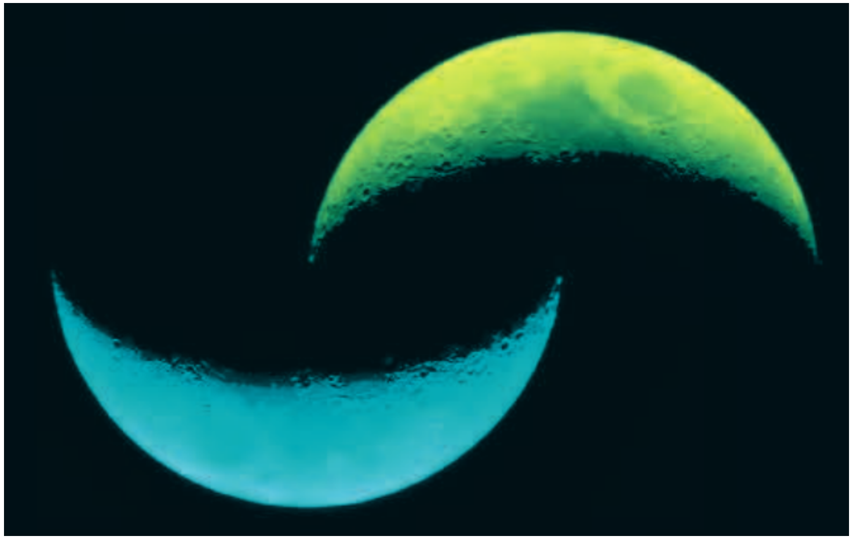
Semi-Supervised Learning

Machine Learning Meetup Aachen
Andrei Ionita
5th September, 2017



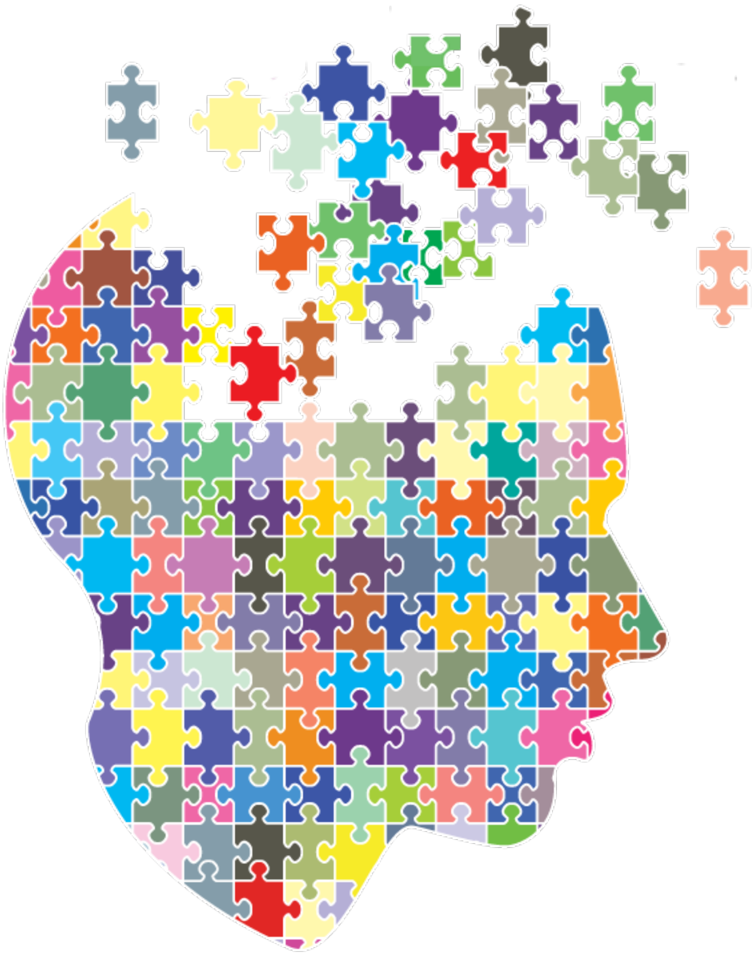
digitalHUB
Aachen

Start-ups


SONAH
DRIVE
HOME
NOT
CIRCLES
WELCOME TO A NEW ERA,
WHERE PARKING LOTS WERE FOUND,
NOT SEARCHED FOR.
START THE SHOW

Overview

1. What and why SSL?
2. Definition
3. Self-Training
4. Co-Training
5. Graph-based learning
6. SSL in Nature
7. Conclusion



What is it?

