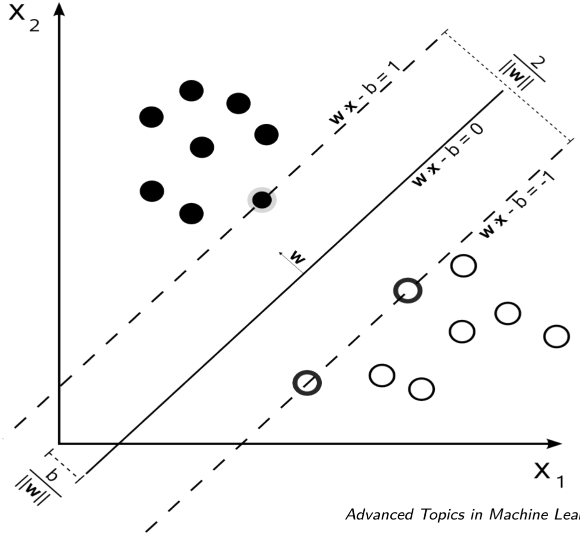
Primal estimated sub-gradient solver for SVM

Lei Zhong

Advanced Topics in Machine Learning

Nov. 4, 2014

Motivating example



Support Vector Machine

Definition

In machine learning many models to do classification and regression analysis are of the following form. Given a training set $(\boldsymbol{x}_i, y_i)_{i=1}^n$ with $\boldsymbol{x}_i \in \mathbb{R}^d$, $y_i \in \{-1, +1\}$, learning is formulated as the task of minimizing the following objective function:

$$f(\boldsymbol{w}) := \frac{1}{n}\sum_{i=1}^{n} L_i(\boldsymbol{w}) + \frac{\lambda}{2}\|\boldsymbol{w}\|^2$$

Loss Function and Subgradient

Definition

- Loss: $L_i := \ell(\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle, y_i)$
- Subgradient: $l'(\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle, y_i)$

Use the notation $z = \langle \boldsymbol{w}, \boldsymbol{x}_i \rangle$, sample loss functions:

Loss function	Subgradient
$l(z, y_i) = \max\{0, 1 - y_i z\}$	ss
$l(z, y_i) = \log(1 + e^{-y_i z})$	1
$l(z, y_i) = \max\{0, \lvert y_i - z \rvert - \epsilon\}$	1

Stochastic Gradient Descent

Description

Following the basic Pegasos algorithm, $\boldsymbol{w}$ is set to be 0 initially. In each round, we pick a random training example $(\boldsymbol{x}_i, y_i)$ in which i is picked with probability p_i, s.t. $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$.

Basic Pegasos Algorithm

Algorithm

1. Choose $\boldsymbol{w}_1 = 0 \in \mathbb{R}^d$
2. Iterate for $t = 1, 2, \ldots, T$
   1. Choose $A_t \subset S = \{1, 2, \ldots, n\}, |A_t| = b$, uniformly at random
   2. Set stepsize $\eta_t \leftarrow \frac{1}{\lambda t}$
   3. Update $w^{(t+1)} \leftarrow w^{(t)} - \eta$

Theorem For $\overline{\boldsymbol{w}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{w}_t$, we have:

$$\mathbb{E}[f(\overline{\boldsymbol{w}})] \leq f(w^*) + c \log(T) \times \frac{1}{\lambda T}$$

where $c = (\sqrt(\lambda) + 1)^2$.

- Technically
- xx
- Methodologically
- xx

Thank You!

Q&A

Acknowledgement:

Thanks to Martin for helpful discussions, suggestions and chips!!!

Reference

[1] Shalev-Shwartz, Shai, et al. "Pegasos: Primal estimated sub-gradient solver for svm." Mathematical programming 127.1 (2011): 3-30.

[2] Lacoste-Julien, Simon, Mark Schmidt, and Francis Bach. "A simpler approach to obtaining an o (1/t) convergence rate for the projected stochastic subgradient method." arXiv preprint arXiv:1212.2002 (2012).

[3] Takáč Martin, et al. "Mini-batch primal and dual methods for SVMs." arXiv preprint arXiv:1303.2314 (2013).

[4] Zhao, Peilin, and Tong Zhang. "Stochastic optimization with importance sampling." arXiv preprint arXiv:1401.2753 (2014).