Optimization

L1-Norm SVM

N Chakradhar[1] Havish[2]

[1]EE16BTECH11022

[2]EE16BTECH11023

March 6, 2019

Introduction

The paper presents a comparitive analysis of how L1, L2 regularization can be applied for learning SVM parameters and also the advantages of using L1 regularization over L2 regularization

Note on SVM

An SVM model is a representation of the dataset as points in space, mapped so that the data points of the separate categories are divided by a clear gap that is as wide as possible.

The problem formulation

- The classifier is given by $f_i(x_i) = w^T x_i + w_0$
- Objective function:

$$\Sigma_{i=1}^{i=n} max(0, 1 - y_i f_i(x)) \tag{1}$$

- Constraint:

$$||w||_p \leq k \tag{2}$$

- Using lagrange multipliers, we can convert this to unconstrained optimization problem as follows

$$min_{w_0, w_j} \lambda ||w||_p + \Sigma_{i=1}^{i=n} max(0, 1 - y_i f_i(x)) \tag{3}$$

The convexity of the problem

- Clearly, the given constraint set is a convex set. This can be trivially proved by using triangle inequality and the definition of convex set.
- The objective function itself is nothing but maximum of two convex functions which is also convex. This can be proved by using the bi-implication that a function is convex iff epigraph set is convex.

Use of regularization

- Regularization helps in shrinking the weights such that they do not have high variance
- Also, in the case of L1-regularization, we can tune the constraint parameter such that the some of the weights can be exactly made 0. This can be interpreted as some sort of feature selection.

Implementation

- In the paper, the two approaches were tested on cancer dataset which has very less number of training samples.
- We would like to frame both the optimization problems using cvxpy, compare their performance on a similar dataset and also observe how tuning of the constraints affects the weights.