- Between Supervised and Unsupervised ML
- Training Phase with
  - Examples $x_1, ..., x_l$ that have labels $y_1, ..., y_l$
  - unlabeled examples $x_{l+1}, ..., x_{l+u}$
  - $l << u$

- Take advantage of the unlabeled data in the Test Phase

Why we need it?

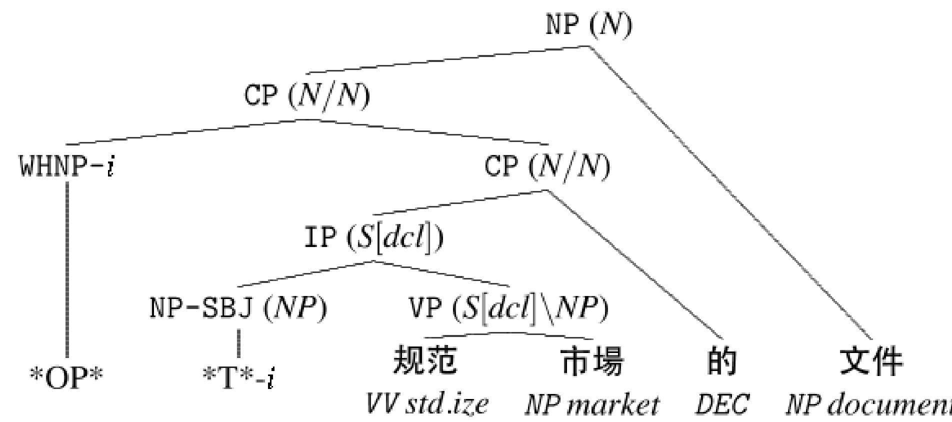
- Lots of available data that is unlabeled
  - Images on the web
  - Categories of documents, keywords, etc.

- Labeling is costly

- Unlabeled data is abundant and cheap

Applications

- Natural Language Parsing
  - Penn Chinese Treebank
  - 2 years for 4000 sentences
- Document classification
- Spam detection
- Handwriting detection
- Video/Audio recognition
- many others, more than you think



