Modified Regularized Least Squares Classification

Jason D. M. Rennie
jrennie@gmail.com

February 14, 2005

Zhang and Oles [1] note that parameters for the Support Vector Machine cannot be optimized direcly (using traditional, gradient-descent-type approaches) due to the use of the hinge loss function,

$$g(z) = \max(0, 1 - z). \tag{1}$$

The hinge loss function has a discontinuous derivative at $z = 1$. They propose an alternate, smooth loss based on the squared loss,

$$h(z) = \begin{cases} (1-z)^2 & \text{if } z \leq 1 \\ 0 & \text{if } z > 1 \end{cases} \tag{2}$$

This "modified squared loss" is similar to the hinge loss. In particular, it serves as an upper bound to the step function that is tight at $z = 0$ and $z \geq 1$. See Figure 1 for plots of the two loss functions. Unlike the hing loss, the modified squared loss is sufficiently smooth to be optimized via simple, gradient-descent-type algorithms.

The full minimization objective for Modified Regularized Least Squares Classification (MRLSC) is

$$J_{MRLSC} = \sum_{i=1}^{n} h(y_i \vec{x}_i^T \vec{w}) + \frac{\lambda}{2} \vec{w}^T \vec{w}, \tag{3}$$

where $\{\vec{x}_1, \dots, \vec{x}_n\}$, $\vec{x}_i \in \mathbb{R}^d$ are the training examples, $\vec{y} \in \mathbb{R}^n$ are the training labels, $\vec{w} \in \mathbb{R}^d$ is the weight vector and λ is the regularization parameter. The gradient is

$$\frac{\partial J_{MRLSC}}{\partial \vec{w}} = \sum_{i=1}^{n} y_i \vec{x}_i h'(y_i \vec{x}_i^T \vec{w}) + \lambda \vec{w}, \tag{4}$$

where $h'(z) = \begin{cases} 2(z-1) & \text{if } z \leq 1 \\ 0 & \text{if } z > 1 \end{cases}$.

References

[1] T. Zhang and F. J. Oles. Text categorization based on regularized linear classification methods. *Information Retrieval*, 4:5–31, 2001.

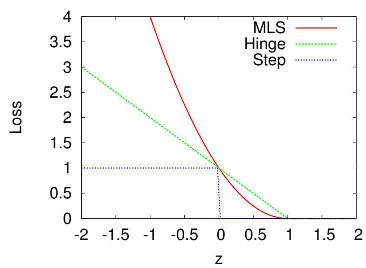

Figure 1: Shown are the Modified Least Squares (MLS), Hinge and Step loss functions.