Welcome to

DS3010:
DS-III: Computational Data Intelligence
Semi-supervised Learning

Prof. Yanhua Li

Time: 11:00am – 12:50pm M & R
Location: HL 114
D-term 2022

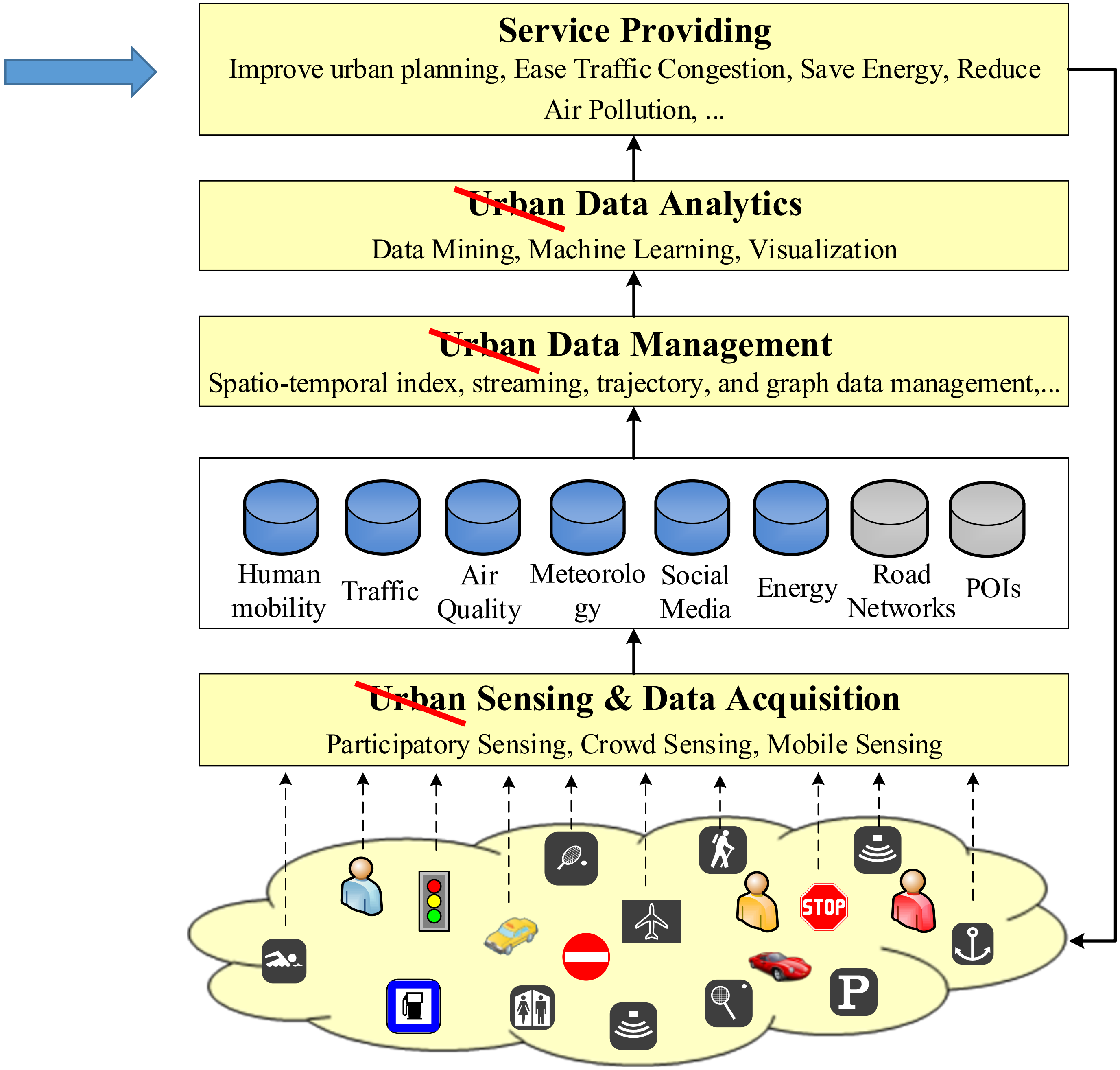
Class arrangement

- **Week 6 (4/18 M):** No Class. Patrios' Day Holiday.
  Note: Project 2 is due.
- **Week 6 (4/21 R):** Unsupervised Learning, Transfer Learning (slides), Review for final exam.
  Note: Project 2 presentations (two students).
- **Week 7 (4/25 M):** Final exam.
- **Week 7 (4/28 R):** Reinforcement Learning.
- **Week 8 (5/2 M):** Project 3 presentations (Zoom or in-person?; We will take a survey in Canvas.)
  Note: Project 3 is due.

