Optimization of a Locally Convex Objective on Convex Regions

Jason D. M. Rennie
jrennie@csail.mit.edu

March 11, 2005

Abstract

We consider the problem of maximizing an objective that is locally convex on a set of convex regions. That is, within each convex region, the objective is convex, but across regions, the objective may not be convex. We cannot say anything about how the objective behaves across regions. We formulate the problem and give an algorithm for finding a local maximum of the objective.

1 Definitions

- Let $W \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R}^d$ be indexed by examples, then classes, then components. W is assumed fixed and known.
- Let $\vec{y} = \{y_1, \ldots, y_n\}$, $y_i \in \{1, \ldots, m\}$, be the class labels of the examples.
- Let $\vec{x} \in \mathbb{R}^d$ be the parameter vector.
- Let $Z_i(\vec{x}) = \sum_{j=1}^m \max_{k \in \{1,\ldots,l\}} \exp\left(W_{ijk}^T \vec{x}\right)$. Our objective is

$$J(\vec{x}) = \prod_{i=1}^n \frac{1}{Z_i(\vec{x})} \exp\left(\max_{k \in \{1,\ldots,l\}} W_{iy_ik}^T \vec{x}\right) \tag{1}$$

$$\log J(\vec{x}) = \sum_{i=1}^n \max_{k \in \{1,\ldots,l\}} W_{iy_ik}^T \vec{x} - \log Z_i(\vec{x}) \tag{2}$$

- Let $R \in \{1, \ldots, l\}^{n \times m}$ specify a local objective. Define $P(\vec{x}; i, j, R) = \frac{1}{Z_i(\vec{x};R)} \exp\left(W_{ijR_{ij}}^T \vec{x}\right)$, where $Z_i(\vec{x}; R) = \sum_{j=1}^m \exp\left(W_{ijR_{ij}}^T \vec{x}\right)$. Then, our local objective is

$$J_R(\vec{x}) = \prod_i P(\vec{x}; i, y_i, R). \tag{3}$$

R_{ij} is the active component for example i, label j. We abuse notation by also using R to define a region. The region R is defined as follows: $R = \{\vec{x} | J_R(\vec{x}) = J(\vec{x})\}$. Note that a point may belong to multiple regions.

- Let $f_R(\alpha; \vec{x}, \vec{d}) = \log J_R(\vec{x} + \vec{d}\alpha)$.

$$\frac{\partial f_R}{\partial \alpha} = \sum_{i=1}^{n} \left(\vec{d}^T W_{iy_i R_{iy_i}} - \sum_{j=1}^{m} d^T W_{ijR_{ij}} P(\vec{x}; i, j, R) \right) \tag{4}$$

To find a local minimum, we choose a direction for which the objective is non-decreasing, and follow that direction to a local maximum or a change in the set of active regions.

2 Function Definitions

Here we define some useful subroutines.

- LineSearch$(R, \vec{d}, \vec{x}, \alpha_0, \alpha_1, \epsilon)$: Assume $\alpha_0 < \alpha_1$. Assume $\vec{x} + \vec{d}[\alpha_0, \alpha_1] \in R$. Let $\alpha_2 = (\alpha_0 + \alpha_1)/2$. Find the maximum of J_R along the interval $\vec{x} + \vec{d}[\alpha_0, \alpha_1]$ using bisection search. Define $f(\alpha) = J_R(\vec{x} + \alpha\vec{d})$. If $\alpha_1 - \alpha_0 < \epsilon$ or $\frac{\partial f_R}{\partial \alpha} = 0$, Return α_2. If $\frac{\partial f_R}{\partial \alpha} > 0$, Return LineSearch$(R, \vec{d}, \vec{x}, \alpha_2, \alpha_1, s, \epsilon)$. If $\frac{\partial f_R}{\partial \alpha} < 0$, Return LineSearch$(R, \vec{d}, \vec{x}, \alpha_0, \alpha_2, s, \epsilon)$.

- Function NextConstraint$(R, \vec{d}, \vec{x})$: Finds the smallest $\alpha \geq 0$ such that $\vec{x} + \alpha\vec{d} \notin R$. For all i, j and $k \neq R_{ij}$, such that $\vec{d}^T W_{ijk} > d^T W_{ijR_{ij}}$, let

$$\alpha^*_{ijk} = -\frac{\vec{x}^T (W_{ijR_{ij}} - W_{ijk})}{\vec{d}^T (W_{ijR_{ij}} - W_{ijk})}. \tag{5}$$

  Return $\min \alpha^*_{ijk}$.

- Function DirectionSearch$(\vec{x})$: Determine a direction in which to search. Let $\mathcal{R} = \text{ActiveRegions}(\vec{x})$. For $i = \{1, \ldots, |\mathcal{R}|\}$, let $\vec{d}_i = \frac{\partial f_{\mathcal{R}_i}}{\partial \alpha}$, let $\alpha_i =$ NextConstraint$(R, \vec{d}_i, \vec{x})$. If $\alpha_i > 0$ for some i, then Return $\arg\max_{\vec{d} \in \{d_1, \ldots, d_{|\mathcal{R}|}\}} J_R(\alpha_i \vec{d})$. Else, project one of the gradients onto the space of the active regions. Let $\vec{d}$ be one of the gradients. Consider the case that there are two active regions, $R^{(1)}$ and $R^{(2)}$. Also, assume that $R^{(1)}_{ij} = R^{(2)}_{ij}$ $\forall i, j$ except for i', j'. Let $r_1 = R^{(1)}_{i'j'}$ and $r_2 = R^{(1)}_{i'j'}$. Hence, the space of the two active regions is (at least locally) defined by $W_{i'j'r_1} = W_{i'j'r_2}$. Let $\vec{w}_1 = W_{i'j'r_1}$ and $\vec{w}_2 = W_{i'j'r_2}$. Then, we want to find the unit-length vector, $\vec{x}$, that maximizes $\vec{x}^T \vec{d}$ such that $\vec{x}^T \vec{w}_1 = \vec{x}^T \vec{w}_2$. This is simply $\vec{d}$ minus the projection of $\vec{d}$ onto $(\vec{w}_1 - \vec{w}_2)$. I.e.

$$\vec{x} = \vec{d} - \left(\frac{\vec{d}^T (\vec{w}_1 - \vec{w}_2)}{\|\vec{w}_1 - \vec{w}_2\|} \right)^T (\vec{w}_1 - \vec{w}_2) \tag{6}$$

For more than two regions, we simply subtract out projections of a set of pairs.

- Function ActiveRegions($\vec{x}$): Return the set of regions for $\vec{x}$. For each i, j, find $\max_k \vec{x}^T W_{ijk}$.

Need additional constraint ($\leq$ constraint according to region boundaries). Otherwise, there is no max/min. But, I don't know the direction... And, I can't add inequality constraints that are not independent of the equality constraints.

We thank John Barnett for valuable comments and discussion.