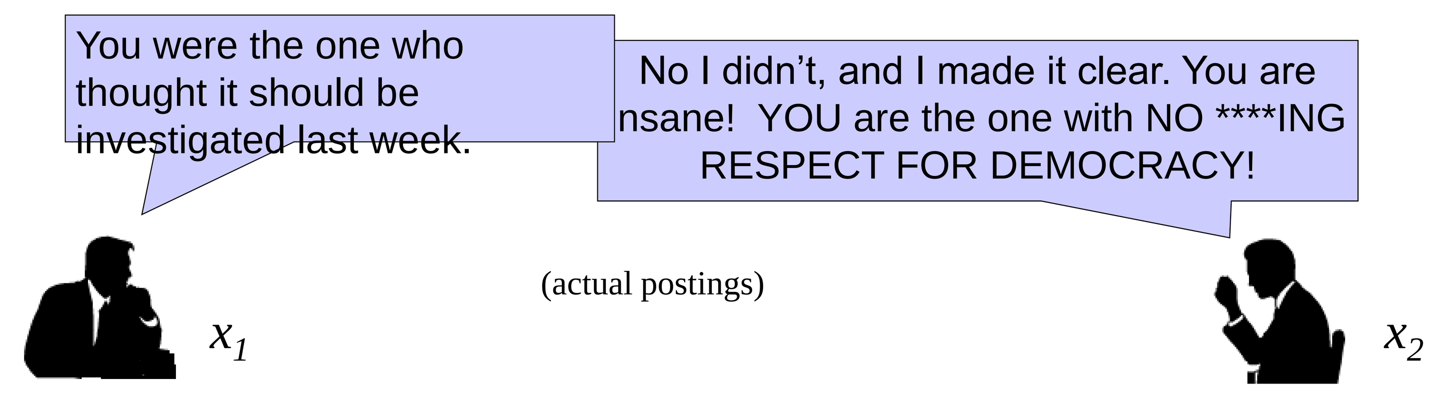
# Dissimilarity in Graph-Based Semi-Supervised Classification


Andrew B. Goldberg, Xiaojin Zhu, Stephen Wright

Computer Sciences Department, University of Wisconsin, Madison, WI, USA. {goldberg, jerryzhu, swright}@cs.wisc.edu


Example: Predict political party from Web blogs




They disagree. ➔ $y_1 \neq y_2$, known as **cannot-links** in clustering.

Our contribution: A convex formula that incorporates both cannot-links and must-links for binary and multiclass classification.

## Binary Classification

Existing graph-based semi-supervised learning requires a graph $W$

- For example, a kNN graph over data points
- $w_{ij}$ is the edge weight between $x_i$ and $x_j$
- Discriminant $f$ regularized by $\frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$.
- Can be written as $\mathbf{f}^\top \mathcal{L}\mathbf{f}$
- $w_{ij}$ is essentially must-links in clustering

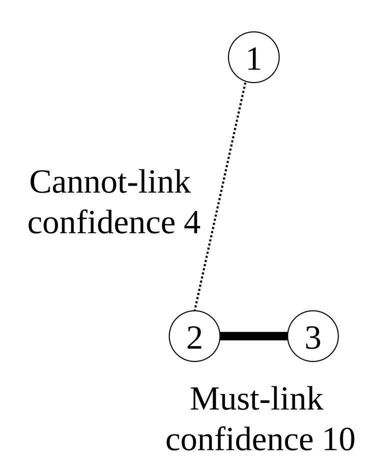
We want to add cannot-links.

Things that do not work:
- Small or zero $w$: no-link instead of cannot-link
- Negative $w$: unbounded solution; non-convex problem

**Our solution**: encode cannot-links between $x_i$ and $x_j$ as

$$w_{ij}(f(\mathbf{x}_i) + f(\mathbf{x}_j))^2.$$

Both cannot-links and must-links can be represented by a mixed graph, where each edge has two variables $s_{ij}$ (1 if must-link, -1 if cannot-link) and $w_{ij}$ (confidence, non-negative).




$$S = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 0 & -1 & 0 \\ 2 & -1 & 0 & 1 \\ 3 & 0 & 1 & 0 \end{array} \qquad W = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 0 & 4 & 0 \\ 2 & 4 & 0 & 10 \\ 3 & 0 & 10 & 0 \end{array}$$

The new regularizer is

$$\mathbf{f}^\top \mathcal{M}\mathbf{f} = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(f(\mathbf{x}_i) - s_{ij}f(\mathbf{x}_j))^2.$$

The "mixed graph Laplacian" is

$$\mathcal{M} = \mathcal{L} + (\mathbf{1} - S) \bullet W,$$

where $M$ is positive semi-definite, and reverts to the standard graph Laplacian $L$ if there are no cannot-links.

