Multiplicative Updates for Nonnegative Quadratic Programming

Fei Sha,Yuanqing Lin,Lawrence K.Saul Daniel D.Lee(2007)

Feiyun Wang

October 25, 2012

Table of contents

Non-negative Quadratic Programming

$$\min_{v} F(v) = \frac{1}{2}v^T A v + b^T v \\ s.t.\ v \geq 0 \tag{1}$$

The constraint indicates that the variable v is confined to the non-negative orthant. Assuming that the matrix A is symmetric and positive definite(may have negative elements off the diagonal),so that the objective function $F(v)$ is convex having one global minimum and no local minima in particular.

- Presenting the multiplicative updates for the basic problem of NQP in eq.(1)
- Extending the multiplicative updates with upper-bound constraints $v \leq l$

Splitting Matrix A

Expressing the multiplicative updates for NQP in terms of the positive and negative components of the matrix A.

- Positive components of A

$$A_{ij}^{+} = \begin{cases} A_{ij} & \text{if } A_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- Negative components of A

$$A_{ij}^{-} = \begin{cases} \|A_{ij}\| & \text{if } A_{ij} < 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

It follows that $A = A^{+} - A^{-}$.

Decomposing the Objective Function

- Combination of three terms.

$$F(v) = F_a(v) + F_b(V) - F_c(v) \tag{4}$$

- The first and third terms split the quadratic pieces of $F(v)$ and the second term captures the linear piece:

$$\begin{aligned} F_a(v) &= \frac{1}{2} v^T A^+ v, \\ F_b(v) &= b^T v \\ F_c(v) &= \frac{1}{2} v^T A^- v. \end{aligned} \tag{5}$$

Corresponding Derivatives

- Derivative of $F_a(v)$

$$a_i = \frac{\partial F_a}{\partial v_i} = (A^+ v)_i \tag{6}$$

- Derivative of $F_b(v)$

$$b_i = \frac{\partial F_b}{\partial v_i} \tag{7}$$

- Derivative of $F_c(v)$

$$c_i = \frac{\partial F_c}{\partial v_i} = (A^- v)_i \tag{8}$$

Multiplicative Updates

Expressing the multiplicative updates in terms of partial derivatives

$$v_i \longleftarrow \left[\frac{-b_i + \sqrt{b_i^2 + 4a_i c_i}}{2a_i}\right] \tag{9}$$

Note that the partial derivatives a_i and b_i are guaranteed to be non-negative when evaluated the vectors v in the non-negative orthant,so that the optimization remains confined to the feasible region for NQP.

Upper-Bound Constraints

Extending the multiplicative updates in Eq.(9) to incorporate upper-bound constraints of the form $v \leq l$.A sample way of enforcing such constraints is to clip the output of the updates in Eq.(10):

$$v_i \longleftarrow \min \left\{ l_i, \left[\frac{-b_i + \sqrt{b_i^2 + 4a_i c_i}}{2a_i} \right] v_i \right\} \tag{10}$$

Analysis of Auxiliary Function

The auxiliary function $G(v, v_t)$ for the objective function in Eq.(1) has two critical properties:

1. $F(v) \leq G(v, v^t)$
2. $F(v) = G(v, v)$

We can derive the update rule $v^{t+1} = arg \min_v G(v, v^t)$,which never increases the objective function $F(v)$:

$$F(v^{t+1}) \leq G(v^{t+1}, v^t) \leq G(v^t, v^t) \leq F(v^t) \tag{11}$$

Constructing Auxiliary Functions

- Auxiliary Function for $F_a(v)$

$$G_a(v, v^t) = \frac{1}{2}\sum_i \frac{(A^+ v^t)}{v_i^t} v_i^2 \tag{12}$$

$$F_a(v) \leq G_a(v, v^t) \tag{13}$$

- Auxiliary Function for $-F_c(v)$

$$G_c(v, v^t) = -\frac{1}{2}\sum_{ij} A_{ij}^- v_i v_j (1 + \log \frac{v_i v_j}{v_i^t v_j^t} \tag{14}$$

$$-F_c(v) \leq G_c(v, v^t) \tag{15}$$

Combination of two Auxiliary Function

Combining the auxiliary functions $G_a(v, v^t)$ and $G_c(v, v^t)$ to generate the final auxiliary function:

$$
\begin{aligned}
G(v, v^t) &= G_a(v, v^t) + G_c(v, v^t) + F_b(v) \\
&= \frac{1}{2}\sum_i \frac{(A^+ v^t)_i}{v_i^t} v_i^2 - \frac{1}{2}\sum_{ij} A_{ij}^- v_i^t v_j^t (1 + \log \frac{v_i v_j}{v_i^t v_j^t}) + \sum_i b_i v_i \\
&= \frac{1}{2}\sum_i \frac{(A^+ v^t)_i}{v_i^t} v_i^2 - \sum_i (A^- v^t)_i v_i^t \log \frac{v_i}{v_i^t} \\
&\quad + \sum_i b_i v_i - \frac{1}{2}\sum A_{ij}^- v_i^t v_j^t
\end{aligned}
\tag{16}
$$

Then $G(v, v^t)$ is an auxiliary function for $F(v)$ satisfying $F(v) \leq G(v, v^t)$ and $F(v) = g(v, v)$.