Project 2

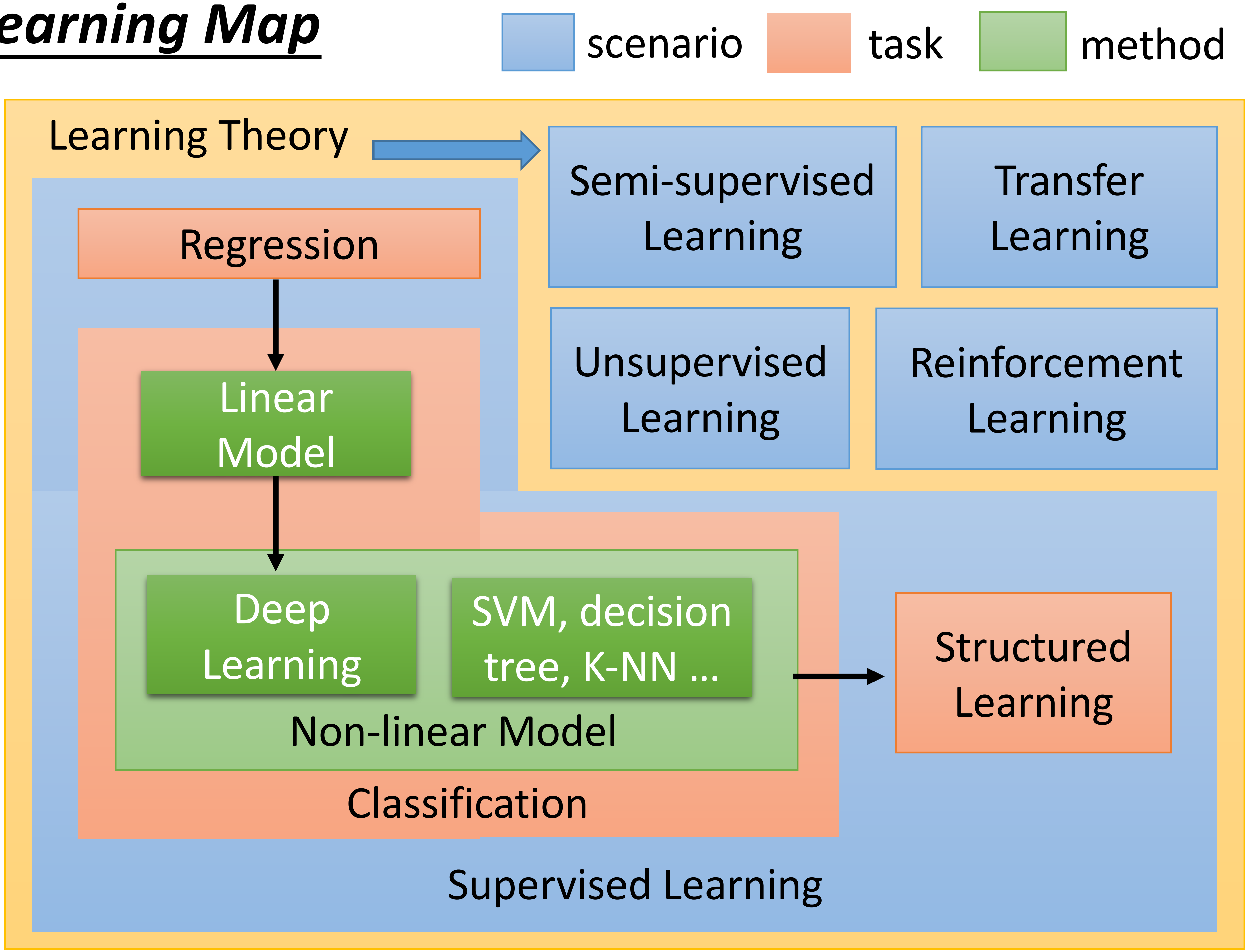
- Implement and compare SVM, Logistic regression, and Multi-layer perceptron (MLP) on the mobile phone data.
  - somemodel.score(x,y) for accuracy evaluation.
- For MLP, you also need to compare different structures, using hidden_layer_sizes parameter.
  - hidden_layer_sizes is the parameter tuple of the MLP classifier.
  - hidden_layer_sizes = 25,11,7,5,3, that means 5 hidden layers with each layer of 25,11,7,5,3, neurons, respectively.
- You define the implementation on your choice:
  - Cross-validation, feature scaling, or not

