Homework 4

Tao Hu
Department of Computer Science
Peking University
No.5 Yiheyuan Road Haidian District, Beijing, P.R.China
taohu@pku.edu.cn

1 Problem ScreenShot

Homework 6 for "Algorithms for Big-Data Analysis"

May 11, 2017

Note: Please write up your solutions independently. If you get significant help from others, write down the source of references. A formal mathematical proof for all your claims is required.

1. Given a $n \times p$ matrix X. Derive the optimal solution for the following problem:

$$\min_{Z,V} \|X - ZV\|_F^2, \text{ s.t. } \quad V^T V = I, Z^T 1 = 0,$$

   where Z is a $n \times q$ matrix and V is a $q \times p$ matrix.

2. Derive the dual optimization problem for

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|w\|_2^2 + C_1 \sum_{i=1}^{n} \xi_i + C_2 \sum_{i=1}^{n} \xi_i^2 \\ \text{s.t.} \quad & y_i \cdot (x_i \cdot w + b) \geq 1 - \xi_i, \forall i = 1, \ldots, n \\ & \xi_i \geq 0, \forall i = 1, \ldots, n \end{aligned}$$

3. Properties of Submodular Functions

   (a) Prove that any non-negative submodular function is also subadditive, i.e. if $F : 2^X \to \mathbb{R}_+$ is submodular then $F(S \cup T) \leq F(S) + F(T)$ for any $S, T \subseteq X$. Here, $\mathbb{R}_+ = \{x \in \mathbb{R}, x \geq 0\}$.

   (b) Prove that a function $F : 2^X \to \mathbb{R}_+$ is submodular if and only if for any $S, T \subseteq X$, the marginal contribution function $F_S(T) = F(S \cup T) - F(S)$ is subadditive.

4. Given finite ground set X , and given $w_d \in [0, 1]$ for all $d \in X$, define

$$F(S) = \Pi_{d \in S} w_d,$$

   where $F(\emptyset) = 1$. Is this submodular, supermodular, modular, or neither?

Author Info: Taohu ,Peking University

2 Solution

2.1 Solution 1

2.2 Solution 2

Firstly,Let's recall the soft SVM[1] problem:

$$min_{w,b,\xi}\frac{1}{2}||w||_2^2 + C\sum_{i=1}^n \xi_i$$
$$S.t \quad y_i(wx_i+b) \geq 1-\xi_i$$
$$\xi_i \geq 0$$
$$i = 1,...,n$$

let's conduct how to acquire the dual form of the soft-threshold SVM is :

Lagrangian polynomials is :

$$L(w,b,\xi,\lambda,\mu) = \frac{1}{2}||w||_2^2 + C\sum_{i=1}^n \xi_i - \sum_{i=1}^n \lambda_i[y_i(wx_i+b) - 1 + \xi_i] - \sum_{i=1}^n \mu_i\xi_i$$

obtain derivation for w,b,ξ_i

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \lambda_i y_i x_i^T$$
$$\frac{\partial L}{\partial b} = -\sum_{i=1}^n \lambda_i y_i = 0$$
$$\frac{\partial L}{\partial \xi_i} = C - \lambda_i - \mu_i = 0$$

fuse the three formula into the Lagrangian polynomial:

$$L(w,b,\xi,\lambda,\mu) = \frac{1}{2}||\sum_{i=1}^n \lambda_i y_i x_i^T||_2^2 + C\sum_{i=1}^n \xi_i - \sum_{i=1}^n(\lambda_i y_i x_i(\sum_{j=1}^n \lambda_j y_j x_j^T) + \lambda_i y_i b - \lambda_i + \lambda_i\xi_i) - \sum_{i=1}^n \mu_i\xi_i$$

here we have to notice that:

$$||\sum_{i=1}^n \lambda_i y_i x_i^T||_2^2 = \sum_{i=1}^n(\lambda_i y_i x_i(\sum_{j=1}^n \lambda_j y_j x_j^T))$$

so based on the above result ,just neaten the formula:

$$L(w,b,\xi,\lambda,\mu) = \frac{1}{2}\sum_{i=1}^n(\lambda_i y_i x_i(\sum_{j=1}^n \lambda_j y_j x_j^T)) - \sum_{i=1}^n \lambda_i$$

so,the dual problem of the soft-threshold SVM is :

$$min_{\lambda_i,\mu_i}\frac{1}{2}\sum_{i=1}^n(\lambda_i y_i x_i(\sum_{j=1}^n \lambda_j y_j x_j^T)) + \sum_{i=1}^n \lambda_i$$
$$S.t \quad \lambda_i \geq 0$$
$$\sum_{i=1}^n \lambda_i y_i = 0$$
$$C - \lambda_i - \mu_i = 0$$
$$\mu_i \geq 0$$

