A Rating Formulation of the SVM

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

October 21, 2004

Abstract

The standard SVM formulation can handle binary labels. Extensions to multiple classes, regression and structured data have been made. However, none of these formulations are appropriate when labels are integer ratings (such as movie or restaurant "star" ratings). Here we provide the primal and dual objectives for an SVM that learns ratings. This is useful for collaborative filtering where data usually comes in the form of an integer rating on each (user,item) pair.

1 Introduction

We assume that each user as a real-valued feature vector that identifies his/her preferences, $\vec{x}_i$. For each object (e.g. book, movie), there are integer ratings (e.g. $\in \{1,2,3,4,5\}$) that identify to what degree individual users prefer that object. Not all users give ratings for all objects. The goal is to be able to predict the rating a user would give to an un-rated object and to predict ratings for a new user, given a feature vector for that new user.

2 SVM Formulation

We define:

- U - number of unique ratings (ratings are $\{1,\ldots,U\}$)
- R - set of index pairs of observed ratings
- $i(r)$ - first index of $r \in R$
- $j(r)$ - second index of $r \in R$
- $y(r)$ - rating observed for $r \in R$
- $\vec{x}_i$ - feature vector for user i
- $\vec{w}_j$ - parameter vector for object j

- b_y - bias term ($y \in \{1, \ldots, U-1\}$)
- α_{kl} - Lagrange multiplier, $k \in \{1, \ldots, |R|\}$, $l \in \{1, 2\}$

The primal constraints are

$$\vec{x}_{i(r)} \cdot \vec{w}_{j(r)} + b_{y(r)-1} \geq +1 - \xi_r, \quad \forall r \text{ s.t. } y(r) > 1, \tag{1}$$

$$\vec{x}_{i(r)} \cdot \vec{w}_{j(r)} + b_{y(r)} \leq -1 + \epsilon_r, \quad \forall r \text{ s.t. } y(r) < U, \tag{2}$$

$$\xi_r \geq 0, \quad \forall r, \quad \text{and} \tag{3}$$

$$\epsilon_r \geq 0, \quad \forall r. \tag{4}$$

The SVM primal objective is

$$J_P = \sum_{j=1}^{m} \frac{1}{2}\|\vec{w}_j\|^2 + C\sum_{r \in R}(\xi_r + \epsilon_r) - \sum_{r \in R} \alpha_r(\vec{x}_{i(r)} \cdot \vec{w}_{j(r)} + b_{y(r)-1} - 1 + \xi_r) + \sum_{r \in R} \beta_r(\vec{x}_{i(r)} \cdot \vec{w}_{j(r)} + b_{y(r)} + 1 - \epsilon_r) - \sum_{r \in R} \mu_r \xi_r - \sum_{r \in R} \nu_r \epsilon_r, \tag{5}$$

where $\alpha_r = 0 \Leftrightarrow y(r) = 1$ and $\beta_r = 0 \Leftrightarrow y(r) = U$. The KKT conditions are

$$\frac{\partial J_P}{\partial \vec{w}_j} = \vec{w}_j - \sum_{r \in R | j(r) = j} (\alpha_r - \beta_r)\vec{x}_{i(r)} = 0, \tag{6}$$

$$\frac{\partial J_P}{\partial b_l} = \sum_{r | y(r) = l+1} \alpha_r - \sum_{r | y(r) = l} \beta_r = 0, \tag{7}$$

$$\frac{\partial J_P}{\partial \xi_r} = C - \alpha_r - \mu_r = 0, \tag{8}$$

$$\frac{\partial J_P}{\partial \epsilon_r} = C - \beta_r - \nu_r = 0, \tag{9}$$

$$\vec{x}_{i(r)} \cdot \vec{w}_{j(r)} + b_{y(r)-1} - 1 + \xi_r \geq 0, \tag{10}$$

$$-\vec{x}_{i(r)} \cdot \vec{w}_{j(r)} - b_{y(r)} - 1 + \epsilon_r \geq 0, \tag{11}$$

$$\xi_r \geq 0, \tag{12}$$

$$\epsilon_r \geq 0, \tag{13}$$

$$\alpha_r \geq 0, \tag{14}$$

$$\beta_r \geq 0, \tag{15}$$

$$\mu_r \geq 0, \tag{16}$$

$$\nu_r \geq 0, \tag{17}$$

$$\alpha_r(\vec{x}_{i(r)} \cdot \vec{w}_{j(r)} + b_{y(r)-1} - 1 + \xi_r) = 0, \tag{18}$$

$$\beta_r(\vec{x}_{i(r)} \cdot \vec{w}_{j(r)} + b_{y(r)} + 1 - \epsilon_r) = 0, \tag{19}$$

$$\mu_r \xi_r = 0, \quad \text{and} \tag{20}$$

$$\nu_r \epsilon_r = 0. \tag{21}$$

These conditions imply

$$\vec{w}_j = \sum_{r \in R | j(r) = j} (\beta_r - \alpha_r) \vec{x}_{i(r)}, \tag{22}$$

$$\sum_{r | y(r) = l+1} \alpha_r = \sum_{r | y(r) = l} \beta_r, \tag{23}$$

$$0 \leq \alpha_r \leq C, \quad \text{and} \tag{24}$$

$$0 \leq \beta_r \leq C. \tag{25}$$

The dual objective is

$$J_D = \sum_{r \in R} (\alpha_r + \beta_r) - \frac{1}{2} \sum_{r \in R} \sum_{s | j(s) = j(r)} (\alpha_r - \beta_r)(\alpha_s - \beta_s) \vec{x}_{i(r)} \cdot \vec{x}_{i(s)} \tag{26}$$

Recall that $\alpha_r = 0 \Leftrightarrow y(r) = 1$ and $\beta_r = 0 \Leftrightarrow y(r) = U$. In other words, minimization of the dual objective is only over $\{\alpha_r\}$ such that $y(r) > 1$ and $\{\beta_s\}$ such that $y(s) < U$.