Data pipeline



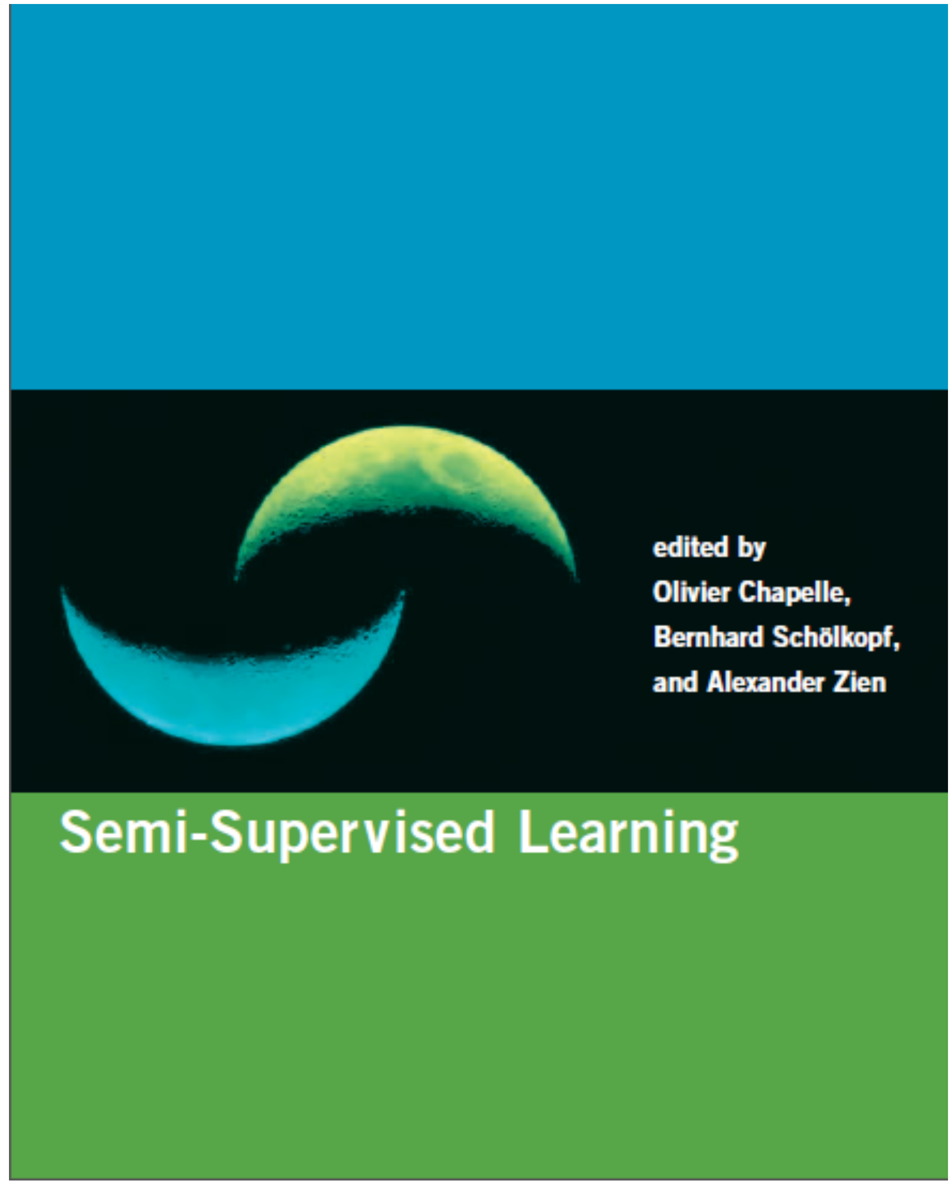
Urban Computing: concepts, methodologies, and applications.
Zheng, Y., et al. *ACM transactions on Intelligent Systems and Technology*.

