Optimization for Data Science
September 4, 2018

1. (6 POINTS) Describe in depth sample average approximation and stochastic gradient method. Furthermore, focus on empirical risk minimization, and explain why stochastic gradient is well-suited in this context.

2. (7 POINTS) Describe in depth the Alternating Direction Method of Multipliers and its application for the solution of global consensus problems.

3. (7 POINTS) Suppose a company has m warehouses and n retail outlets. Products are to be shipped from the warehouses to the outlets. Each warehouse i has a given level of supply a_i, and each outlet j has a given level of demand b_j. We are also given the transportation costs between every pair of warehouse i and outlet j, and these costs are linear (i.e., unitary cost is c_{ij}). Describe the linear programming problem to determine an optimal transportation scheme between the warehouses and the outlets, subject to the specified supply and demand constraints. Furthermore, build up the related dual and try to give an economic interpretation of the problem.

4. (8 POINTS) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function with Lipschitz continuous gradient having Lipschitz constant $L > 0$. Denote $g_l = \|\nabla f(x_l)\|$ and define

   $$g_k^* = \min_{1 \le l \le k} g_l.$$

   Prove that the gradient method with constant stepsize $\alpha_k = \frac{1}{L}$ satisfies the following inequality:

   $$g_k^* \le \sqrt{\frac{2L\,(f(x_1) - f^*)}{k}},$$

   with f^* minimum for $f(x)$.

5. (8 POINTS) Consider the problem of projecting a point $v \in \mathbb{R}^n$ over the ℓ_1-ball:

   $$\begin{array}{ll} \min\limits_{w \in \mathbb{R}^n} & \|w - v\|_2^2 \\ \text{s.t.} & \|w\|_1 \le r \end{array} \tag{1}$$

   with $r > 0$. Let w be an optimal solution of Problem (1). Prove that for all $i = 1, \dots, n$, we have $w_i v_i \ge 0$.
   Taking into account the theoretical result described above, and assuming that an efficient procedure for projecting over the simplex

   $$\Delta = \{w \in \mathbb{R}^n : \; e^\top w = r, \; w \ge 0\}$$

   is available, describe a method for efficiently projecting over the ℓ_1-ball.