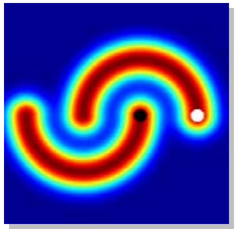
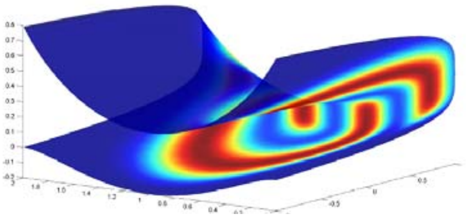
Learning the intrinsic structure from labeled & unlabeled data





Apr. 10, 2009 , Spring CS Seminar

Chunsheng Fang (Victor)

Advisor: Prof. Anca Ralescu

CS Dept, Univ. of Cincinnati

University of Cincinnati

A little bit about me

- 2nd year PhD, RA w/ Prof. Anca Ralescu in CS, UC
- CSGSA secretary (2008-2009)

VC-bir
Powered by Victor Fang, UC 2008
UC bir

- Current research interest:
  - Machine learning
    - Similarity Measure (Probability based)
    - Image Retrieval & Annotation (CBIR, etc)

- Attended projects before 2007:
  - Computer Vision & Pattern Recognition
    - Video object tracking
    - Multiple object classification

www.VictorFang.com

UNIVERSITY OF Cincinnati

Major references

- [1] Dengyong Zhou, Olivier Bousquet, Thomas N. Lal, Jason Weston, Bernhard Schoelkopf , "Learning with Local and Global Consistency"
- [2] Dengyong Zhou, Jason Weston, Arthur Gretton, Olivier Bousquet, Bernhard Schoelkopf , "Ranking on Data Manifolds“
  - Above 2 are from NIPS (Neural Information Processing Systems) 2003

Neural Information
Processing Systems
Conference

- [3] Jerry Zhu, Semi-supervised learning survey, UW-Madison, 2008

- [4] Matthias Hein and Ulrike von Luxburg, “Introduction to Graph-based Semi-supervised Learning”, MLSS 2007

Roadmap

Part 1: Intro to Machine Learning

- Machine Learning family
- Machine Learning researchers
- Examples

Part 2: Learning the intrinsic structure from labeled & unlabeled data

- Semi-supervised learning (SSL)
- Data manifold
- Graph theory
- Algorithm: Local & global consistency
- Applications

Part 3: Conclusions

Note: This talk conveys only conceptual ideas, not mathematically rigor enough. Read references for details.

Part 1: Intro to Machine Learning

- Machine Learning family
- Machine Learning researchers
- Examples

Machine Learning family

- Subfield of AI
- “Automatically produce models from data”

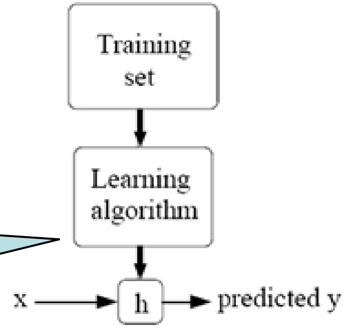


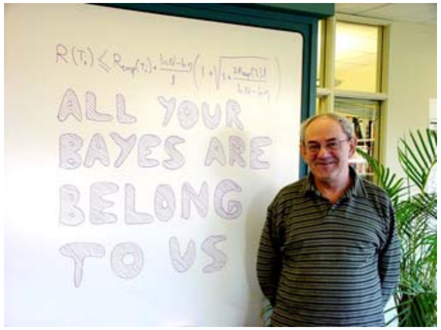
Figure by Andrew Ng, Stanford

- Supervised learning
  – Learn from labeled data
- Unsupervised learning (clustering, feature extraction...)
  – Learn from unlabeled data
- Semi-supervised learning
  – Learn from both labeled & unlabeled data!

- More: Reinforcement learning, Transduction, etc.

Machine Learning researchers

- The author: Dengyong Zhou
  - Microsoft Research, Redmond.



