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.

   Solution 1. See Notes Section 4.7.

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

   Solution 2. See Notes Section 5.8.

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.

   Solution 3. The decision variables x_{ij} (all non-negative) represent the size of the shipment from warehouse i to outlet j. Consider the shipment from warehouse i to outlet j. For any pair i, j, the transportation cost per unit is c_{ij} and the size of the shipment is x_{ij} . Since we assume that the cost function is linear, the total cost of this shipment is given by $c_{ij} \cdot x_{ij}$. Summing over all i and all j we get the overall transportation cost for all warehouse-outlet combinations. That is, our objective function is:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}.$$

   Consider warehouse i. The total outgoing shipment from this warehouse is the sum $\sum_{j=1}^{n} x_{ij}$. Since the total supply from warehouse i is a_i, the total outgoing shipment cannot exceed a_i. That is, we must require

$$\sum_{j=1}^{n} x_{ij} \leq a_i.$$

   Consider outlet j. The total incoming shipment at this outlet is the sum $\sum_{i=1}^{m} x_{ij}$. Since the demand at outlet j is b_j, the total incoming shipment should not be less than b_j. That is, we must require

$$\sum_{i=1}^{m} x_{ij} \geq b_j.$$

   The final linear programming problem we get is the following:

$$\begin{array}{lll} \min & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} & \\ \text{s.t.} & \sum_{j=1}^{n} x_{ij} \leq a_i, & i = 1, \ldots, m \\ & \sum_{i=1}^{m} x_{ij} \geq b_j, & j = 1, \ldots, n \\ & x_{ij} \geq 0, & i = 1, \ldots, m,\ j = 1, \ldots, n. \end{array}$$

In order to get the dual, we rewrite the problem as follows:

$$\begin{array}{lll} \min & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} & \\ \text{s.t.} & -\sum_{j=1}^{n} x_{ij} \geq -a_i, & i = 1,\dots,m \\ & \sum_{i=1}^{m} x_{ij} \geq b_j, & j = 1,\dots,n \\ & x_{ij} \geq 0, & i = 1,\dots,m,\ j = 1,\dots,n. \end{array}$$

Introducing dual variables u_i ($i = 1,\dots,m$) for the first group of constraints and v_j ($j = 1,\dots,n$) for the second group of constraints, we can write the problem as follows:

$$\begin{array}{lll} \max & -\sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j & \\ \text{s.t.} & -u_i + v_j \leq c_{ij}, & i = 1,\dots,m,\ j = 1,\dots,n \\ & u_i \geq 0, & i = 1,\dots,m \\ & v_j \geq 0, & j = 1,\dots,n. \end{array}$$

The shadow market is represented by a new agent that wants to buy all products in our warehouses and sell them back to us at the outlets (i.e., agent takes care of shipping goods). The unitary price for buying products at the warehouse i is u_i, The unitary price for selling products at outlet j is v_j. Dual constraints ensure that offer is appealing to us: for each warehouse i and outlet j, the difference between the unitary selling price related to warehouse i and the unitary buying price for outlet j cannot exceed the unit cost c_{ij}. The objective function models agent's goal, that is maximizing the profit.

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

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

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

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

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

Solution 4. Taking into account Remark 4.12 we can write

$$f(x_{k+1}) - f(x_k) \leq -\frac{1}{2L}\|\nabla f(x_k)\|^2.$$

Summing up these inequalities for iterations $1,\dots,k$, and considering definition of g_k^*, we can write:

$$k \cdot (g_k^*)^2 \leq \sum_{i=1}^{k} \|\nabla f(x_i)\|^2 \leq 2L\left(f(x_1) - f(x_{k+1})\right) \leq 2L\left(f(x_1) - f^*\right).$$

Thus getting our bound:

$$g_k^* \leq \sqrt{\frac{2L\left(f(x_1) - f^*\right)}{k}}.$$

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

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

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

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

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

Solution 5. We first prove the theoretical result. Assume by contradiction that the claim does not hold. Thus, we have an index i such that

$$w_i v_i < 0.$$

Let $\tilde{w}$ be a vector such that $\tilde{w}_i = 0$ and for all $j \neq i$ we have $\tilde{w}_j = w_j$. We hence get that

$$\|\tilde{w}\|_1 = \|w\|_1 - |w_i| \leq r,$$

and $\tilde{w}$ is feasible for the problem. We further can write

$$\|w - v\|_2^2 - \|\tilde{w} - v\|_2^2 = (w_i - v_i)^2 - (0 - v_i)^2 = w_i^2 - 2w_i v_i > w_i^2 > 0,$$

thus contradicting optimality of w and proving our result. Defining projection over the ℓ_1-ball is then pretty straightforward. We call $P_\Delta(v)$, $P_{\ell_1}(v)$ the projection of vector v over the unit simplex and ℓ_1-ball, respectively. We have that

$$P_{\ell_1}(v) = \begin{cases} v & \text{if } \|v\|_1 \leq r \\ \text{sgn}(v) \cdot x & \text{otherwise,} \end{cases}$$

where the product $\cdot$ is intended component-wise and $x = P_\Delta(|v|)$.