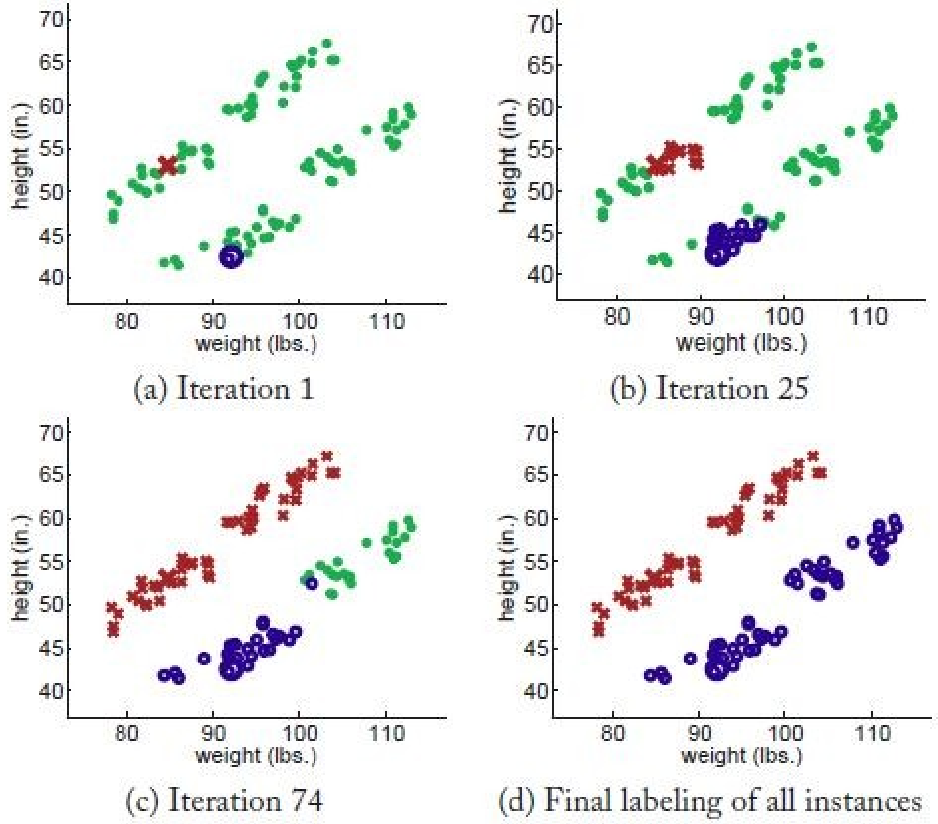
Transductive vs. Inductive

- Vapnik (70s) introduced
  - Transductive learning: infer the correct labels for the given unlabeled data
  - Inductive learning: infer the correct mapping X to Y
- Most SSL algorithms are inductive

Self-Training

1. Classifier C is trained with the labeled data
2. C is used to classify the unlabeled data
3. Add determined labeled data to the training set
   - Variations:
