Classification using sparse combination of base functions

Kornél Kovács and András Kocsor

Tasks in machine learning often lead to classification and regression problems where applying models using convex objective functions could be beneficial. Consider the problem of classifying *n* points in a compact set $\mathcal{X}$ over $\mathbb{R}^m$, represented by $\mathbf{x}_1, \ldots, \mathbf{x}_n$, according to membership of each point $\mathbf{x}_i$ in the classes $+1$ or -1 as specified by $y_1, \ldots, y_n$. First, let S denote a finite set of continuous base functions

$$S = \{f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})\} \qquad f_i : \mathcal{X} \to \mathbb{R}.$$

Second, consider the following general convex optimization problem of the classification task:

$$\inf_{f(\mathbf{x}) \in Span(S)} \sum_{i=1}^{n} L\left(f(\mathbf{x}_i) y_i\right), \tag{3}$$

where $L : \mathbb{R} \to \mathbb{R}$ a method dependent convex loss function, and $Span(S)$ denotes the linear space generated by the base functions

$$Span(S) = \left\{ h : \mathcal{X} \to \mathbb{R} \mid h(\mathbf{x}) = \sum_{i=1}^{k} \alpha_i f_i(\mathbf{x}),\ \mathbf{x} \in \mathcal{X} \right\}.$$

Taking into account the fact that $f(\mathbf{x}) \in Span(S)$, i.e. $f(\mathbf{x}) = \sum_{i=1}^{k} \alpha_i f_i(\mathbf{x})$, Eq. (3) then has the following form:

$$\inf_{\boldsymbol{\alpha}} g\left(\boldsymbol{\alpha}\right), \tag{4}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^T$ and

$$g\left(\boldsymbol{\alpha}\right) = \sum_{i=1}^{n} L\left(\sum_{j=1}^{k} \alpha_j f_j(\mathbf{x}_i) y_i\right).$$

We show that the optimization problem defined in Eq. (4) includes several well-known machine learning algorithms, such as certain variants of boosting methods [1, 2] and Support Vector Machines [3, 4]. The nonlinear Gauss-Seidel (GS) method can be applied to optimize Eq. (4), which alters model parameters one at a time. If $\nabla g\left(\boldsymbol{\alpha}\right)$ has the Lipschitz continuity property in Eq. (4) the convergence of GS can be proved. The GS method has low memory requirements during optimization, but in large real-life problems, the solution is practically infeasible due to the numerous iteration steps.

That is why the application of (heuristic) methods providing approximate solution seem important here. We define a set of heuristic methods which quickly and efficiently determines adequately functioning suboptimal solutions in a classification sense. The algorithms are based on the methods of feature selection, a special field in machine learning. The methods used here are called Sequential Forward Selection, Plus *l*-Take Away *r* and Sequential Forward Floating Selection.

The proposed algorithms looks for solutions that have a predefined number of nonzero components among the model parameters. We provide a justification for them by solving several tasks using data taken from the UCI Repository [5] which is widely used for testing machine learning algorithms.

References

[1] FREUND, Y. AND SCHAPIRE, R.E. *A decision-theoretic generalization of on-line learning and an application to boosting,* J. Comput. Syst. Sci., vol. 55/1, pp. 119-139, 1997.

[2] FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R. *Additive logistic regression: A statictical view of boosting,* The Annals of Statistics, vol. 28/2, pp. 337-407, 2000.

[3] LEE, Y.-J. AND MANGASARIAN, O. L. *SSVM: A Smooth Support Vector Machine for Classification,* Computational Optimization and Applications, vol. 20/1, pp. 5-22, 2001.

[4] SUYKENS, J.A.K. AND VANDEWALLE, J. *Least squares support vector machine classifiers,* Neural Processing Letters, 1999.

[5] BLAKE, C. L. AND MERZ, C. J. *UCI repository of machine learning databases,* http://www.ics.uci.edu/ mlearn/MLRepository.html, 1998.