- Vladimir Vapnik
  - NEC Lab(2002 - now), AT&T Lab
  - Statistical Learning Theory, SVM

- Tom Mitchell
  - Chair of ML Dept, CMU
  - ML Textbook , fMRI human brain

- Michael I. Jordan
  - UC Berkeley
  - Beyesian networks, EM, linking ML with statistics

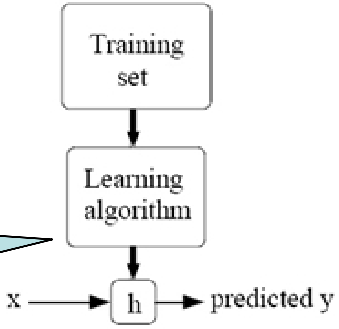
- A lot more… (Research labs, Universities, etc)

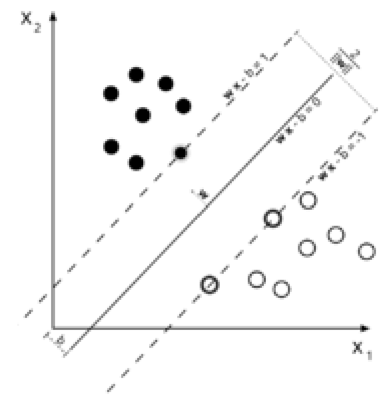
- Related keyword:
  - Statistical learning
  - Pattern recognition
  - Computer vision, ….

Supervised learning : examples

- Object classification, Face recognition
  - Require a lot of labeled data (training set)
  - Learn the optimal model from the training set
  - Use the learned model to classify unknown data
- Problems…
  - A practical system have to label >100,000 training samples, depending on application.
  - Expensive, tedious to label data….

Figure by Andrew Ng, Stanford

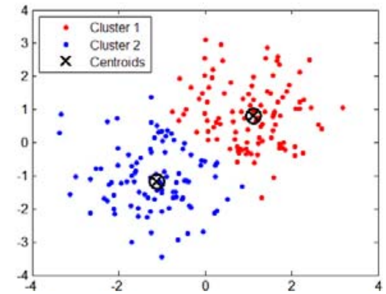




UNIVERSITY OF Cincinnati

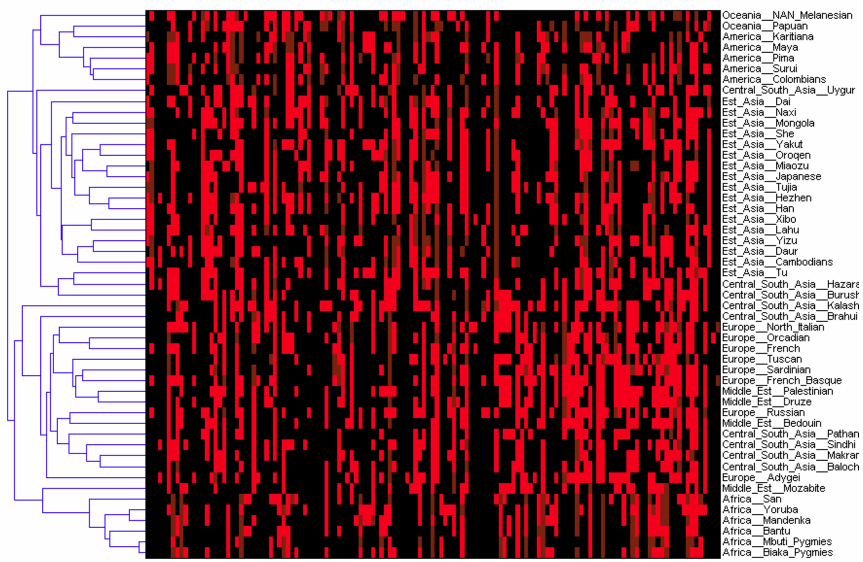
Unsupervised learning : examples

- Find clusters in unlabeled data



- My final project in Comp Genomics (Dr. Cheng):
- Hierarchical clustering SNP data to explore human race migration

