Untitled

Paballo Moeletsi

January 28, 2017

Recall that the primal problem for non linearly separable data of the svm takes the form of the following quadratic optimisation problem.

$\min_{\beta,\beta_0} \quad \|\beta\|_2^2 + C\sum_{i=1}^n \xi_i$

subject to :

$$y_i(\beta^T x + b) + \xi_i \leq 1$$

where the data set corresponds to the pairs x_i, y_i

the weights of the maximum margin hyperplane are β

the ξ_i are the misclassifications

The SVM does not automatically perform feature selection. However there are various ways agument the learning procedure so that feature selection is undertaken. These methods include L_1 norm regularization, recursive feature elimination, tree based feature selection etc.

The goal of this appoarch is to perform feature selection as part of the solving the optimisation problem very similar to the L_1 norm appoach.

One way of achiveing this is associating each each coeffient with a binary variable z_i and then forcefully requring a select number of the z_i to be zero which would inturn make the weight zero and remove the i^{th} feature. One way to achieve this is with the following :

$\min_{\beta,\beta_0} \quad \|\beta\|_2^2 + C\sum_{i=1}^n \xi_i$

subject to :

$$y_i(\beta^T x + b) + \xi_i \leq 1$$
$$|\beta_i| \leq M z_i$$
$$\sum_{i=1}^n z_i = \frac{x}{100} p$$

.If $z_i = 0$ then we have $|\beta| \leq z_i M \implies \beta_i = 0$.The restriction that$\sum_{i=1}^n z_i = \frac{x}{100}p$ ensures that **at most** $\frac{x}{100}p$ β_i will be non zero.Hence for small p only a small subset of the $\langle\beta\rangle$ vector will be non zero. We choose M be some very large positive integer.Using the definition of the absolute value, the second inequality of (1) can written as two inequalities which gives the following equivalent optimisation problem with only linear constraints:

$\min_{\beta,\beta_0} \quad \|\beta\|_2^2 + C\sum_{i=1}^n \xi_i$

subject to :

$$y_i(\beta^T x + b) + \xi_i \leq 1$$
$$\beta_i - z_i M \leq 0$$
$$\beta_i + z_i M \geq 0$$
$$\sum_{i=1}^n z_i = \frac{x}{100} p$$

This is a integer quadratic programming problem with linear constraints. This the optimisation problem that is solved in the python code.