Decomposing Auxiliary Function

1. We identify the auxiliary function with

$$G(v, v^t) = \sum_i G_i(v_i) - \frac{1}{2} v^T A^- v \tag{17}$$

$$G_i(v_i) = \frac{1}{2} \frac{(A^+ v)_i}{v_i^t} v_i^2 - (A^- v)_i v_i \log \frac{v_i}{v_i^t} + b_i v_i \tag{18}$$

   where $G_i(v_i)$ is a single-variable function of v_i.
2. Note that the minimizer of $G(v, v^t)$ can be easily found by minimizing each $G_i(v_i)$ separately: $v_i = arg \min_{v_i} G_i(v_i)$.
3. Besides, $G_i(v_i)$ is strictly convex in v_i:

$$G_i^{''}(v_i) = \frac{(A^+ v^t)_i}{v_i^t} + \frac{(A^- v^t)_i}{v_i^2} v_i^t \geq 0. \tag{19}$$

Locality Sensitive Discriminant Analysis(LSDA)

We only consider the problem of mapping all the data points $X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m)^T$ to a line.Here,we get a set of points $\mathbf{y} = (y_1, y_2, \cdots, y_m)^T$.

- Objective function of within-class graph:

$$\min \sum_{ij} (y_i - y_j)^2 W_{w,ij} \tag{20}$$

- Objective function of between-class graph:

$$\max \sum_{ij} (y_i - y_j)^2 W_{b,ij} \tag{21}$$

What do we want to do?

Adding sparseness and non-negativity to LSDA and applying them in hashing.

- Objective function of sparseness:

$$\min \left\|\mathbf{y}\right\|_1 \tag{22}$$

- Constraint C1: $y^T D_b y = 1$
- Constraint C2: $\forall i, y_i \geq 0$

Rewriting the First Objective Function

By simple algebra formulation,the objective function (20) can be reduced to

$$
\begin{aligned}
&\frac{1}{2}\sum_{ij}(y_i - y_j)^2 W_{w,ij} \\
&= \frac{1}{2}\sum_{ij}(y_i^2 + y_j^2 - 2y_i y_j) W_{w,ij} \\
&= \sum_{i} y_i^2 W_{w,ij} - \sum_{ij} y_i y_j W_{w,ij} \\
&= y^T D_w y - y^T W_w y \\
&= y^T L_w y
\end{aligned}
\tag{23}
$$

Reorganizing the Expression of the Second Objective Function

Similarly,the objective function (21) can be reduced to

$$\begin{aligned}
&\frac{1}{2}\sum_{ij}(y_i - y_j)^2 W_{b,ij} \\
&= \frac{1}{2}\sum_{ij}(y_i^2 + y_j^2 - 2y_i y_j) W_{b,ij} \\
&= \sum_{i} y_i^2 W_{b,ij} - \sum_{ij} y_i y_j W_{b,ij} \\
&= y^T D_b y - y^T W_b y
\end{aligned} \tag{24}$$

Simplifying the Objective Functions above

- The objective function (20) is equivalent to:

$$\min_{y} y^T L_w y \tag{25}$$

- Under the constraint C1,the objective function (21) becomes the following:

$$\min_{y} y^T W_b y \tag{26}$$

- And the objective function (22) can be rewritten as follows under the constrain C2:

$$\min_{y} by. \tag{27}$$

where b is a row vector with all ones.

Final Objective Function

Combining the objective functions (25),(26),(27) and the constraints together,we get the final objective function:

$$\begin{aligned}\min_{y} F(y) &= \frac{1}{2}y^T(\alpha L_w + (1-\alpha)W_b)y + \frac{\eta}{2}(1 - y^T D_b y) + \lambda by \\ &= \frac{1}{2}y^T(\alpha L_w + (1-\alpha)W_b - \eta D_b)y + \frac{\eta}{2} + \lambda by\end{aligned} \tag{28}$$

Constraint 1: $y \geq 0$
Constraint 2: $0 \leq \alpha \leq 1$
Constraint 3: $\eta \geq 0$
Constraint 4: $\lambda \geq 0$

