Empirical Bayes Matrix Completion

Takeru Matsuda and Fumiyasu Komaki
The University of Tokyo, RIKEN Brain Science Institute

In various applications, we encounter problems of estimating the unobserved entries of a matrix from the observed entries. For example, in the famous Netflix problem, we have a matrix of movie ratings by users and aim to predict the preference for movies of each user for recommendation. This problem is called the matrix completion problem and many studies have investigated its theoretical properties and developed efficient algorithms.

In the matrix completion problems, the low-rank property of the underlying matrix plays a central role. Indeed, existing matrix completion algorithms succeed in estimating the unobserved entries by assuming the low-rankness. Note that low rank matrices have sparse singular values since the rank of a matrix is equal to the number of its nonzero singular values. The sum of singular values of a matrix is called the nuclear norm and employed by many existing algorithms for regularization.

In practice, the data matrix often contains observation noise and we aim to recover the true underlying matrix. If the data matrix is fully observed with the Gaussian observation noise, then the matrix completion problem reduces to the estimation of the mean matrix parameter of a matrix-variate normal distribution. For this problem, [1] developed an empirical Bayes estimator and proved that its minimaxity under the Frobenius loss. Based on this idea, [2] developed singular value shrinkage priors as a natural generalization of the Stein prior. The singular value shrinkage priors are superharmonic and the Bayes estimators based on them are minimax estimators with similar properties to the Efron–Morris estimator.

In this study [3], we develop an empirical Bayes (EB) algorithm for matrix completion. The EB algorithm is a natural extension of the Efron–Morris estimator and based on the following hierarchical model:

$$M \sim \mathrm{N}_{p,q}(0, I_p, \Sigma),$$

$$Y \mid M \sim \mathrm{N}_{p,q}(M, \sigma^2 I_p, I_q).$$

Namely, we assume that each row vector $m_i^\top \in \mathbb{R}^q$ of $M = (m_1, \cdots, m_p)^\top \in \mathbb{R}^{p \times q}$ has the distribution $m_i \sim \mathrm{N}_q(0, \Sigma)$ independently. Since only part of the entries of Y are observed, we use the EM algorithm to estimate the hyperparameters Σ and σ^2. The EB algorithm does not require heuristic parameter tuning other than tolerance. Numerical experiments demonstrate the effectiveness of the EB algorithm compared with existing algorithms. Specifically, the EB algorithm works well when the difference between the number of rows and columns is large. Application to real data also shows the practical utility of the EB algorithm.

References

[1] B. Efron and C. N. Morris. Empirical Bayes on vector observations: an extension of Stein's method, *Biometrika* **59**, 335–347, 1972.

[2] T. Matsuda and F. Komaki. Singular value shrinkage priors for Bayesian prediction. *Biometrika* **102**, 843–854, 2015.

[3] T. Matsuda and F. Komaki. Empirical Bayes Matrix Completion. arXiv:1706.01252.