Nonparametric Maximum Margin Similarity for Semi-Supervised Learning

Yingzhen Yang, Xinqi Chu, Zhangyang Wang, Thomas S. Huang

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

Introduction

- Label Propagation has been proven to be effective for semi-supervised learning problems.
  - It encourages local smoothness of the labels in accordance with the similarity graph.
- We relate Label Propagation to a novel nonparametric **Maximum Margin Similarity** framework, with the concept of similarity margin.
  - The formulation of LP becomes a special case when the separation parameter is sufficiently large.
  - Theoretical guarantee: the hinge loss of similarity margin is the upper bound for the expected hinge loss of a linear classifier in a transformed space.
  - Semi-supervised learning algorithm: Maximum Margin Similarity Graph

Maximum Margin Similarity

- Given the data $\{\mathbf{x}_i\}_{i=1}^n \subseteq \mathbb{R}^d$, where the first l points are labeled with $y_i \in \{1, -1\}$ for $i = 1 \dots l$, and the labels of $\{\mathbf{x}_i\}_{i=l+1}^n$ are missing.
- Label Propagation with the given similarity matrix W:

$$\min_f \quad \sum_{i,j=1}^{n} W_{ij}(f_i - f_j)^2$$
$$s.t. \quad f_i = y_i, i = 1 \dots l$$

- Label Propagation encourages similar data to have similar labels and has a harmonic solution.
- Relating Label Propagation to a new Maximum Margin Similarity framework

Maximum Margin Similarity

- The similarity margin of the datum $\mathbf{x} \in \mathbb{R}^d$ is defined as the difference of sum of $\mathbf{x}$'s similarity to the data with the same label as $\mathbf{x}$, and the sum of $\mathbf{x}$'s similarity to the data with different label:

$$\gamma_{\mathbf{x}} = \frac{1}{n}\Big(\sum_{j:y_j=y(\mathbf{x})} S(\mathbf{x}, \mathbf{x}_j) - \sum_{j:y_j \neq y(\mathbf{x})} S(\mathbf{x}, \mathbf{x}_j) \Big)$$

  where $y(\mathbf{x})$ is the label of $\mathbf{x}$.

- The hinge loss of the similarity margins is

$$H_{\gamma,\mathcal{D}} = \frac{1}{n} \sum_{i=1}^{n} \max\{0, 1 - \frac{\gamma_i}{\gamma}\}$$

  and γ is the separation parameter.

Maximum Margin Similarity

- Theoretical Guarantee: the hinge loss of similarity margin is the upper bound for the expected hinge loss of a linear classifier in a transformed space.

Theorem

Define the mapping $F_{\mathcal{D}}(\mathbf{x}) = \frac{1}{\sqrt{n}}(S(\mathbf{x},\mathbf{x}_1), S(\mathbf{x},\mathbf{x}_2), \ldots, S(\mathbf{x},\mathbf{x}_n))$. *For* $\delta_1, \delta_2, \delta_3 > 0$, *with probability at least* $1 - \delta_1 - \delta_2 - \delta_3$ *over the data* $\mathcal{D}$, *there exists a linear classifier in the transformed space induced by* $F_{\mathcal{D}}$ *such that this classifier has hinge loss at most* $H_0 = H_{\gamma,\mathcal{D}} + \frac{\sqrt{\frac{2}{n}\log\frac{n}{\delta_1}}}{\gamma} + \sqrt{\frac{2}{n\gamma^2}\log\frac{1}{\delta_2}} + \delta_3(1 + \frac{1}{\gamma})$ *with respect to the margin* γ. *Namely, there exists a vector* $\boldsymbol{\beta} \in \mathbb{R}^n$ *such that* $\mathbb{E}_{(\mathbf{x},y)\sim P}\left[\max\{0, 1 - \frac{y\langle\boldsymbol{\beta}, F_{\mathcal{D}}(\mathbf{x})\rangle}{\gamma}\}\right] \le H_0$.

- We design a new Maximum Margin Similarity Graph for semi-supervised learning, which minimizes the hinge loss of the similarity margin.