The *convex* binary classification problem is

$$\min_{f \in \mathcal{H}} \sum_{i=1}^{l} c(y_i, f(\mathbf{x}_i)) + \lambda_1 \|f\|_{\mathcal{H}}^2 + \lambda_2 \mathbf{f}^\top \mathcal{M}\mathbf{f}.$$

with any convex loss function $c()$.

## Multiclass Classification

It is not trivial to incorporate cannot-links into multiclass semi-supervised classification.

Things that do not work:
- 1-vs-rest: cannot-links become must-links in "rest."
- 1-vs-1: cannot determine which unlabeled points to participate.
- Warped kernel in multiclass kernel machine.

We use Lee, Lin & Wahba (2004) multiclass SVM encoding, which encodes $y=j$ in a $k$ class problem as the zero-sum vector

$$\begin{array}{c} 1 \\ \vdots \\ j \\ \vdots \\ k \end{array} \begin{bmatrix} -\frac{1}{k-1} \\ \vdots \\ 1 \\ \vdots \\ -\frac{1}{k-1} \end{bmatrix}$$

If we want a cannot-link between $x_s$ and $x_t$, the "good" and "bad" $y$'s, when summed up, are

"good" $y_s \neq y_t$

$$\begin{bmatrix} -\frac{1}{k-1} \\ 1 \\ -\frac{1}{k-1} \\ \vdots \\ -\frac{1}{k-1} \end{bmatrix} + \begin{bmatrix} -\frac{1}{k-1} \\ \vdots \\ -\frac{1}{k-1} \\ 1 \\ -\frac{1}{k-1} \end{bmatrix} = \begin{bmatrix} -\frac{2}{k-1} \\ \frac{k-2}{k-1} \\ -\frac{2}{k-1} \\ \frac{k-2}{k-1} \\ -\frac{2}{k-1} \end{bmatrix}$$

"bad" $y_s = y_t$

$$\begin{bmatrix} -\frac{1}{k-1} \\ \vdots \\ 1 \\ \vdots \\ -\frac{1}{k-1} \end{bmatrix} + \begin{bmatrix} -\frac{1}{k-1} \\ \vdots \\ 1 \\ \vdots \\ -\frac{1}{k-1} \end{bmatrix} = \begin{bmatrix} -\frac{2}{k-1} \\ \vdots \\ 2 \\ \vdots \\ -\frac{2}{k-1} \end{bmatrix}$$

So we do not want any element in $f(x_i) + f(x_j)$ larger than $(k-2)/(k-1)$. This is achieved with the convex regularizer

