, the log-likelihood is

$$L(\boldsymbol{\gamma}) = \ln \prod_{i,j} \left(\frac{\gamma_i}{\gamma_i + \gamma_j} \right)^{y_{ij}} = \sum_{i,j} y_{ij} [\ln \gamma_i - \ln(\gamma_i + \gamma_j)].$$

We estimate parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ via MLE.

1. Is the log-likelihood function $L(\boldsymbol{\gamma})$ concave?

2. Derive a minorization-maximization algorithm for maximizing $L(\boldsymbol{\gamma})$. (Hint: minorize the $-\ln(\gamma_i + \gamma_j)$ term by its supporting hyperplane.)

3. Implement the MM algorithm derived in 2. For identifiability, you can set $\gamma_1 = 1$.

4. If we reparameterize $\gamma_i = e^{\lambda_i}$, where $\lambda_i \in \mathbb{R}$, is the log-likelihood function concave in new parameters $\boldsymbol{\lambda}$?

5. Denote the reparameterized log-likelihood function $L(\boldsymbol{\lambda})$. Derive the score and observed information of $L(\boldsymbol{\lambda})$.

6. Implement a Newton algorithm for maximizing $L(\boldsymbol{\lambda})$. Again for identifiability, you can set $\lambda_1 = 0$.

7. Consider the case p is large, e.g., the number of users in a popular online game. What structure in the Hessian can be exploited in the Newton method?

8. Consider the ESPN's MLB Standings Grid for a win-loss matchups of any two MLB teams for the 2013 season `http://espn.go.com/mlb/standings/grid/_/year/2013`.

   Retrieve the data for the American League and apply your functions to rank the teams within the American League. Note the last column contains a team's overall record versus the National League. Since we are ranking teams only within the American League, we ignore this information.

9. Please, present your results (ranking, strengths, win probabilities, etc) in a nice, human-readable way. Imagine you're presenting your rankings and predictions in your blog ☺