     - Add most confident labeled object
     - Add all labeled objects
     - Add all labeled objects, weighed by confidence
4. Back to step 1; repeat as long as feedback loop is positive

Self-Training (2)



Propagating 1-nearest-neighbor applied to the 100-little-green-alien data.

Self-Training: Problem

- A mistake can be used to reinforce itself

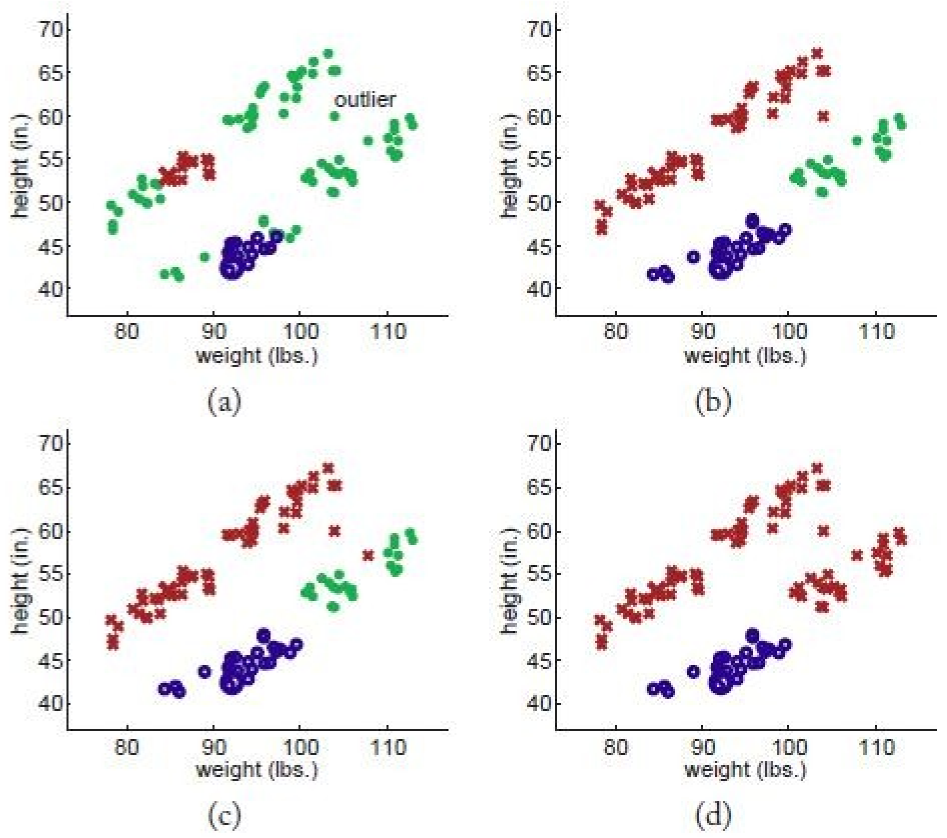
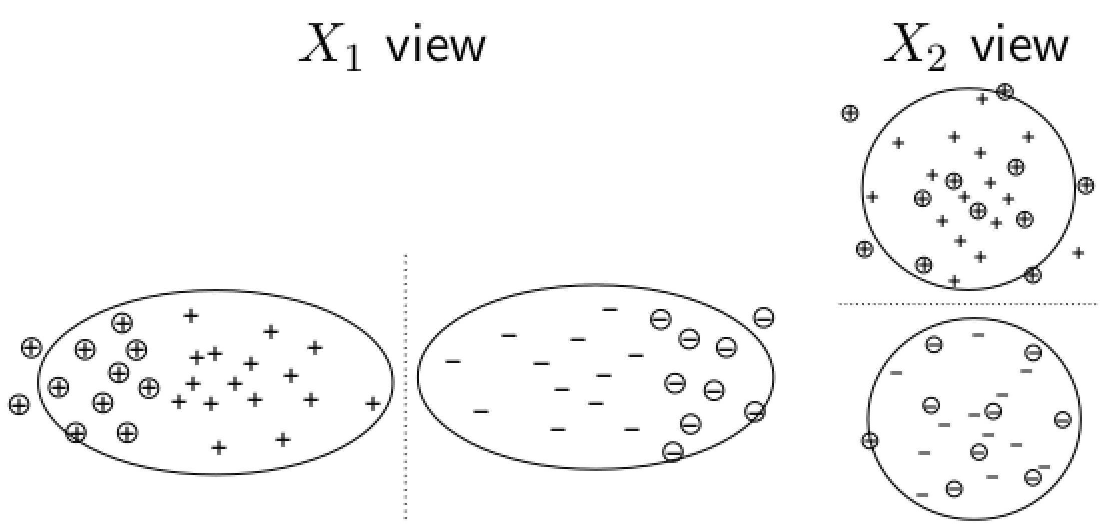