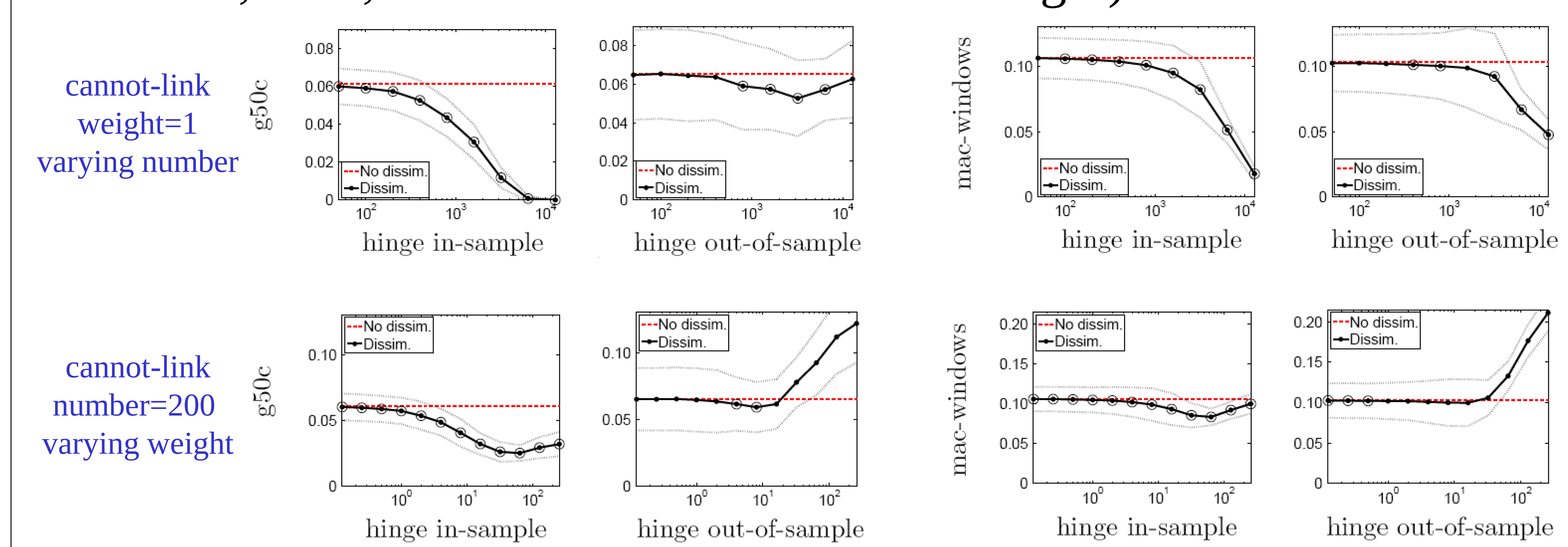
$$\sum_{(s,t)\in\mathcal{D}} \sum_{j=1}^{k} \left( f_j(\mathbf{x}_s) + f_j(\mathbf{x}_t) - \frac{k-2}{k-1} \right)_+^p ,$$

The convex multiclass SVM classification problem is

$$\begin{aligned} \min \quad & \tfrac{1}{l}\textstyle\sum_{i=1}^{l} L_{i\cdot}(\mathbf{f}(\mathbf{x}_i) - \mathbf{y}_i)_+ + \lambda_1 \sum_{j=1}^{k} \|h_j\|_{\mathcal{H}}^2 \\ & + \tfrac{\lambda_2}{|\mathcal{D}|} \textstyle\sum_{(s,t)\in\mathcal{D}} \sum_{j=1}^{k} \left( f_j(\mathbf{x}_s) + f_j(\mathbf{x}_t) - \frac{k-2}{k-1} \right)_+^p \\ \text{s.t.} \quad & \textstyle\sum_{j=1}^{k} f_j(\mathbf{x}_i) = 0, \quad i = 1 \cdots n, \end{aligned}$$

## Experiments

**Binary classification, oracle cannot-links** (g50c, mac-window, $n \approx 1000$, $l=50$, kNN must-links with RBF weight)




**Multiclass classification, oracle cannot-links** (USPS, $n \approx 2000$, $k=10$, $l=50$)

| Dissim. | Overall | In-sample | Out-of-sample |
|---|---|---|---|
| baseline 0 | 24.48 | 24.48 | 24.48 |
| 10 | 24.41 | 20.47 | 24.40 |
| 20 | 24.32 | 23.53 | 24.33 |
| 40 | 24.27 | 24.17 | 24.27 |
| 80 | **23.96** | 23.57 | **23.99** |
| 160 | **23.63** | 24.49 | **23.48** |
| 320 | **23.30** | 23.57 | **23.20** |

**Binary classification, real cannot-links** (politics.com, $n=184$, $l=50$)

Cannot-link(A,B) if twice, A or B quotes the other, and text next to quote has ??, or !!, or ALL CAPS.

| Classifier | Base error rate | SSL error rate | Δ |
|---|---|---|---|
| SVM | 45.67 ± 3.28 | 40.15 ± 4.95 | 5.5% |
| RLS | 45.60 ± 3.94 | 37.99 ± 1.88 | 7.6% |