Seeking the optimal weights in Multitask Multiclass Support Vector Machine

Asunción Jiménez Cordero

OASYS Group (`oasys.uma.es`)

1 Goal

The goal of this document is to formally state the optimization problem which is solved in the example named as *M2SVM_Optimal_Weights*.

2 Problem Formulation

Let us assume given a set of time periods $t = 1, \ldots, T$. For each t, we consider the pair (x_t, y_t), where $x_t = (x_{t1}, \ldots, x_{tB}) \in \mathbb{R}^B$ denotes the vector of features, and $y_t \in \{1, \ldots, Q\}$ is the class label associated to each time period. The aim is to associate a weight ω_b, $b = 1, \ldots, B$ to each one of the features of x_t, $\forall t$ in order to obtain a classification of the data as accurate as possible. To do this, the following optimization problem is solved:

$$\left\{\begin{array}{lll} \min\limits_{\substack{\omega_b, \forall b,\\ r_t, \forall t}} & \sum\limits_{t=1}^{T} r_t & \\ & & \\ \text{s.t.} & r_t \geq -\delta_{q,y_t} + 1 + \frac{1}{2\beta} \sum\limits_{s=1}^{T} (\delta_{q,y_s} - \alpha_{qs} - \delta_{y_s,y_t} + \alpha_{y_t\, s}) K_\omega(x_t, x_s), & q = 1, \ldots, Q, \\ & & t = 1, \ldots, T \\ & \omega_b \geq 0, \forall b. & \end{array}\right. \tag{1}$$

where $\delta_{i,j}$ denotes the Dirac delta function, which is equal to 1 if $i = j$, and 0 otherwise. In addition, α_{qt} are the optimal solution of the Multitask Multiclass Support Vector Machine (M2SVM) optimization problem in (2), β is the regularization parameter, and $K_\omega(x_t, x_s)$ is the Gaussian kernel function that can be seen in (3).

$$
\begin{cases}
\min\limits_{\alpha} & \sum\limits_{q=1}^{Q}\sum\limits_{t=1}^{T}\alpha_{qt}\delta_{q,y_t} + \dfrac{1}{4\beta}\sum\limits_{q=1}^{Q}\sum\limits_{t,\,s=1}^{T}(\alpha_{qt}-\delta_{q,y_t})(\alpha_{qs}-\delta_{q,y_s})K_\omega(d_t,d_s) \\
\text{s.t.} & \sum\limits_{q=1}^{Q}\alpha_{qt} = 1, \quad t=1,\dots,T \\
 & \alpha_{qt} \geq 0, \quad q=1,\dots,Q,\ t=1,\dots,T
\end{cases}
\tag{2}
$$

$$
K_\omega(x_t,x_s) = \exp\left(-\sum_{b=1}^{B}\omega_b(x_{tb}-x_{sb})^2\right) \tag{3}
$$

It is important to notice that the optimal solution of (2) is assumed to be known, since the objective of this project is to check the performance of the different nonlinear solvers, Therefore the only optimization problem which is solved in this case is Problem (1) whose nonlinearity is given via the exponential function of the kernel in (3).