Learning Map



Reference

Semi-supervised Learning

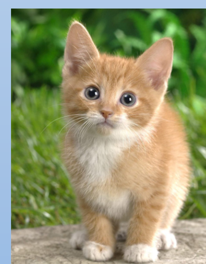


http://olivier.chapelle.cc/ssl-book/

Semi-supervised Learning

Introduction

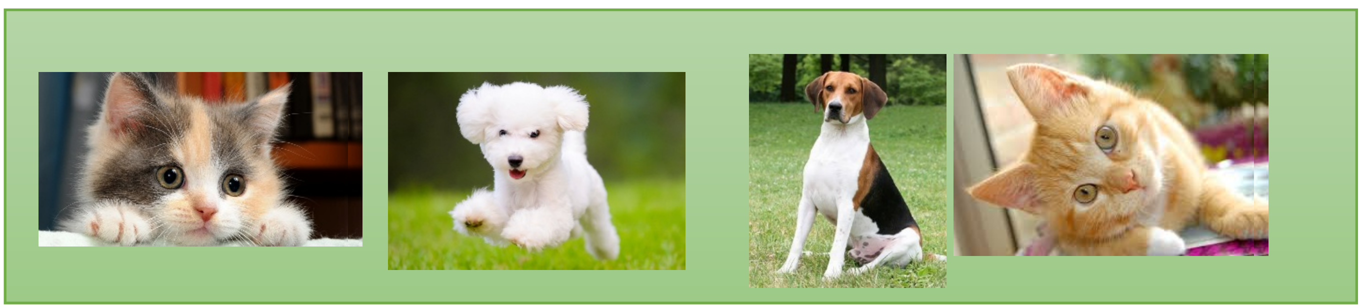
Labelled data



cat

dog

Unlabeled data

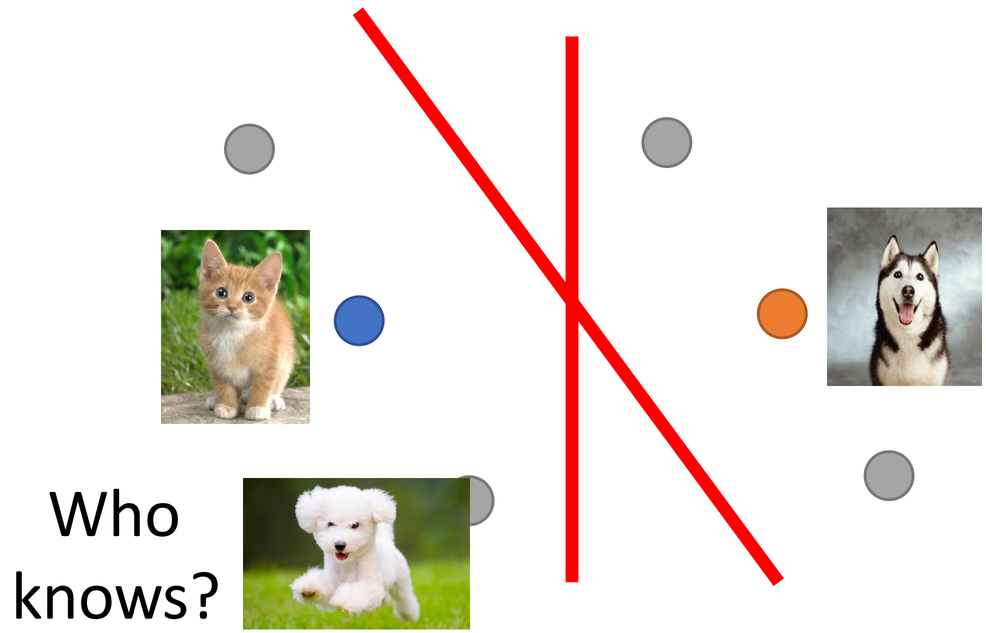


(Image of cats and dogs without labeling)

Introduction

- Supervised learning: $\{(x^r, \hat{y}^r)\}_{r=1}^{R}$
  - E.g. x^r: image, $\hat{y}^r$: class labels
- Semi-supervised learning: $\{(x^r, \hat{y}^r)\}_{r=1}^{R}$, $\{x^u\}_{u=R+1}^{R+U}$
  - A set of unlabeled data, usually U >> R
  - Transductive learning: unlabeled data is the testing data
  - Inductive learning: unlabeled data is **not** the testing data
- Why semi-supervised learning?
  - Collecting data is easy, but collecting "labelled" data is expensive
  - We do semi-supervised learning in our lives

Why semi-supervised learning helps?