Figure 2.4: Propagating 1-nearest-neighbor illustration featuring an outlier: (a) after first few iterations, (b,c) steps highlighting the effect of the outlier, (d) final labeling of all instances, with the entire rightmost cluster mislabeled.

Co-Training

- Two different and independent perspectives
- Two features of an object
  - e.g. content and subject of emails
  - Task: spam detection



Co-Training (2)

1. Train two models independently on the same training pool
2. Determine and add most confident k emails for each
   - Add positive and negative labels for both models
   - Keep proportion of positive:negative labels
3. Go back to Step 1
4. Continue until all are labeled

The result surpasses the ones where each model is trained separately and applied on a test set.

Applications of Co-Training?

- Actor/Actress Recognition using Audio and Video sequence

Co-EM

1. Train two classifiers on features A and B
2. Probabilistically label all unlabeled data by applying A
3. Train B on A's tentatively labeled data, re-label all initially unlabeled (probabilistically)
4. Re-training A on B's tentatively labeled data, re-label all initially unlabeled (probabilistically)
5. Repeat until the classifiers converge

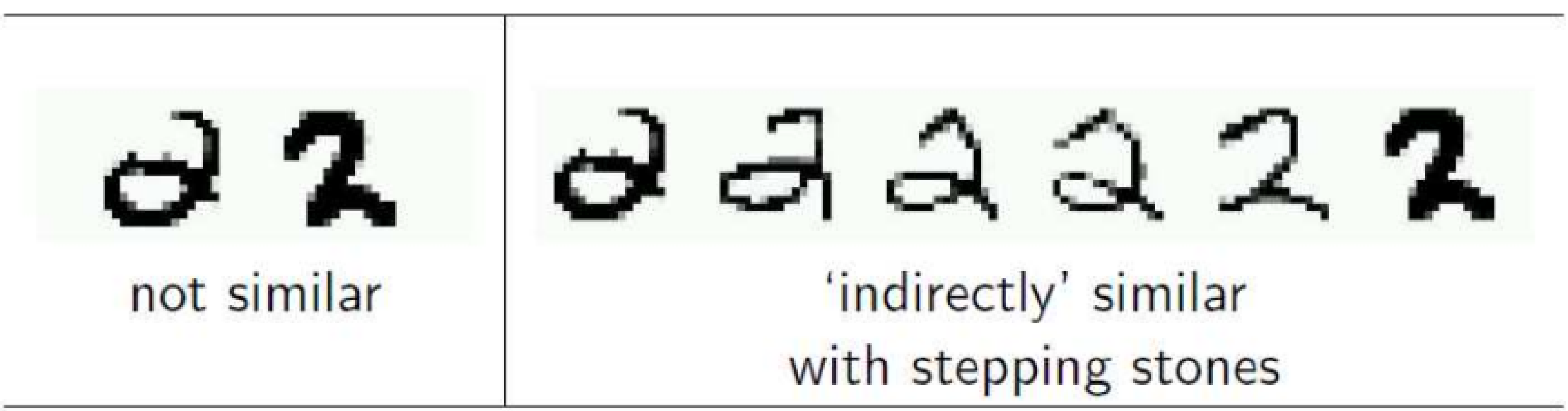
Co-EM improves on Co-Training

Extensions

- Even when the features split is random, results are better
- Find a split that achieves maximum independency!
- The algorithms are more robust against the underlying assumptions that the classifiers make
- Multiview: train multiple classifiers on different features
  - Classify unlabeled data with all classifiers
  - Add majority vote label

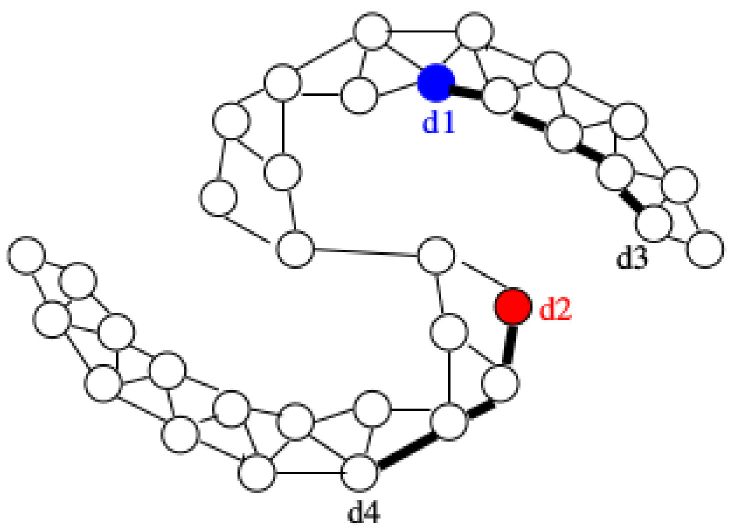
Graph-based regularization

- Represent each example (labeled/unlabeled) as vertex in a graph
- Nearby vertices should have similar labels (label smoothness property)



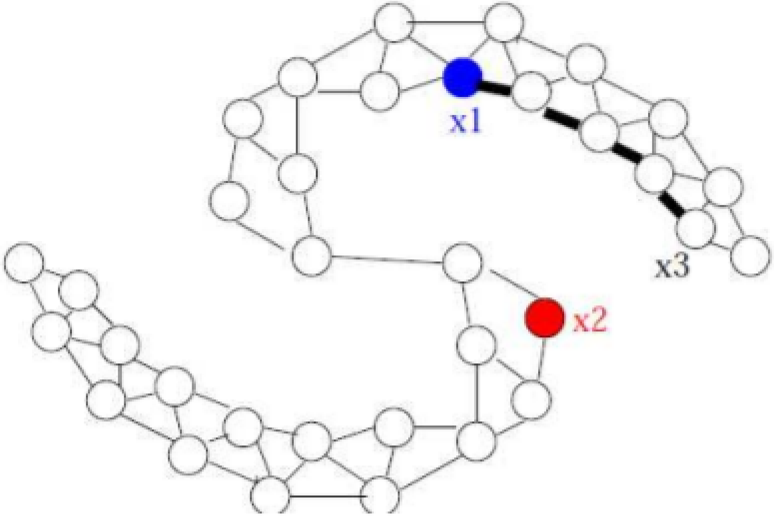
Graph-based learning

- Edges = similarity weights computed from features
- Heavy edges correspond to vertices that have the same label



Graph-based learning (2)

- Objective: calculate similarity on all paths
- Predict labels that
  - (1) minimize the loss on labeled data
  - (2) ensures the smoothness of labeled and unlabeled data

