Support Vector Regression using Deflected Subgradient Methods

Elia Piccoli
Nicola Gugole

February 6, 2021

A project presented for the
Computational Mathematics for Learning and Data Analysis
course

University of Pisa
Artificial Intelligence
A.Y. 2020/2021

Abstract

Project aim is developing the implementation of a model which follows an SVR-type approach including various different kernels. The implementation uses as optimization algorithm a dual approach with appropriate choices of the constraints to be dualized, where the Lagrangian Dual is solved by an algorithm of the class of deflected subgradient methods.

1 Introduction

SVR objective is predicting a unidimensional real-valued output y through the use of an *objective function* built by optimization using an ε-insensitive loss function. Another fundamental aspect about SVR is keeping the function *as flat as possible* through the tuning of a C parameter in order to avoid overfitting and generating a correct tradeoff between accuracy and generalization.
The resulting function can be generically described as:

$$f(x) = wx + b \tag{1}$$

Keeping the above function *as flat as possible* is equivalent to an optimization problem formulated as having minimum $\|w\|$, or, for a more convenient mathematical derivation, minimum $\|w\|^2$. This does not change the semantics of the problem.
This brings us to a convex minimization problem, which will be called *primal problem*:

$$\min_{w,\xi_i,\xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_i (\xi_i + \xi_i^*) \tag{2}$$

Where ξ and ξ^* are called *slack variables*, used in conjunction with C to create a *regularization factor* and consequently a *penaltymeasure* to elements which are not part of the ε-tube. Slack variables allow the definition of constraints applicable to (2):

$$y_i - w^T\phi(x_i) - b \leq \varepsilon + \xi_i, \tag{3a}$$
$$b + w^T\phi(x_i) - y_i \leq \varepsilon + \xi_i, \tag{3b}$$
$$\xi_i, \xi_i^* \geq 0 \tag{3c}$$

x_i input, y_i output

2 Dual Representation

As expressed in the abstract, the implementation will follow a dual approach, which in SVR models is preferred due to the applicability and efficiency of the use of *kernels*. *Dual problem* formulation can be achieved defining the *Lagrangian* function:

$$\begin{aligned}\mathcal{L}(\alpha, \alpha^*, \mu, \mu^*) = \frac{1}{2}\left\|w\right\|^2 \\ + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \\ + \sum_{i=1}^{m}(\alpha_i(y_i - w^T\phi(x_i) - b - \varepsilon - \xi_i)) \\ + \sum_{i=1}^{m}(\alpha_i^*(w^T\phi(x_i) + b - y_i - \varepsilon - \xi_i^*)) \\ - \sum_{i=1}^{m}(\mu_i\xi_i + \mu_i^*\xi_i^*)\end{aligned} \tag{4}$$