Three Viewpoints Toward Exemplar SVM

Takumi Kobayashi
National Institute of Advanced Industrial Science and Technology, Japan

Classification plays a key role in various pattern recognition problems such as image recognition and object detection. While multi-class classification is mainly addressed by decomposing multiple categories into independent categories in a *one-vs-rest* manner, an extreme classification formulation called *exemplar SVM* [1] has been recently proposed to discriminate *each* instance sample (exemplar) from the others; that is, a category is divided into samples at the finest resolution. The exemplar SVM (ex-SVM), however, has difficulty in determining (tuning) two regularization parameters that balance effects of positive and negative classification costs, due to the highly biased classification problem where only one positive sample is contrasted with a large number of negative samples. Besides, classification scores of multiple exemplar SVMs are required to be calibrated so as to be comparable among an ensemble of heterogeneous classifiers which are individually trained.

This paper presents three viewpoints for ex-SVM; one is original formulation while the other two are novel. From these viewpoints, we can give light on an intrinsic structure of exemplar SVM. Specifically, the regularization parameters, which are hard to be determined in the original ex-SVM, are reduced into only one parameter whose role is also clearly provided so as to intuitively determine the parameter value. And, we can clarify how the classifier geometrically works, which frees us from carefully calibrating the classification scores. As a result, exemplar SVM can be utilized more simply without careful tuning. In addition, those viewpoints lead to novel feature transformation using exemplar SVM. Although exemplar SVM has been mainly applied to detection tasks so far, the proposed feature transformation contributes to improve performance on general classification tasks.

Exemplar SVM (ex-SVM) is originally formulated in [1] as a binary classification problem to discriminate only one target positive sample $\boldsymbol{x}\in\mathbb{R}^d$ from the other negatives $\{\boldsymbol{\xi}_i\}_{i=1}^N$ (Fig. 1a). It is formulated as

$$\min_{\boldsymbol{w},b}\frac{1}{2}\|\boldsymbol{w}\|^2+C_p\mathsf{h}(1-\boldsymbol{w}^\top\boldsymbol{x}-b)+C_n\sum_{i=1}^N\mathsf{h}(1+\boldsymbol{w}^\top\boldsymbol{\xi}_i+b),\qquad(1)$$

where $\mathsf{h}(x)$ is a hinge loss function and $\{C_p,C_n\}$ are regularization parameters for balancing the positive and negative costs. Its dual is given by

$$\min_{\hat{\boldsymbol{\alpha}}}\frac{1}{2}\hat{\boldsymbol{\alpha}}^\top\hat{\boldsymbol{K}}\hat{\boldsymbol{\alpha}}-\mathbf{1}^\top\hat{\boldsymbol{\alpha}},\ s.t.\ \alpha_0-\sum_{i=1}^N\alpha_i=0,\ 0\le\alpha_0\le C_p,\ 0\le\alpha_i\le C_n,\quad(2)$$

where $\hat{\boldsymbol{\alpha}}=[\alpha_0,\alpha_1,\cdots,\alpha_N]^\top$ in which α_0 is for positive and $\boldsymbol{\alpha}=\{\alpha_i\}_{i>0}$ are for negatives. Ex-SVM is characterized by this highly unbalanced formulation in which only one positive sample is contrasted with plenty of negative samples drawn such as from the categories other than the target one. Thus, it is required to carefully tune the regularization parameters C_p and C_n.

The first novel viewpoint is related to one-class SVM [2]. We reformulate ex-SVM by rewriting the dual problem (2) into

$$\min_{\bar{\boldsymbol{\alpha}}}\frac{1}{2}\bar{\boldsymbol{\alpha}}^\top\bar{\boldsymbol{K}}\bar{\boldsymbol{\alpha}},\ \ s.t.\ \sum_{i=1}^N\bar{\alpha}_i=1,\ 0\le\bar{\alpha}_i=\frac{\alpha_i}{\alpha_0^*}\le\frac{C_n}{\alpha_0^*}=C,\qquad(3)$$

