UNIVERSITAT POLITÈCNICA DE CATALUNYA

FACULTAT D'INFORMÀTICA DE BARCELONA

Extension Report: Multi-Class

Analysis of the SVM-RFE algorithm for feature selection

Author:
Robert PLANAS

Director:
Luis A. BELANCHE

Bachelor Degree in Informatics Engineering
Specialization: Computing

June 2, 2021

Contents

Chapter 1

Multi-class classification

In this section we extend SVM-RFE to the multi-class classification problem.

1.1 Description and reasoning

When it comes to extending SVM to handle a multi-class problems two common methods exists, OvR (One-vs-Rest) and OvO (One-vs-One). In both cases the idea is to divide the problem in a set of binary classification problems and use a joint decision function that operates on the results of each of these. Because we're not really making any predictions during the SVM-RFE procedure, we can not use this joint decision function. Instead, we must find a way to merge the ranking criteria obtained form each problem to find a joint ranking criteria.

We know that the ranking criteria is an estimator of the importance of some feature for a given binary decision problem. It can be the case that a feature is very useful to distinguish between two classes but useless for the rest. In this case a joint ranking criteria formed by taking the mean, the median or the sum will result in poor selections. A better idea would be to take the maximum. However, it may also be desirable to estimate the joint importance a feature has by considering its individual importance in more than one problem, that is, a feature that is important in more than one binary classification problem is more important than another that is only important in one such classification problem even if the second feature has a greater individual importance. A way to perform such ranking would be, for instance, the sum of the squares. For both methods proposed a normalization of all feature rankings is probably adequate.

Note that `sklearn` only supports OvO, and is therefore the option we will use.

1.2 Pseudocode formalization

Definitions:

- $X_0 = [\vec{x}_0, \vec{x}_1, \ldots, \vec{x}_k]^T$ list of observations.

- $\vec{y} = [y_1, y_2, \ldots, y_k]^T$ list of labels.

Algorithm 1: SVM-RFE for multi-class classification problems

   Input: t `// t = step`

   Output: $\vec{r}$

   Data: $X_0, \vec{y}$

1   $\vec{s} = [1, 2, \ldots, n]$ `// subset of surviving features`

2   $\vec{r} = []$ `// feature ranked list`

3   **while** $|\vec{s}| > 0$ **do**

         `/* Restrict training examples to good feature indices    */`

4      $X = X_0(:, \vec{s})$

         `/* Compute the joint ranking criteria                     */`

5      $\vec{c} = [0, 0, \ldots,]$

6   **for** $\vec{X}l \subseteq \vec{X}, \vec{y}l \subseteq \vec{y}$ *with* $\vec{X}l$ *and* $\vec{y}l$ *being an instance of OvO* **do**

            `/* Train the classifier                                */`

7         $\vec{\alpha} = \texttt{SVM-train}(\vec{X}l, \vec{y}l)$

            `/* Compute the weight vector of dimension length` $|\vec{s}|$ `   */`

8         $\vec{w} = \sum_k \vec{\alpha}_k \vec{y}l_k \vec{X}l_k$

            `/* Compute the joint ranking criteria                  */`

9         $\vec{c} = \left[\max(c_i, (w_i)^2) \text{ for all } i\right]$

10   **end**

         `/* Find the` t `features with the smallest ranking criterion   */`

11     $\vec{f} = \texttt{argsort}(\vec{c})(: t)$

         `/* Iterate over the feature subset                        */`

12   **for** $f_i \in \vec{f}$ **do**

            `/* Update the feature ranking list                     */`

13        $\vec{r} = [\vec{s}(f_i), \ldots \vec{r}]$

            `/* Eliminate the feature selected                      */`

14        $\vec{s} = [\ldots \vec{s}(1 : f_i - 1), \ldots \vec{s}(f_i + 1 : |\vec{s}|)]$

15   **end**

16 **end**

1.3   Results