ST790-003, Homework 6

Due Monday, Mar 30, 2015 @ 11:59PM

SDP, GP, MIP

This homework explores applications of SDP (semidefinite programming), GP (geometric programming), and MIP (mixed integer programming) in statistics. This is a *solo* homework. Discussion with fellow students is allowed but you have to write your code and report independently.

1. (Experiment design) Consider a linear model

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n,$$

where ϵ_i are independent Gaussian noises with common variance σ^2. It is well known that the least squares estimate $\hat{\boldsymbol{\beta}}$ has covariance $\sigma^2(\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i^T)^{-1}$. In experimental design, given total number of allowable experiments, we want to choose among a list of m candidate design points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ such that the covariance matrix is minimized in some sense. In mathematical terms, we want to find a probability vector $\boldsymbol{p} = (p_1, \ldots, p_m)$ such that $p_i \geq 0$, $\sum_{i=1}^m p_i = 1$, and the matrix $\boldsymbol{V} = \left(\sum_{i=1}^m p_i \boldsymbol{x}_i \boldsymbol{x}_i^T\right)^{-1}$ is "small".

   (a) (D-optimal design) In *D-optimal design*, we minimize the determinant of $\boldsymbol{V}$. That is

$$\begin{array}{ll} \text{minimize} & \det \left(\sum_{i=1}^m p_i \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \\ \text{subject to} & \boldsymbol{p} \succeq \boldsymbol{0}, \ \boldsymbol{1}^T \boldsymbol{p} = 1 \end{array}$$

   for the optimal $\boldsymbol{p}$. Formulate this problem as an SDP.

   (b) (E-optimal design) In *E-optimal design*, we minimize the spectral norm, i.e., the maximum eigenvalue of $\boldsymbol{V}$

$$\begin{array}{ll} \text{minimize} & \lambda_{\max} \left(\sum_{i=1}^m p_i \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \\ \text{subject to} & \boldsymbol{p} \succeq \boldsymbol{0}, \ \boldsymbol{1}^T \boldsymbol{p} = 1. \end{array}$$

   Statistically we are minimizing the maximum variance of $\sum_{j=1}^p a_j \text{var}(\hat{\beta}_j)$ over all vectors $\boldsymbol{a}$ with unit norm. Formulate this problem as an SDP.

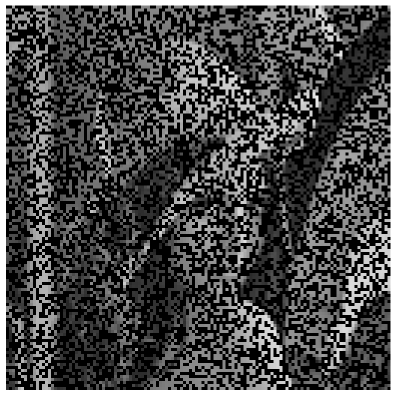
   (c) (A-optimal design) In *A-optimal design*, we minimize the trace of $\boldsymbol{V}$

$$\begin{array}{ll} \text{minimize} & \text{tr} \left(\sum_{i=1}^m p_i \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \\ \text{subject to} & \boldsymbol{p} \succeq \boldsymbol{0}, \ \boldsymbol{1}^T \boldsymbol{p} = 1. \end{array}$$

   Statistically we are minimizing the total variance $\sum_{j=1}^p \text{var}(\hat{\beta}_j)$. Formulate this problem as an SDP.

(d) Consider a 3×4 factorial design with treatment A at levels A1, A2, and A3 and treatment B at levels B1, B2, B3, and B4. It is known that the A3:B4 combination creates undesirable side effects so we omit this design point. Find the optimal D, E, and A designs for this experiment. For ease of grading, please use the dummy coding for factors. Report your solutions in 3×4 tables.

2. (Complete Lena) This time the picture of Lena (`lena128missing.png`) has a lot of missing pixels.



Your goal is to "complete" Lena by a technique called matrix completion (Candès and Recht, 2009; Candès and Tao, 2010). Let $\boldsymbol{Y} = (y_{ij})$ be the gray levels of a 2D image with missing values coded as 0. We complete the matrix $\boldsymbol{Y}$ by solving the optimization problem

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{X}\|_* \\ \text{subject to} & x_{ij} = y_{ij} \text{ for all observed entries } (i,j). \end{array}$$

Here $\|\boldsymbol{M}\|_* = \sum_i \sigma_i(\boldsymbol{M})$ is the nuclear norm. That is we seek the matrix with minimal nuclear norm that agrees with the observed entries. Formulate and solve this problems as an SDP and display your solution.

3. (Best subset regression) We again work on the prostate cancer data in HW4.

(a) Repeat HW4 Q1 parts (a) and (b).

(b) Fit the best subset regression

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\boldsymbol{y} - \beta_0 \boldsymbol{1} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 \\ \text{subject to} & \|\boldsymbol{\beta}\|_0 \le k \end{array}$$

for $k = 0, 1, \ldots, 8$ using MIP on the training data and plot the solution path. Also plot the prediction errors on the test set over k.

4. (Ranking MLB teams) We revisit the sports team ranking problem considered in ST758 (2014 fall) HW8 `http://hua-zhou.github.io/teaching/st758-2014fall/ST758-2014-HW8.pdf`

(a) Retrieve the complete American League 2014 season win-loss data at `http://espn.go.com/mlb/standings/grid/_/year/2014`.

(b) The ranking method described in ST758 HW8 is the classical Bradley-Terry model. Formulate the maximum likelihood estimation (MLE) problem for the Bradley-Terry model as a geometric program (GP) and find the MLE for the 2014 American League data using a convex optimization software. Display the team strength parameters γ_i and the ranking.

(c) Alternatively we model each team's ability by a parameter $a_i \in [0, 1]$, $i = 1, \ldots, p$. Team i beats team j with probability $\mathbf{P}(a_i - a_j > v)$ where v is a standard normal random variable. Formulate the MLE problem for this model as a convex optimization problem. Find the MLE for the 2014 American League data using a convex optimization software. (Hint: `cvx` has a built-in function `log_normcdf`.) Compare the ranking to that by the Bradley-Terry model.

References

Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772.

Candès, E. J. and Tao, T. (2010). The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inf. Theor.*, 56(5):2053–2080.