where α_0 is *fixed* to its optimizer α_0^* of (2), the constant term regarding α_0^* is ignored, variables $\boldsymbol{\alpha}$ are rescaled by α_0^* into $\bar{\boldsymbol{\alpha}}$, and $\bar{\boldsymbol{K}}\in\mathbb{R}^{N\times N}$ is the kernel Gram matrix *centered at* $\boldsymbol{x}$; in the case of linear kernel, $\bar{K}_{ij}=(\boldsymbol{\xi}_i-\boldsymbol{x})^\top(\boldsymbol{\xi}_j-\boldsymbol{x})$. This is exactly the same as the dual of one-class SVM (oc-SVM) [2] with a regularization parameter C. Thus, we can insist that exemplar SVM is intrinsically reduced to one-class SVM in the feature space centered at the target sample $\boldsymbol{x}$ as shown in Fig. 1b. This insight into ex-SVM is also practically useful in the following three points regarding parameter issues; (i) Two parameters C_p and C_n are reduced to only one parameter C, (ii) the paramter C is limited in $\frac{1}{N}\le C\le 1$, and (iii) C controls the number of support vectors (outliers) which is greater than $\frac{1}{C}$ [2]. Thus, in the case that negative samples are all drawn from the categories other than the target one, we can simply set $C=1$ to minimize outliers.

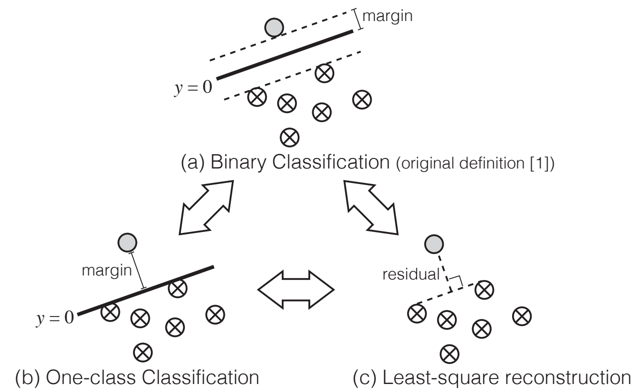

Figure 1: Three viewpoints for exemplar SVM: (a) Original formulation of large margin binary classification [1], (b) the first novel viewpoint by one-class large margin classifier and (c) the second novel viewpoint by least-square reconstruction. Gray circle indicates the target positive sample, while the others are negative samples.

We then show the second novel viewpoint for exemplar SVM in the framework of least squares. We consider a problem to reconstruct $\boldsymbol{x}$ by using (restricted) convex combination of $\{\boldsymbol{\xi}_i\}_{i=1}^N$;

$$\min_{0\le\bar{\boldsymbol{\alpha}}\le C,\mathbf{1}^\top\bar{\boldsymbol{\alpha}}=1}\frac{1}{2}\|\boldsymbol{x}-\sum_{i=1}^N\bar{\alpha}_i\boldsymbol{\xi}_i\|_2^2\ \Leftrightarrow\ \min_{0\le\bar{\boldsymbol{\alpha}}\le C,\mathbf{1}^\top\bar{\boldsymbol{\alpha}}=1}\frac{1}{2}\bar{\boldsymbol{\alpha}}^\top\bar{\boldsymbol{K}}\bar{\boldsymbol{\alpha}},\qquad(4)$$

where the optimum coefficients $\bar{\boldsymbol{\alpha}}^*$ are obtained by the least-square method. Since this primal problem (4) is the same quadratic programming (QP) as (3), this least-square formulation is connected to ex-SVM via the dual problem in oc-SVM. The least-square formulation is also found in (unsupervised) similarity metric learning methods [3] which employ a locally linear model to approximate a sample vector by non-negative convex construction of the other samples. They are different from the least-square formulation for ex-SVM in that the optimized non-negative coefficients $\bar{\boldsymbol{\alpha}}_i^*$ are utilized to construct similarity measure between $\boldsymbol{x}$ and $\boldsymbol{\xi}_i$, while we make use of the residual vector for an ex-SVM classifier. In that sense, the similarity learning methods and ex-SVM are complementary to each other, through the identical QP formulation.

We conclude that the above three formulations produce the same ex-SVM classifier (except for the scaling derived from α_0^*) and thus they reveal various intrinsic characteristics of ex-SVM from different aspects. The novel two viewpoints have advantages not only in reducing parameters but also in unifying LDA version of ex-SVM, facilitating an ensemble of ex-SVM detectors, and particularly establishing ex-SVM based feature transformation. In the experiments on object detection and image classification, the proposed methods regarding ex-SVM exhibit favorable performance.

[1] Tomasz Malisiewicz, Abhinav Gupta, and Alexei A. Efros. Ensemble of exemplar-svms for object detection and beyond. In *International Conference on Computer Vision (ICCV)*, pages 89–96, 2011.

[2] B. Scholkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13:1443–1471, 2001.

[3] F. Wang and C. Zhang. Label propagation through linear neighborhoods. *IEEE Transactions on Knowledge and Data Engineering*, 20(1): 55–67, 2008.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.