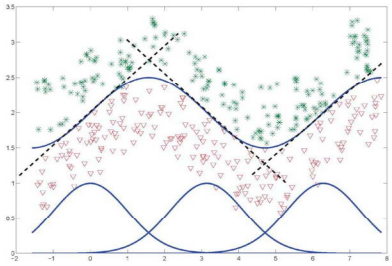
Quasi-Linear Support Vector Machine for Nonlinear Classification

Bo ZHOU, Weite LI, Benhui CHEN and Jinglu HU
Graduate School of Information Production and Systems, WASEDA University

1. Motivation

- An SVM with a composite quasi-linear kernel, which realizes a multi-local linear classifier with interpolation, so as to prevent the potential overfitting problem.



2. SVM with a Quasi-Linear Kernel

$$f_P(x)=\sum_{j=1}^{M}(\Omega_j^T x+b_j)R_j(x)+b$$

$$\Phi(x)=\left[R_1(x),x^T R_1(x),\ldots,R_M(x),x^T R_M(x)\right]^T$$
$$\Theta=[b_1,\Omega_1^T,\ldots,b_M,\Omega_M^T]^T$$

$$f_P(x)=\Theta^T\Phi(x)+b$$

$$\max_\alpha \mathcal{J}_D(\alpha)=-\frac{1}{2}\sum_{k,l=1}^{N}y_k y_l K(x_k,x_l)\alpha_k\alpha_l+\sum_{k=1}^{N}\alpha_k$$

$$s.t.\begin{cases}\sum_{k=1}^{N}\alpha_k y_k=0\\ 0\le\alpha_k\le c, k=1,\ldots,N\end{cases}$$

$$K(x_k,x_l)=\Phi^T(x_k)\Phi(x_l)=(1+x_k^T x_l)\sum_{j=1}^{M}R_j(x_k)R_j(x_l)$$

3. Quasi-Linear Kernel

$$K(x_k,x_l)=\Phi^T(x_k)\Phi(x_l)=(1+x_k^T x_l)\sum_{j=1}^{M}R_j(x_k)R_j(x_l)$$

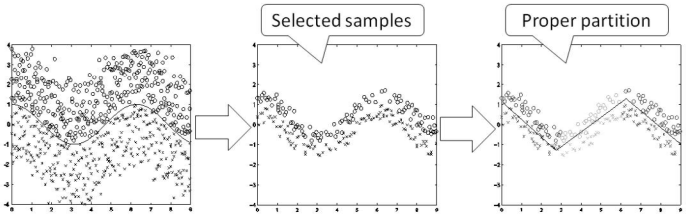
Linear kernel $K(x_k,x_l)=1+x_k^T x_l$ (M=1) ... (M→∞) Nonlinear kernel, e.g. RBF kernel function $K(x_k,x_l)=k_n(x_k,x_l)$

- It provides a flexible and adjustable kernel, filling the gap between linear and nonlinear kernels.

4. Implementation (Partitioning)

- Guided partition method with modified K-means
  - B. Zhou, B. Chen, and J. Hu, "Quasi-linear support vector machine for nonlinear classification," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 97, no. 7, pp. 1587–1594, 2014.

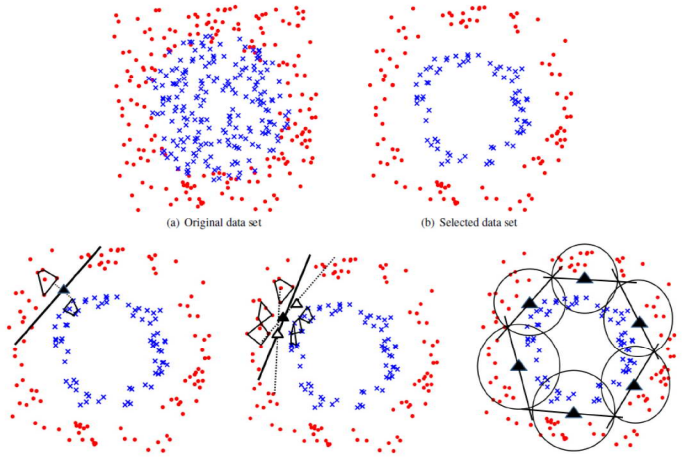
$$\min\{\sum_{j=1}^{k}\sum_{i=1}^{n}\|X_i-C_j\|^2+\lambda_1\sum_{j=1}^{k}|\sum_{i=1}^{n}Z_{i,j}Y_i|+\lambda_2\sum_{j=1}^{k}D_j\}$$



Original training data — Separating boundary detection — K-Means clustering

4. Implementation (cont'd)

- A geometry-based method to detect local linear partition
  - W. Li and J. Hu, "Geometric Approach of Quasi-linear Kernel Composition for Support Vector Machine", in *Proc. of 2015 IEEE International Joint Conference on Neural Networks* (IJCNN'2015) (Killarney), July, 2015.

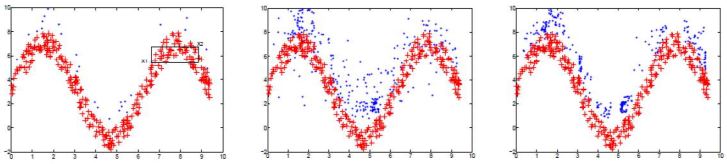


(a) Original data set (b) Selected data set

(c) A classifier separating two convex hulls (d) Multiple classifiers separating the potential separation boundary. (e) The local linear partitions extracted and the local linear classifiers.

5. Related Applications

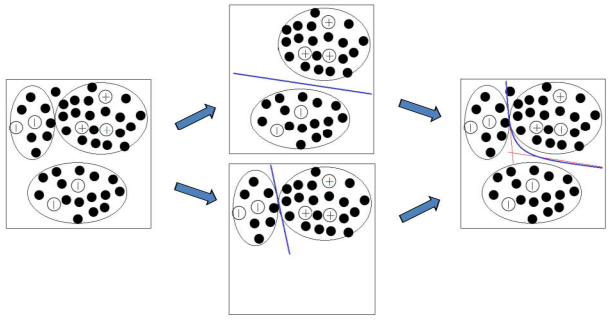
- A local linear based oversampling for imbalanced data classification
  - B. Zhou, C. Yang, H. Guo, and J. Hu, "A quasi-linear SVM combined with assembled SMOTE for imbalanced data classification," in *Proc. of 2013 IEEE International Joint Conference on Neural Networks* (IJCNN'2013) (Dallas), August 2013, pp. 2351–2357.



(a) Imbalanced dataset (b) Standard SMOTE with k = 10 (c) The proposed segmented SMOTE

Figure 3: Overlapping problems in generating synthetic samples by a standard SMOTE

- Transductive SVM with quasi-linear kernel for semi-supervised classification
  - Composite Density Information of Unlabeled Samples in quasi-linear kernel.
  - B. Zhou and J. Hu "A Transductive SVM with quasi-linear kernel based on cluster assumption for semi-supervised classification." in *Proc. of 2015 IEEE International Joint Conference on Neural Networks* (IJCNN'2015), 2015.
  - Building Adjustable Model in TSVM Training
  - B. Zhou, C. Hu and J. Hu "A Transductive Support Vector Machine with adjustable quasi-linear kernel for semi-supervised data classification." in *Proc. of 2014 IEEE International Joint Conference on Neural Networks* (IJCNN'2014), 2014.



古月研究室
早稲田大学大学院情報生産システム研究科
Email: {zhoubo_bird@toki; liweite@akane; jinglu@}waseda.jp.
Tel/Fax: 093-692-5271