Nonparametric Pairwise Similarity for Discriminative Clustering

Yingzhen Yang, Xinqi Chu, Thomas S. Huang

Beckman Institute for Advanced Science and Technology,
University of Illinois at Urbana-Champaign

December 9, 2013

Introduction

- Pairwise clustering
  - Group the data solely by pairwise similarity between the data points, such as Spectral Clustering and Pairwise Clustering and Graphical Models.
  - Often without estimating complex hidden variables or parameters which is difficult for high dimensional data
- The success of pairwise clustering largely depends on the pairwise similarity.
- Challenge: how to model pairwise similarity?
  - Pairwise similarity based on unsupervised classification

Unsupervised Classification by Plug-In Classifier

- Given the data $\{\mathbf{x}_l\}_{l=1}^n \overset{i.i.d.}{\sim} P_X, \{\mathbf{x}_l\}_{l=1}^n \subset \mathbb{R}^d$, clustering is equivalent to searching for the optimal hypothetical labeling $\{\hat{y}_l\}$.

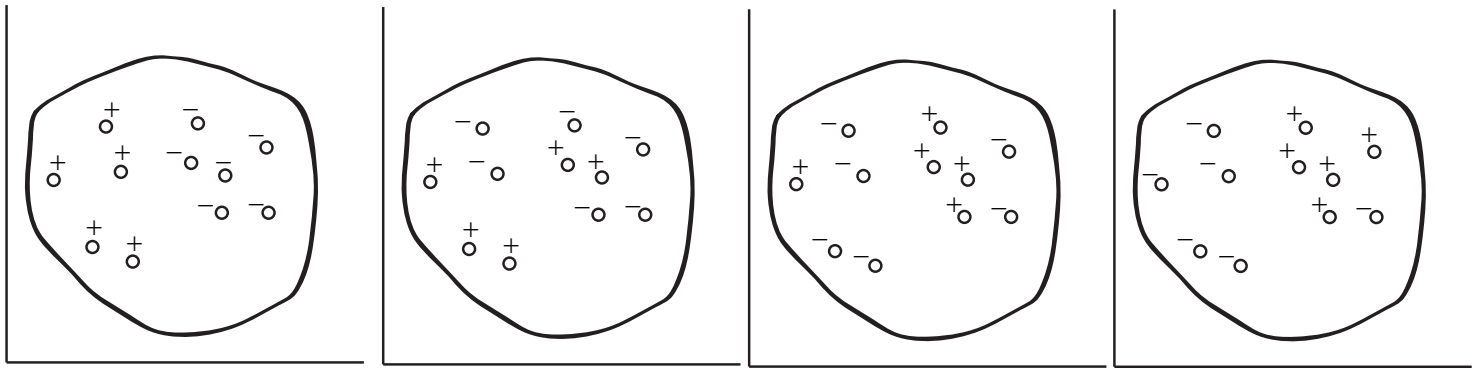

Figure 1: Illustration of binary hypothetical labelings

- For each hypothetical labeling, a plug-in classifier is learned from the training data $S_{\hat{\mathbf{y}}} = \{\mathbf{x}_l, \hat{y}_l\}$:

$$\mathrm{PI}_{S_{\hat{\mathbf{y}}}}(X) = \arg\max_{1 \le i \le Q} \hat{\eta}_{n,h_n}^{(i)}(X) \quad \hat{\eta}_{n,h_n}^{(i)}(x) = \frac{\sum_{l=1}^{n} K_{h_n}(x - \mathbf{x}_l) \mathbb{I}_{\{\hat{\mathbf{y}}_l = i\}}}{n \hat{f}_{n,h_n}(x)}$$

Unsupervised Classification by Plug-In Classifier

- Under some mild conditions on the class-conditional densities and the marginal density, the main theorem of this paper states that there exists n_0, when $n > n_0$, with high probability the generalization error of the plug-in classifier $\mathrm{PI}_{S_{\hat{\mathbf{y}}}}$ satisfies

$$R\left(\mathrm{PI}_S\right) \leq \frac{1}{n^2} \sum_{l,m} \theta_{lm} G_{lm,h_n} + \mathcal{O}\left(n^{-\frac{(1-\varepsilon)\gamma}{d+2\gamma}}\right) \tag{1}$$

  where $\theta_{lm} = \mathbb{1}_{\{\mathbf{y}_l \neq \mathbf{y}_m\}}, G_{lm,h_n} = \dfrac{K_{h_n}(\mathbf{x}_l - \mathbf{x}_m)}{\hat{f}^{\frac{1}{2}}_{n,h_n}(\mathbf{x}_l)\hat{f}^{\frac{1}{2}}_{n,h_n}(\mathbf{x}_m)}$

- We design a pairwise clustering algorithm to minimize the generalization error bound (1).