- More details: www.VictorFang.com



UNIVERSITY OF Cincinnati

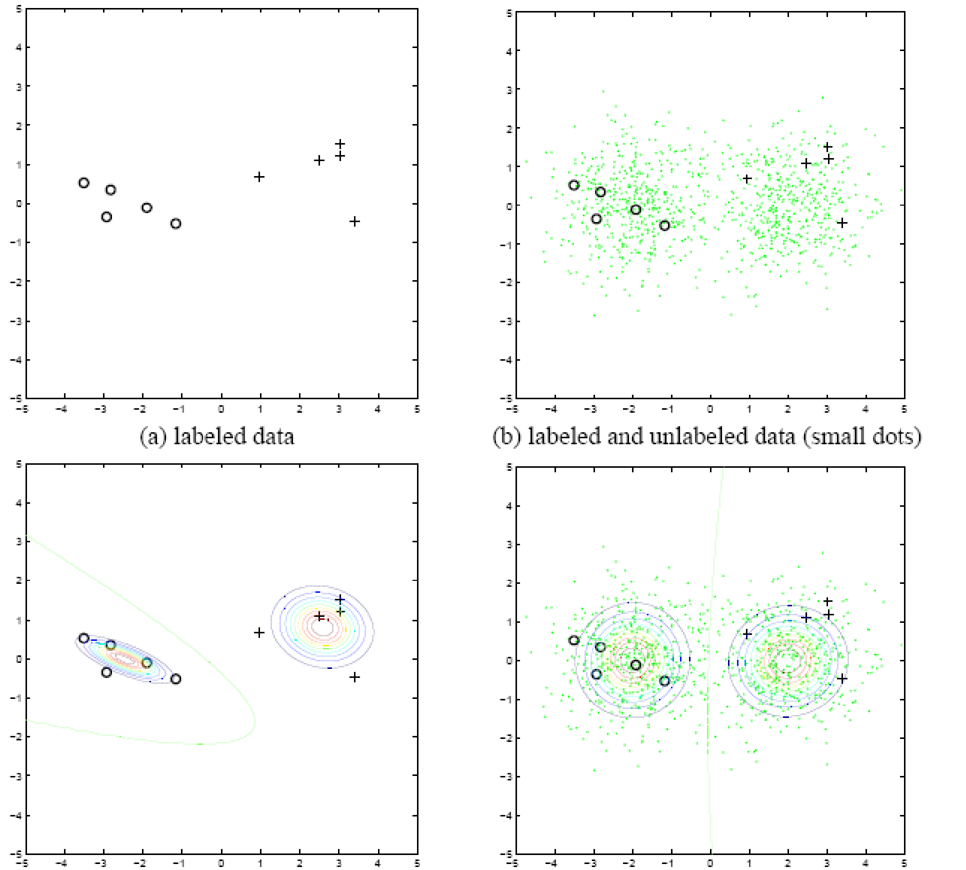
Part 2: Learning the intrinsic structure from labeled & unlabeled data

- Semi-supervised learning (SSL)
- Data manifold
- Construct manifold using graph theory
- Algorithm: Local & global consistency
- Applications

Semi-supervised Learning (SSL)

- Why?
  - Human labels can be expensive and time consuming;
  - a lot of unlabeled data around us;
  - The knowledge about the unlabeled data should be helpful.

- Intuitive idea:
  - From Dr. Jerry Zhu's SSL survey, UW-Madison, 2008



(a) labeled data

(b) labeled and unlabeled data (small dots)

(c) model learned from labeled data

(d) model learned from labeled and unlabeled data

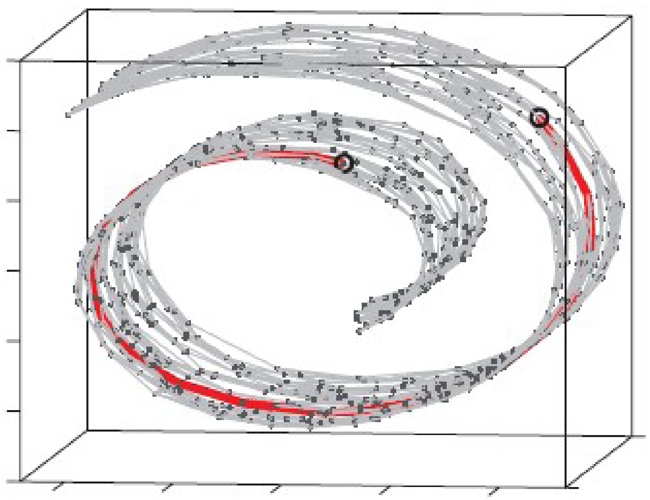
UNIVERSITY OF Cincinnati

Consistency assumption

- The key to SSL is the prior assumption of consistency:
  - (1) nearby points are likely to have the same label; (local)
  - (2) points on the same structure (typically referred to as a cluster or a manifold) are likely to have the same label. (global)

