Laplacian SSL: The Limit of Infinite Unlabeled Data

Boaz Nadler (Weizmann), Nati Srebro (TTI-C), Xueyuan Zhou (UChicago)

Semi-Supervised Learning with Graph Laplacian Regularization:

[Zhu Ghahramani Lafferty 03],**[D.Zhou** *et al* **04]**,**[Bengio** *et al* **05]**,…

$$\min_f \overbrace{\sum_{\text{all } i,j} W_{i,j}(f(x_i)-f(x_j))^2}^{\text{Laplacian regularization term}} + \overbrace{\sum_{\text{labeled } i} \text{loss}(f(x_i);y_i)}^{\text{data fit term}}$$

$$W_{i,j} = \exp\left(-\frac{||x_i-x_j||^2}{\sigma^2}\right)$$

With #labeled fixed, as #unlabeled$\to\infty$, we would hope things get better…

BUT: We show that as #unlabeled$\to\infty$,
solution degenerates to an uninformative constant!

- Theoretical analysis + empirical examples demonstrating the phenomenon
- In 1D (and *only* in 1D): Things are OK, and at the limit we get RKHS regularization with sensible density-dependent kernel
- Also: What about related Laplacian Eigenmaps method? Does it also break down when #unlabeled$\to\infty$? What is it doing? (Answers at poster)