Let's back to the original problem. The Lagrangian problem is:

$$L(w,b,\xi,\lambda,\mu) = \frac{1}{2}||w||_2^2 + C_1\sum_{i=1}^n \xi_i + C_2\sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \lambda_i[y_i(x_i w + b) - 1 + \xi_i]\sum_{i=1}^n \mu_i\xi_i$$

acquire the derivation of w, b, ξ_i

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \lambda_i y_i x_i^T = 0$$
$$\frac{\partial L}{\partial b} = -\sum_{i=1}^n \lambda_i y_i$$
$$\frac{\partial L}{\partial \xi_i} = C_1 + 2C_2\xi_i - \lambda_i - \mu_i = 0$$

further reduction based on the derivation above:

$$L(w,b,\xi,\lambda,\mu) = \frac{1}{2}||\sum_{i=1}^2 \lambda_i y_i x_i||_2^2 + C_1\sum_{i=1}^n \xi_i + C_2\sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \lambda_i y_i x_i(\sum_{j=1}^n \lambda_j y_j x_j^T) - \sum_{i=1}^n \lambda_i y_i b + \sum_{i=1}^n(\lambda_i - \lambda_i\xi_i - \mu_i\xi_i)$$

$$= -\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j y_i y_j x_i^T x_j + C_1\sum_{i=1}^n \xi_i + C_2\sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n[\lambda_i - (C_1 + 2C_2\xi_i)\xi_i]$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j y_i y_j x_i^T x_j + \sum_{i=1}^{n}\lambda_i - \sum_{i=1}^{n} C_2\xi_i^2$$

$$= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j y_i y_j x_i^T x_j + \sum_{i=1}^{n}\lambda_i - \sum_{i=1}^{n}\frac{(\lambda_i+\mu_i-C_1)^2}{4C_2}$$

so the dual problem of the original problem is:

$$\boxed{\begin{aligned} max_{\lambda_i,\mu_i} & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j y_i y_j x_i^T x_j + \sum_{i=1}^{n}\lambda_i - \sum_{i=1}^{n}\frac{(\lambda_i+\mu_i-C_1)^2}{4C_2} \\ & S.t \quad \lambda_i,\mu_i \geq 0 \\ & \sum_{i=1}^{n}\lambda_i y_i = 0 \\ & \lambda_i+\mu_i-C_1 \geq 0 \end{aligned}}$$

2.3 Solution 3

(a).from the definition of the submodular :

$$F(S\cap T) + F(S\cup T) \leq F(S) + F(T)$$

because it's non-negative submodular,so $F(S\cap T)\geq 0$,further induction:

$$F(S\cup T) \leq F(S) + F(T)$$

(b).

$$\begin{aligned} & F_S(A\cup B) \leq F_S(A) + F_S(B) \\ \Leftrightarrow & F(A\cup B\cup S) - F(S) \leq F(A\cup S) - F(S) + F(B\cup S) - F(S) \\ \Leftrightarrow & F(A\cup B\cup S) + F(S) \leq F(B\cup S) + F(A\cup S) \end{aligned}$$

according to submodule definition:

$$F(B\cup S) + F(A\cup S) \geq F(A\cup B\cup S) + F(S\cup(A\cap B))$$

however, the relationship of $F(S\cup(A\cap B))$ and $F(S)$ is not determined since F may not be monotone increasing. so the conclusion is not correct.

2.4 Solution 4

it's a supermodule. Let

$a = \{F(M)|M = A - B\}$,
$c = \{F(M)|M = B - A\}$,
$b = \{F(M)|M = A\cap B\}$

then we have:

$$\begin{aligned} & F(A) + F(B) \leq F(A\cup B) + F(A\cap B) \\ & ab + bc \leq b + abc \\ & (1-c)(1-a) \geq 0 \end{aligned}$$

because every step above can be induced bottom-to-up. and $F(\phi) = 1$,therefore it's a supermodule according to definition.

Acknowledgments

References

[1] Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press, 2001.