The distribution of the unlabeled data tell us ***something***.

Usually with some assumptions

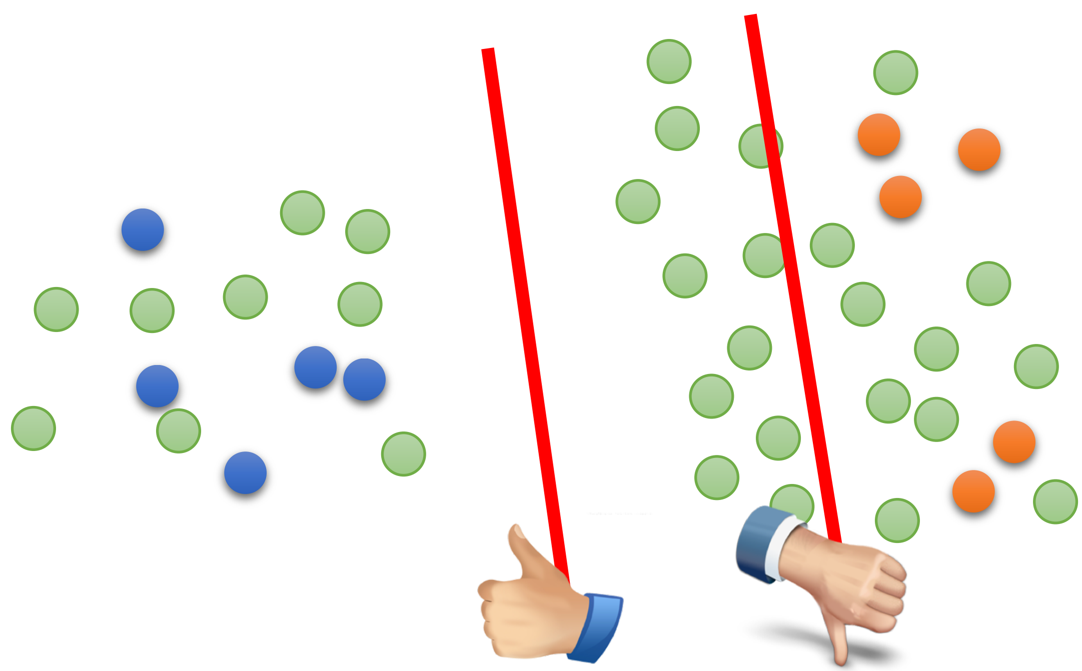
Outline

Low-density Separation Assumption

Smoothness Assumption

Semi-supervised Learning
Low-density Separation

Clear
Separation



Self-training for classification

- Given: labelled data set = $\{(x^r, \hat{y}^r)\}_{r=1}^{R}$, unlabeled data set = $\{x^u\}_{u=1}^{U}$
- Repeat:
  - Train model f^* from labelled data set

You can use any model here.

Regression?

  - Apply f^* to the unlabeled data set
    - Obtain $\{(x^u, y^u)\}_{u=1}^{U}$

Pseudo-label

  - Remove a set of data from unlabeled data set, and add them into the labeled data set

How to choose the data set remains open

You can also provide a weight to each data.

Self-training for Classification with hard labels

- Hard label v.s. Soft label

Considering using neural network

θ^* (network parameter) from labelled data

$x^u \rightarrow \boxed{\theta^*} \rightarrow \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix}$

Hard

Class 1 → New target for x^u is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

It looks like class 1, then it is class 1.

