Precision matrix estimations in high dimension and some related problems

Thi Khuyen LE, Aix Marseille University

Frédéric RICHARD, Aix Marseille University

Caroline CHAUX, Aix Marseille University

Consider the problem of estimating the inverse covariance matrix (a.k.a the precision matrix) in a high dimensional case under an assumption that the observations are independent and identically Gaussian distributed. This problem is concerned by many authors in recent years. Among the proposed methods, GLASSO (Graphical Least Absolute Shrinkage and Selection Operator) [1] is known as an effective method for solving this problem. Based on the GLASSO solutions, we adapt the Linear Discriminant Analysis (LDA) in high dimension by including a sparse estimate of the precision matrix over all population. We further propose a variable selection procedure based on the graph associated with the estimated precision matrix. For that, we define a discriminant capacity of each connected component of the graph and keep variables of the most discriminant components. The adapted-LDA and variable selection methods are both evaluated on synthetic data and applied to real data from PET brain images for the classification of patients with Alzheimer's disease.
We also consider the Fused GLASSO (FGL) problem [2], which is a generalization of the GLASSO, for estimating simultaneously two precision matrices when data belongs to two distinct classes. FGL problem is involved by two regularization terms: the first one penalizes the sparsity level of both estimated matrices, the second one penalizes their similarity. The estimators obtained by solving FGL problem are sparse and they tend to have the same values when the involved regularization parameters are large enough. By considering the intersection of two FGL models, we propose a test for the significance of each new point which enters to this model when the involved parameters change. Under the null hypothesis is that each connected component of the true model is contained within a connected component of the estimated model, we show that the test is asymptotically Exponential distribution.

Références

[1] Yuan, M. and Lin, Y., *Model selection and estimation in the Gaussian graphical model*, Biometrika, 2007.

[2] Danaher, P. and Wang, P. and Witten, D. M., *The joint graphical lasso for inverse covariance estimation across multiple classes*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2014.

Prénom NOM, Adresse Longue
`email`