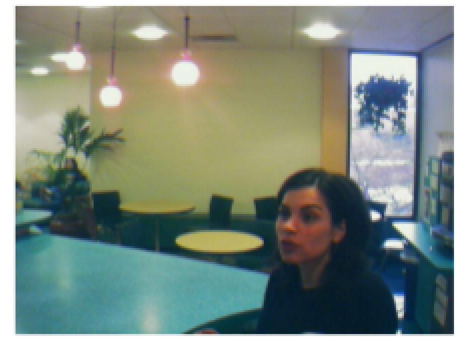


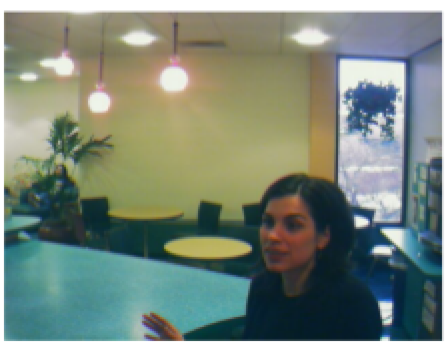
Graph-based learning (3)

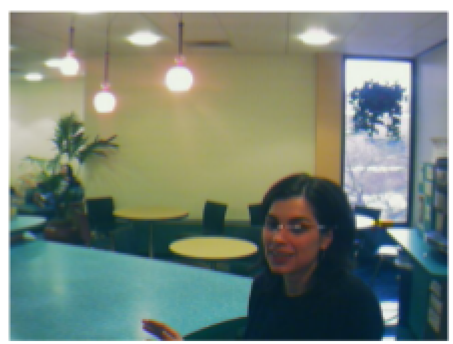
- Person identification: time, color, face edges

