Theoretical Foundations of the Analysis of Large Data Sets

James-Stein Phenomenon

Due 01.06.2017

1. Simulate 500 realizations of the random vector $X = (X_1, \dots, X_p) \sim N(\mu, I)$ where $p = 500$

   a) $\mu = 0$,

   b) μ is obtained by a simulation (just once) from $N(0, 5I)$,

   c) $\mu_1, \dots, \mu_p$ are obtained as iid from $N(20, 5)$ (just once).

   For each of these cases compare the mean square error of the maximum likelihood estimate X, classical James-Stein estimate $\hat{\mu}_{JS} = \left(1 - \frac{p-2}{||X||^2}\right) X$ and the Empirical Bayes estimate $\hat{\mu}_i^{EB} = \bar{X} + \left(1 - \frac{p-3}{S}\right)(X_i - \bar{X})$, where $S = \sum_{i=1}^{p}(X_i - \bar{X})^2$.

2. Simulate 500 realizations of the random vector $X = (X_1, \dots, X_p) \sim N(\mu, \Sigma)$ where $p = 500$, $\Sigma_{i,i} = 1$, for $i \neq j$ $\Sigma_{i,j} = 0.7$ and the vector μ is as in Problem 1.

   Compare the mean square error of the maximum likelihood estimate X with the extension of James-Stein estimate by Mary Ellen Bock (1975)

   $\mu_{MEB} = \left(1 - \frac{\tilde{p}-2}{X^T \Sigma^{-1} X}\right) X$, where $\tilde{p} = \frac{Tr(\Sigma)}{\lambda_{max}(\Sigma)}$.

Malgorzata Bogdan