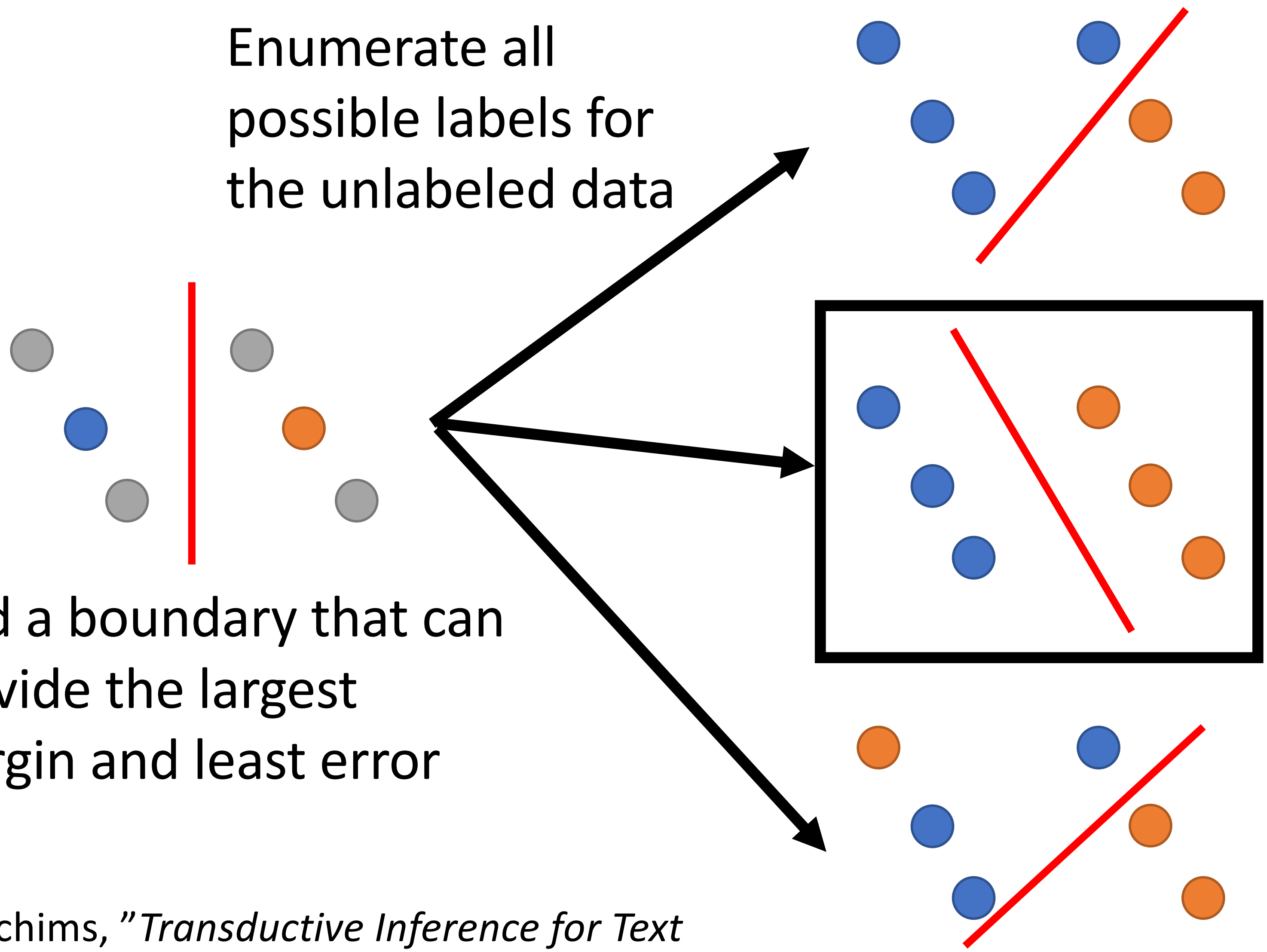
Soft

70% Class 1
30% Class 2 → New target for x^u is $\begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix}$

Doesn't work …

Outlook: Semi-supervised SVM

Enumerate all possible labels for the unlabeled data



Find a boundary that can provide the largest margin and least error

Thorsten Joachims, "*Transductive Inference for Text Classification using Support Vector Machines*", ICML, 1999

Semi-supervised Learning
Smoothness Assumption

“You are known by the company you keep”