Data Manifold

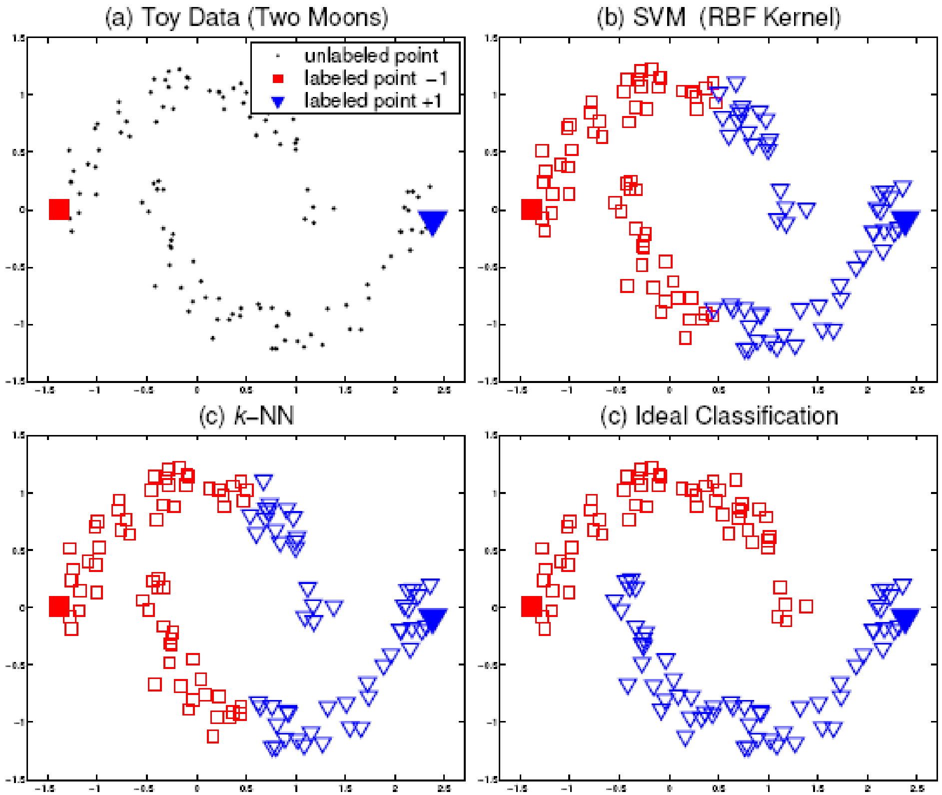
- Definition of manifold (Wolfram MathWorld):
  - A manifold is a topological space that is locally Euclidean
  - (i.e., around every point, there is a neighborhood that is topologically the same as the open unit ball in R^n).
- A friendly illustration:



- A more friendly illustration: Mahjong seat pad!

UNIVERSITY OF Cincinnati

When Euclidean distance fails…

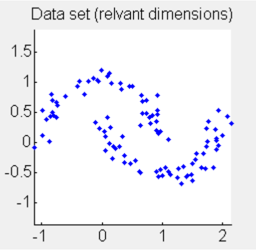


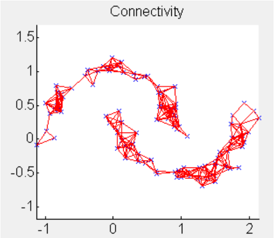
- For ideal classification, we need manifold!
- Transform problem into graph domain!!

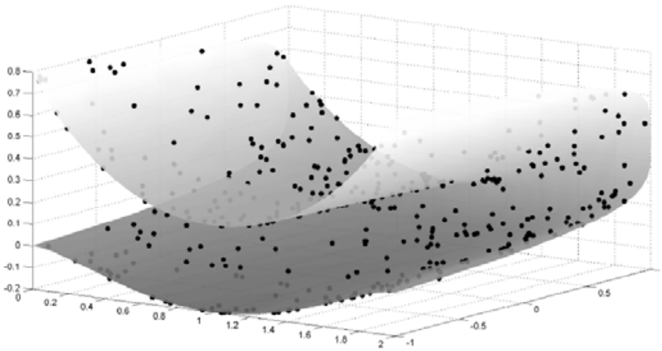
Construct manifold using graph theory

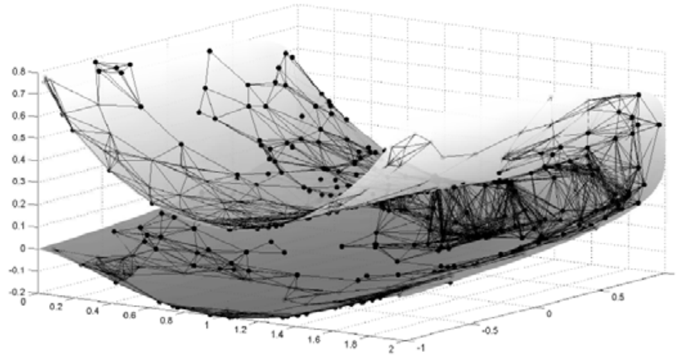
- Transform data points in Euclidean space
- Similarity graph : nodes with weighted edges (only) linking to nearest neighbor !

