Probabilistic Lipschitzness
A niceness assumption for deterministic labels

Ruth Urner and Shai Ben-David

Learning Faster From Easy Data Workshop @ NIPS 2013

December 10, 2013

Theory/Practice discrepancy

Standard learning theory:
Evaluate performance as worst case over all data generating distributions
Phenomenon:
Real world data is more benign than worst-case scenarios
Need:
Adapt theoretical models: Identify realistic data properties and take advantage of such properties

This theory/practice discrepancy is particularly apparent in learning setting that employ unlabeled data.

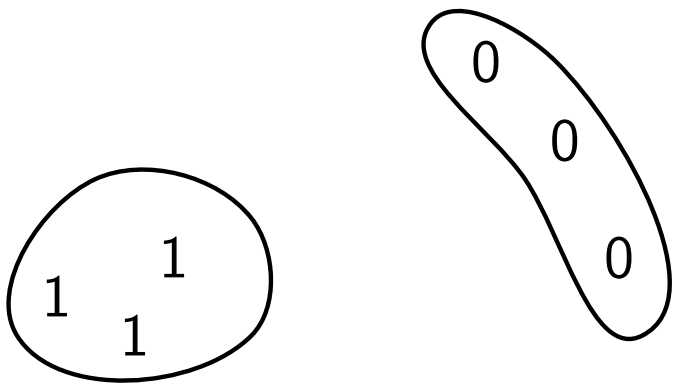
It's not all about bounding noise

Popular niceness assumption: Tsybakov noise condition

However:
A deterministic labeling function does not render learning easy!

Probabilistic Lipschitzness (PL) - Intuition

Intuition: Close points are likely to have the same label.



- Cluster assumption often cited to account for use of unlabeled data
- PL is a relaxation and formalization of this assumption

Probabilistic Lipschitzness

- Captures assumption that is implicit in many Machine Learning paradigms
- Is meaningful measure of niceness for distributions with deterministic labellings
- Yields faster rates for non-parametric (Nearest Neighbor) Learning
- Yields provable label savings in Semi-supervised and Active Learning

See you at the poster!