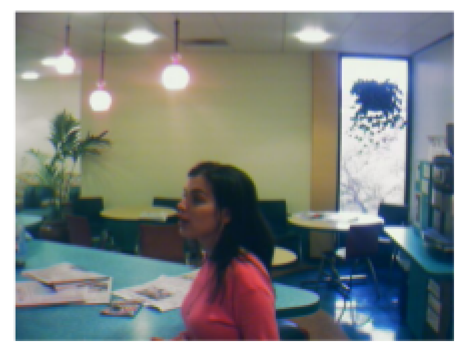

image 4005


neighbor 1: time edge


neighbor 2: color edge

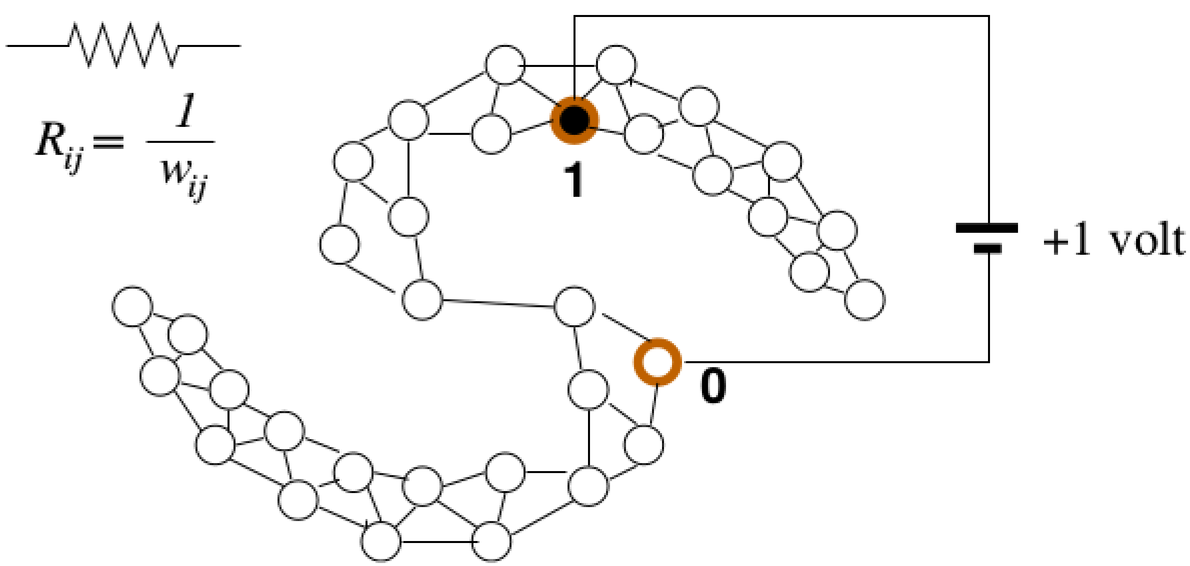

neighbor 3: color edge


neighbor 4: color edge


neighbor 5: face edge

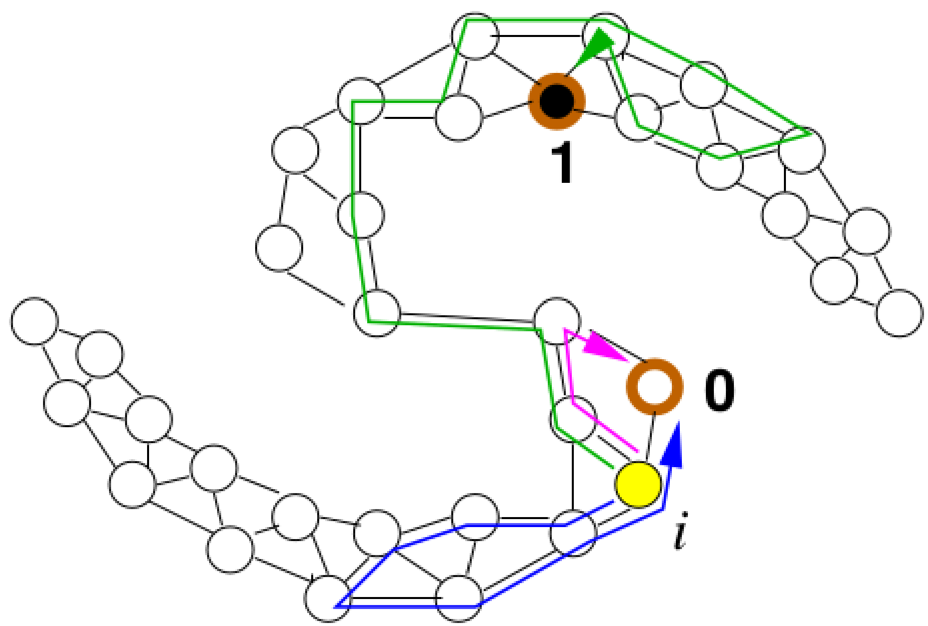
Graph-based learning (4)

- Electric network interpretation
- Edges as resistors with conductance w_{ij}
- 1 Volt battery connects to labeled points
- The voltage at the nodes are the labeled values



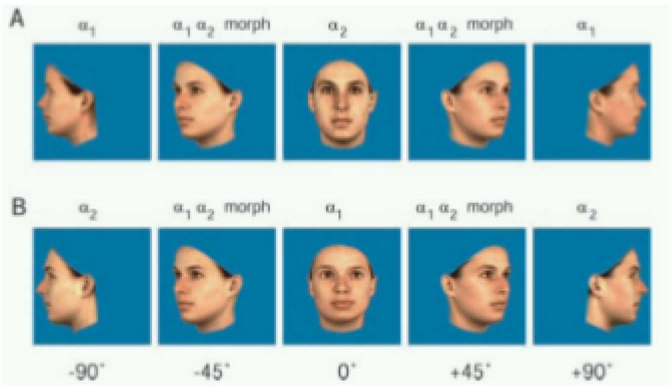
Graph-based learning (5)

- Random walk interpretation
- Randomly walk from node i to j with probability $\frac{w_{ij}}{\sum_k w_{ik}}$
- Stop when arriving at a labeled node



SSL in Nature

- Visual Recognition with temporal association
- A human can recognize a face in motion, even if she knows it from two angles
- However: transforming a face A to another face B by artificially rotating goes unnoticed



Conclusion

- Unlabeled data is everywhere
- SSL combines labeled and unlabeled data
- It achieves a better performance than
  - Supervised machine learning
  - Clustering
- Widely used
  - Document/Image classification
  - Video/Audio recognition
  - Handwriting recognition

References and Image Sources

- Semi-Supervised Learning Tutorial, Xiaojin Zhu, ICML 2007
- Semi-Supervised Learning, Piyush Rai, 2011, University Utah
- Data Mining – Practical Machine Learning , Witten, Frank, Hall
- Semi-Supervised Learning, Chapelle, Schölkopf, Zien
- Semi-Supervised Learning Literature Survey, Xiaojin Zhu, 2008

...Thank you!