- Graph can be expressed as adjacency matrix (affinity matrix)
- Actually it belongs to Spectral graph theory.
- Examples:
  - 2D data
  - 3D data









UNIVERSITY OF Cincinnati

Problem formulation

Given a point set $\mathcal{X} = \{x_1, \ldots, x_l, x_{l+1}, \ldots, x_n\} \subset \mathbb{R}^m$ and a label set $\mathcal{L} = \{1, \ldots, c\}$, the first l points $x_i (i \leq l)$ are labeled as $y_i \in \mathcal{L}$ and the remaining points $x_u (l+1 \leq u \leq n)$ are unlabeled. The goal is to predict the label of the unlabeled points.

Notations:

- Y, nxc binary matrix: initial label information

0	1	0
0	0	1
1	0	0
0	0	0
0	0	0

- F(t), nxc labeling confidence matrix at step t
- α, between [0,1], the relative amount of the information from its neighbors and its initial label information.

Consistency Algorithm

1. Construct the affinity matrix *W* defined by a Gaussian kernel:

$$w_{ik} = \begin{cases} \exp(-\|x_i - x_k\|^2 / 2\sigma^2), & \text{if } i \neq k \\ 0, & \text{if } i = k \end{cases}$$

2. Normalize *W* symmetrically by $S = D^{-1/2} W D^{-1/2}$,

   where D is a diagonal matrix (degrees) with $D_{ii} = \sum_k W_{ik}$.

Spectral Graph Theory

each data point receive “confidence of label info” from its neighbors

retains the initial labels, to guarantee moving in correct direction

3. Iterate $F(t+1) = \alpha S F(t) + (1-\alpha) Y$ until converge.

Trust your nearest neighbors!

4. Let F^* denote the limit of the sequence {*F*(*t*)}. The classification results is :

$$\text{Labeling } x_i \text{ as } y_i = \operatorname{argmax}_{j \le c} F^*_{ij}$$

University of Cincinnati

Convergence on Consistency Algorithm

- From the iteration equation,

$$F(t+1) = \alpha SF(t) + (1-\alpha)Y$$

$$F(0) = Y$$

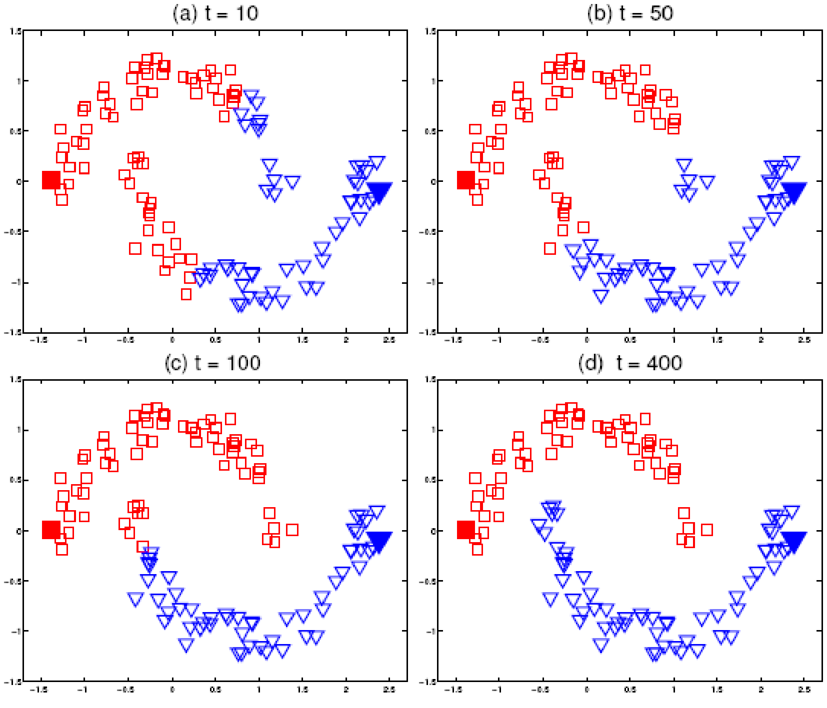
- it is very easy to show that: **(Nice!)**