Extracting and Splitting the Quadratic Parameter of $F(y)$

- Quadratic Parameter of $F(y)$:

$$A = \alpha L_w + (1-\alpha) W_b - \eta D_b \tag{29}$$

- Positive Components of A:

$$A^+ = \alpha D_w + (1-\alpha) W_b \tag{30}$$

- Negative Components of A:

$$A^- = \alpha W_w + \eta D_b \tag{31}$$

Decomposing the Objective Function $F(y)$

- Positive Quadratic Piece of $F(y)$:

$$F_a(y) = \frac{1}{2} y^T A^+ y \tag{32}$$

- Linear Piece of $F(y)$:

$$F_b(y) = \lambda b y + \frac{\eta}{2} \tag{33}$$

- Negative Quadratic Piece of $F(y)$:

$$F_c(y) = \frac{1}{2} y^T A^- y \tag{34}$$

Constructing Auxiliary Functions

- Auxiliary Function for $F_a(y)$:

$$G_a(y, y^t) = \frac{1}{2} \sum_i \frac{(A^+ y^t)_i}{y_i^t} y_i^2 \tag{35}$$

$$F_a(y) \leq G_a(y, y^t) \tag{36}$$

- Auxiliary Function for $-F_c(y)$:

$$G_c(y) = -\frac{1}{2} \sum_{ij} A_{ij}^- y_i^t y_j^t (1 + \log \frac{y_i y_j}{y_i^t y_j^t}) \tag{37}$$

$$-F_c(y) \leq G_c(y) \tag{38}$$

- Generating Global Auxiliary Function $G(y, y^t)$ for $F(y)$:

$$
\begin{aligned}
G(y, y^t) &= G_a(y, y^t) + G_c(y, y^t) + F_b(y) \\
&= \frac{1}{2}\sum_i \frac{(A^+ y^t)_i}{y_i^t} y_i^2 - \frac{1}{2}\sum_{ij} A_{ij}^- y_i^t y_j^t (1 + \log \frac{y_i y_j}{y_i^t y_j^t}) \\
&\quad + \lambda \sum_i y_i + \frac{\eta}{2} \\
&= \frac{1}{2}\sum_i \frac{(A^+ y^t)_i}{y_i^t} y_i^2 - \sum_i (A^- y^t)_i y_i^t \log \frac{y_i}{y_i^t} - \frac{1}{2}\sum_{ij} A_{ij}^- y_i^t y_j^t \\
&\quad + \lambda \sum_i y_i + \frac{\eta}{2}
\end{aligned}
\tag{39}
$$

- Having the following properties:

$$
F(y^{t+1}) \leq G(y^{t+1}, y^t) \leq G(y^t, y^t) = F(y^t) \tag{40}
$$

Splitting Auxiliary Function

- The Auxiliary Function can be reduced to:

$$G(y, y^t) = \sum_i G_i(y_i) - \frac{1}{2}\sum_{ij} A^-_{ij} y^t_i y^t_j + \frac{\eta}{2} \tag{41}$$

- Picking Up the Single-Variable of y_i from $G(y, y^t)$

$$G_i(y_i) = \frac{1}{2}\frac{(A^+ y^t)_i}{y^t_i} y_i^2 - (A^- y^t)_i y^t_i \log \frac{y_i}{y^t_i} + \lambda y_i \tag{42}$$

First and Second Derivative of $G_i(y_i)$

- The First Derivative of $G_i(y_i)$:

$$G_i^{'}(y_i) = \frac{(A^+y^t)_i}{y_i^t}y_i - \frac{(A^-y^t)_i}{y_i^t}y_i + \lambda \tag{43}$$

- The Second Derivative of $G_i(y_i)$:

$$G_i^{''}(y_i) = \frac{(A^+y^t)_i}{y_i^t} + \frac{(A^-y^t)_i}{y_i^2}y_i^t \geq 0 \tag{44}$$

So that $G_i(y_i)$ is convex in y_i.

Deriving Update Rule

- Minimizing each $G_i(y_i)$ separately so as to find the minimizer of $G(y, y^t)$.Simply setting the first derivative of $G_i(y_i)$ to zero:

$$G_i^{'}(y_i) = 0 \Longrightarrow (A^+ y^t)_i y_i^2 + \lambda y_i^t y_i - (A^- y^t)_i {y_i^t}^2 = 0 \quad (45)$$

- Solving the quadratic formula Eq.(45) to get the update rule:

$$y_i = \frac{-\lambda y_i^t + \sqrt{(\lambda y_i^t)^2 + 4(A^+ y^t)_i (A^- y^t)_i {y_i^t}^2}}{2(A^+ y^t)_i} \quad (46)$$