Smoothness Assumption

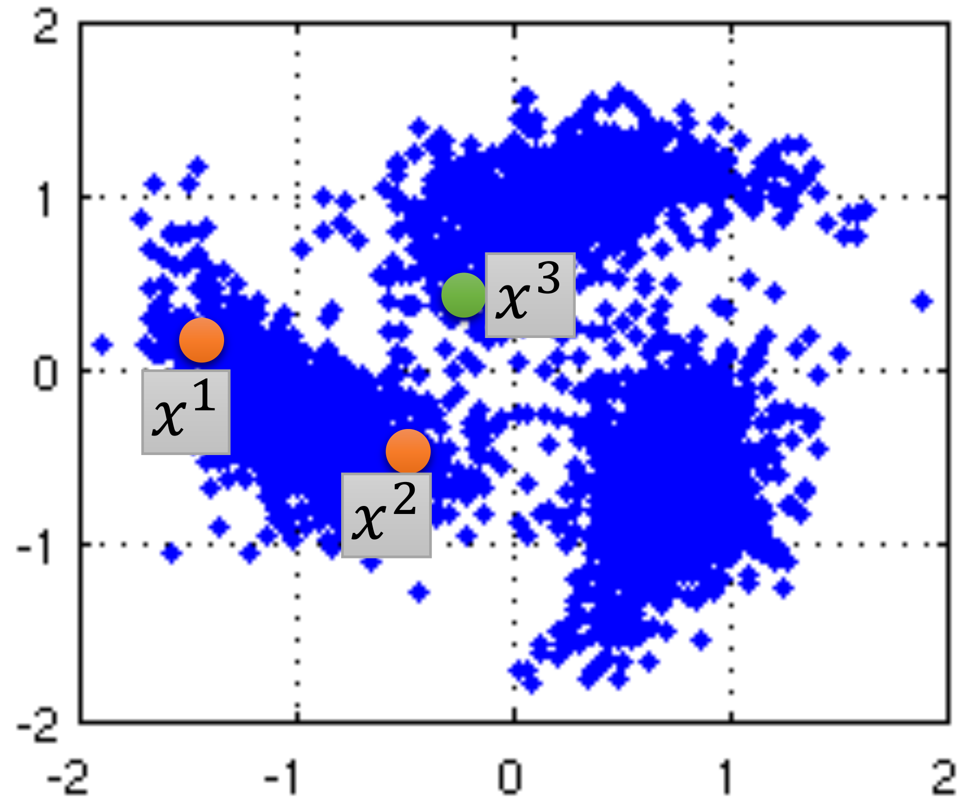
- Assumption: “similar” x has the same $\hat{y}$
- More precisely:
  - x is not uniform.
  - If x^1 and x^2 are close in a high density region, $\hat{y}^1$ and $\hat{y}^2$ are the same.

connected by a high density path

Smoothness Assumption

- Assumption: "similar" x has the same $\hat{y}$
- More precisely:
  - x is not uniform.
  - If x^1 and x^2 are close in a high density region, $\hat{y}^1$ and $\hat{y}^2$ are the same.