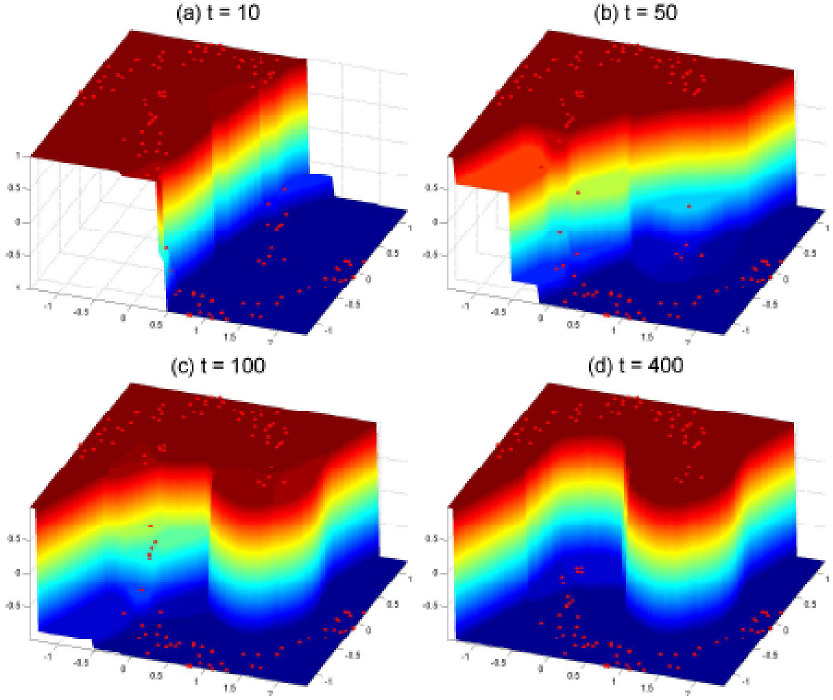
$$F^* = \lim_{t \to \infty} F(t) = (I - \alpha S)^{-1} Y$$

- Therefore:
  - For small-scale problem: we could compute *F** directly. (One-shot)
  - For large-scale problem: we prefer iterate until converges. (like Google)

UNIVERSITY OF Cincinnati

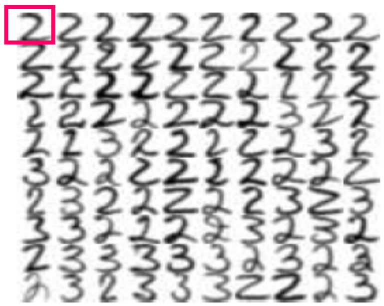
Iteration converges



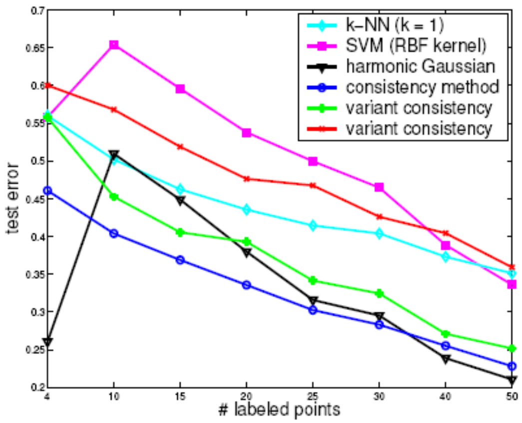


Applications:

- Digit Recognition
  - Euclidean: Proposed:



- Text Classicification
  - On “20 news group” dataset



- Can be applied to more applications…
  - Image retrieval
  - Image annotation

UNIVERSITY OF Cincinnati

Conclusions

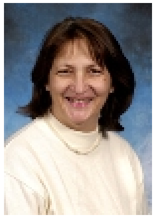
- SSL can make use of both labeled & unlabeled data
- Intrinsic data manifold can be modeled and learned using graph theory

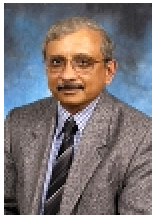
Tips:

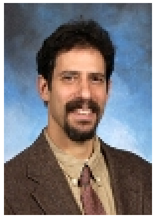
- Well formulate the research problem!
- Transform your problem domain , maybe helpful !

More important…

- Talk to our professors!




……

UNIVERSITY OF Cincinnati

Welcome on board in ML world

Questions??

CS/ECE GSA coming Events

- **04/11/2009 - Research, Jobs & a Movie**
  - **427 ERC** , 10am-5pm
- **04/17/2009 - 4th Annual ECE/CS Poster Competition**
- http://cs.uc.edu/~csgsa/

UNIVERSITY OF Cincinnati