INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 15, Number 1-2, Pages 154–169

A FIXED-POINT PROXIMITY APPROACH TO SOLVING THE SUPPORT VECTOR REGRESSION WITH THE GROUP LASSO REGULARIZATION

ZHENG LI, GUOHUI SONG, AND YUESHENG XU

In Memory of the Professor Ben-yu Guo

Abstract. We introduce an optimization model of the support vector regression with the group lasso regularization and develop a class of efficient two-step fixed-point proximity algorithms to solve it numerically. To overcome the difficulty brought by the non-differentiability of the group lasso regularization term and the loss function in the proposed model, we characterize its solutions as fixed-points of a nonlinear map defined in terms of the proximity operators of the functions appearing in the objective function of the model. We then propose a class of two-step fixed-point algorithms to solve numerically the optimization problem based on the fixed-point equation. We establish convergence results of the proposed algorithms. Numerical experiments with both synthetic data and real-world benchmark data are presented to demonstrate the advantages of the proposed model and algorithms.

Key words. Two-step fixed-point algorithm, proximity operator, group lasso, support vector machine, ADMM

1. Introduction

The support vector machine (SVM) has been widely used in many applications including text/image recognition [8, 35], face detection [29], bioinformatics [4, 6], since its introduction in [13]. In general, we could consider SVM in two main categories [15, 31, 36]: support vector classification (SVC) [16, 18] and support vector regression (SVR) [2, 32, 33]. The standard ℓ^2-norm SVC aims at finding the best hyperplane that has the largest distance to the nearest points of each class. It turns out that this hyperplane is determined by a small fraction of the training points that are called the support vectors. The standard ℓ^2-norm SVR performs in an analogical way. It maximizes the margin from the hyperplane to the furthest point to get the best fitting hyperplane. Similarly, this hyperplane is also determined by only a small subset of the training points. In this paper we shall focus on SVR.

For the purpose of promoting sparsity of the support vectors, the SVM with the ℓ^1-norm regularizer [31, 36, 38] was put forward. It is well received that the ℓ^1-norm regularizer produces sparse solutions [34]. In particular, the ℓ^1-SVM has been proven to be advantageous when there are redundant noise features [38] and to have shorter training time than the standard ℓ^2-SVM [20]. A natural extension of the ℓ^1-norm regularization is the group lasso regularization that could be viewed as a group-wise ℓ^1-norm. It has been shown in [19, 26, 37] that group lasso regularization overwhelms the ℓ^1-norm regularization when the optimal variable has the group structure. The group lasso regularization performs better when the regression problem has the prior information with group structure [14, 26, 37]. On the other hand, applications with cluster structure have been observed in practice

Received by the editors February 2, 2017 and, in revised form, April 12, 2017.
2000 *Mathematics Subject Classification.* 65N30.

[10, 25]. Therefore, in this paper we shall consider the SVM model with the group lasso regularization.

The main challenge of solving the SVM model with the group lasso regularization comes from the non-differentiability of the SVM loss functions and the group lasso regularization term. A popular technique [9, 11] is to solve a smooth approximation of the original model instead. However, it may bring an extra approximation error term and thus we prefer solving the original model rather than a smooth approximation.

The goal of this paper is to develop numerical algorithms of solving the original SVR model with the group lasso regularization. Specifically, we shall employ the techniques of proximity operators to construct a two-step fixed-point proximity algorithm. We point out that fixed-point proximity algorithms have been popular in solving non-differentiable optimization models in image processing [21, 22, 27, 28] and machine learning [1, 23, 24]. We shall first characterize solutions of the non-differential model as fixed-points of certain nonlinear map defined in terms of the proximity operator of the convex functions involved in the objective function. We then employ a matrix splitting technique to derive a class of two-step algorithms to compute the fixed points.

The rest of this paper is organized as follows. In Section 2, we introduce the optimization model of the group lasso regularized SVR. In Section 3, we characterize solutions of the proposed model as the fixed-points of a nonlinear map defined in terms of the proximity operators of the convex functions appearing in the objective function. We develop a class of two-step proximity algorithms for computing the fixed-points and present its convergence analysis in Section 4. We demonstrate the performance of the proposed model and algorithms in Section 5 through numerical experiments with both synthetic data and real-world benchmark data. We draw a conclusion in Section 6.

2. SVR with Group Lasso Regularization

In this section, we shall introduce the model of the SVR with group lasso regularization. To this end, we first recall the models of the standard ℓ^2-norm SVR (ℓ^2-SVR) and the variant ℓ^1-norm SVR (ℓ^1-SVR).

We start with the notation used throughout this paper. We denote by $\mathbb{R}^m$ the usual m-dimensional Euclidean space and define

$$\mathbb{R}^m_+ := \{\boldsymbol{x} \in \mathbb{R}^m : x_i \geq 0\}.$$

For a positive integer $m \in \mathbb{N}$, we set $\mathbb{N}_m := \{1, 2, \ldots, m\}$. The standard inner product is defined for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i \in \mathbb{N}_m} x_i y_i.$$

For $p \in \mathbb{N}_2$, we define the ℓ^p norm for $\boldsymbol{x} \in \mathbb{R}^m$ by

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p}.$$

We next recall the SVR models. Given instances $\{(\boldsymbol{x}_i, y_i) : i \in \mathbb{N}_m\} \subseteq \mathbb{R}^n \times \mathbb{R}$, the standard ℓ^2-norm soft margin SVR aims at finding the best hyperplane that has the largest margin to the farthest training points. This leads to the following