
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:

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

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

2. Derive the dual optimization problem for

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

3. Properties of Submodular Functions

- (a) Prove that any non-negative submodular function is also subadditive, i.e. if $F : 2^X \rightarrow \mathbb{R}_+$ 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 $F : 2^X \rightarrow \mathbb{R}_+$ 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

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

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

2 Solution

2.1 Solution 1

2.2 Solution 2

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

$$\begin{aligned} \min_{w,b,\xi} & \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \xi_i \\ \text{S.t.} & y_i(w x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned}$$

$i = 1, \dots, n$

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

Lagrangian polynomials is :

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

obtain derivation for w, b, ξ_i

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

fuse the three formula into the Lagrangian polynomial:

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

here we have to notice that:

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

so based on the above result ,just neatent 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 :

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

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

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

further reduction based on the derivation above:

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

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

so the dual problem of the original problem is:

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

2.3 Solution 3

2.4 Solution 4

Acknowledgments

References

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