ECP
LSD
2015 - 2016

Selectivity estimation for DBSM

1 Introduction

In database system management (DBSM), every request formulated by a user can be viewed as an event in a probability space $(\Omega, \mathcal{T}, P)$ where Ω is a finite set having N elements. In oder to optimize request fulfillment, it is useful to estimate accurately the probabilities (also called selectivities) associated with the elements of Ω. To do so, rough estimations of the probabilities of a certain number M of events are available. (These estimations are performed based on a history of formulated requests and some a priori knowledge.)

Let $x \in \mathbb{R}^N$ be the vector of sought probabilities and let $b = (b^{(i)})_{1\le i\le M} \in]0,1]^M$ be the vector of estimated probabilities. The problem is equivalent to

$$\underset{x\in C}{\text{minimize}} \;\; q_b(Ax) \tag{1}$$

where

- $C = \{x = (x^{(i)})_{1\le i\le N} \in [0,1]^N \mid \sum_{i=1}^N x^{(i)} = 1\}$;
- $A \in \{0,1\}^{M\times N}$ is a binary matrix establishing the theoretical link existing between the probabilities of each event and the probabilities of the elements of Ω belonging to it ;
- q_b is the *quotient function* defined as

$$\big(\forall y = (y^{(i)})_{1\le i\le M} \in \mathbb{R}^M\big) \qquad q_b(y) = \sum_{i=1}^{M} \theta(y^{(i)}/b^{(i)}),$$

with

$$(\forall \xi \in \mathbb{R}) \qquad \theta(\xi) = \begin{cases} \xi & \text{if } \xi \ge 1 \\ \xi^{-1} & \text{if } 0 < \xi < 1 \\ +\infty & \text{otherwise.} \end{cases}$$

2 Work to be performed

1. Is the quotient function convex ? lower-semicontinuous ?
2. Does there exist a solution to Problem (1) ?

3. Let $\gamma \in]0, +\infty[$. It can be proved that

$$(\forall \xi \in \mathbb{R}) \qquad \operatorname{prox}_{\gamma\theta}(\xi) = \begin{cases} \zeta & \text{if } \xi < 1 - \gamma \\ 1 & \text{if } \xi \in [1-\gamma, 1+\gamma] \\ \xi - \gamma & \text{if } \xi > 1 + \gamma, \end{cases}$$

   where ζ is the unique solution in $]0, 1]$ of the cubic equation

$$\zeta^3 - \xi\zeta^2 = \gamma.$$

   Deduce the expression of the proximal operator of γq_b, for every $b \in]0, 1[^M$.
4. Write a function for computing this operator.
5. Propose a primal-dual algorithm to solve Problem (1).
   (A code is provided implementing the projection onto C.)
6. Apply the algorithm to the following example :

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 0.2114 \\ 0.6331 \\ 0.6312 \\ 0.5182 \\ 0.9337 \\ 0.0035 \end{pmatrix}.$$

7. Would it be possible to apply ADMM on this example ?