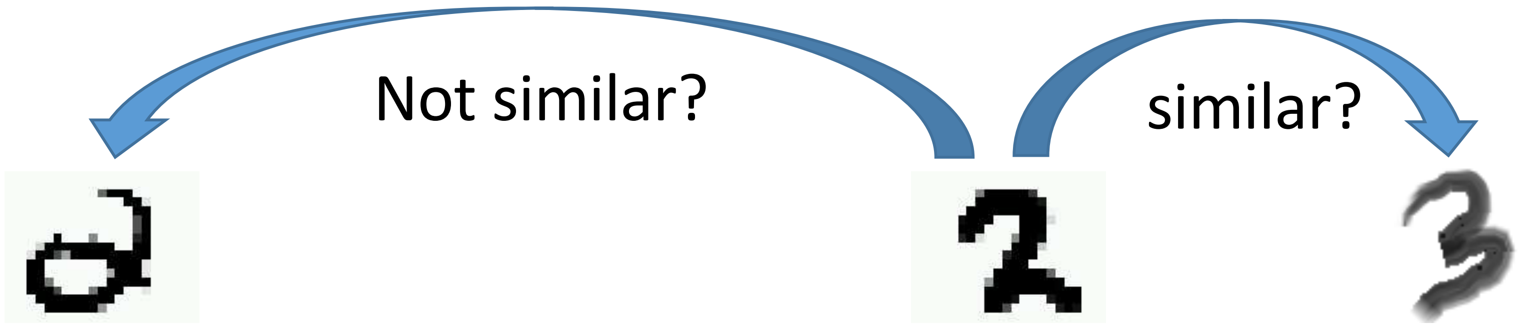
connected by a high density path



x^1 and x^2 have the same label

x^2 and x^3 have different labels

Smoothness Assumption



"indirectly" similar
with stepping stones

Smoothness Assumption

- Classify astronomy vs. travel articles

	d_1	d_3	d_4	d_2
asteroid	●	●		
bright	●	●		
comet		●		
year				
zodiac				
⋮				
airport				
bike				
camp			●	
yellowstone			●	●
zion				●

	d_1	d_3	d_4	d_2
asteroid	●			
bright	●			
comet				
year				
zodiac		●		
⋮				
airport			●	
bike			●	
camp				
yellowstone				●
zion				●

Smoothness Assumption

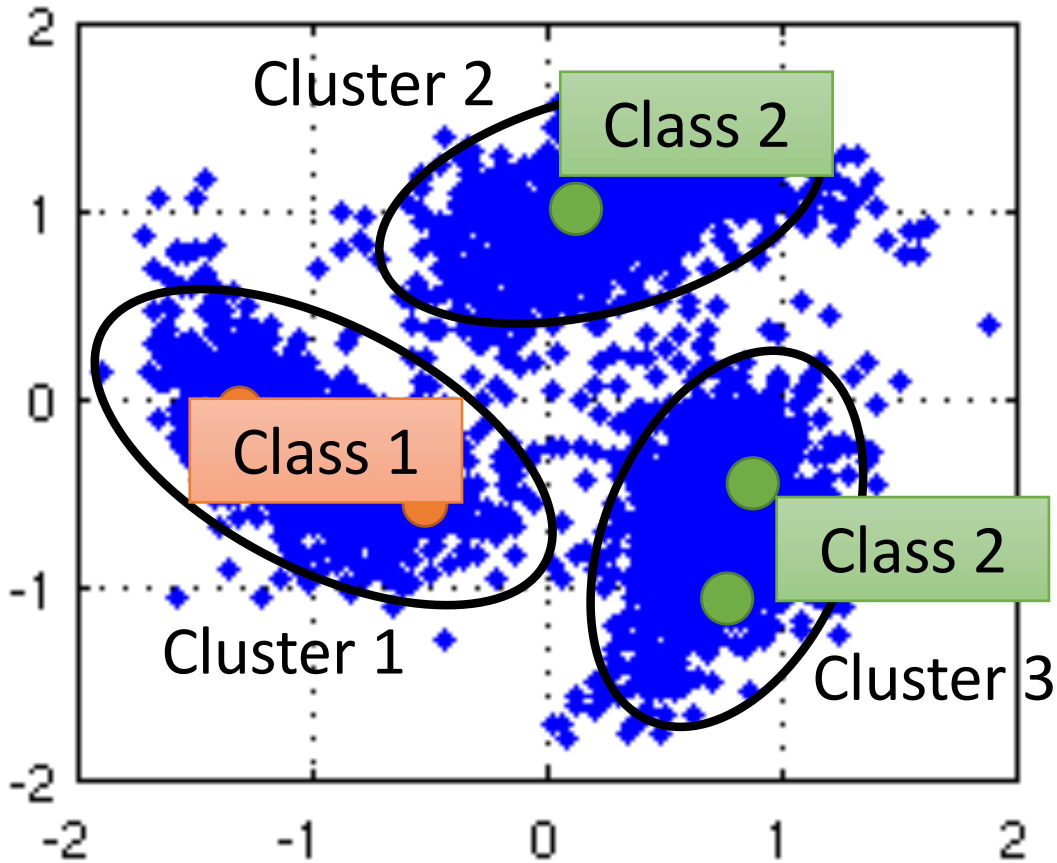
- Classify astronomy vs. travel articles

	d_1	d_5	d_6	d_7	d_3	d_4	d_8	d_9	d_2
asteroid	●								
bright	●	●							
comet		●	●						
year			●	●					
zodiac				●	●				
⋮									
airport						●			
bike						●	●		
camp							●	●	
yellowstone								●	●
zion									●

Cluster and then Label

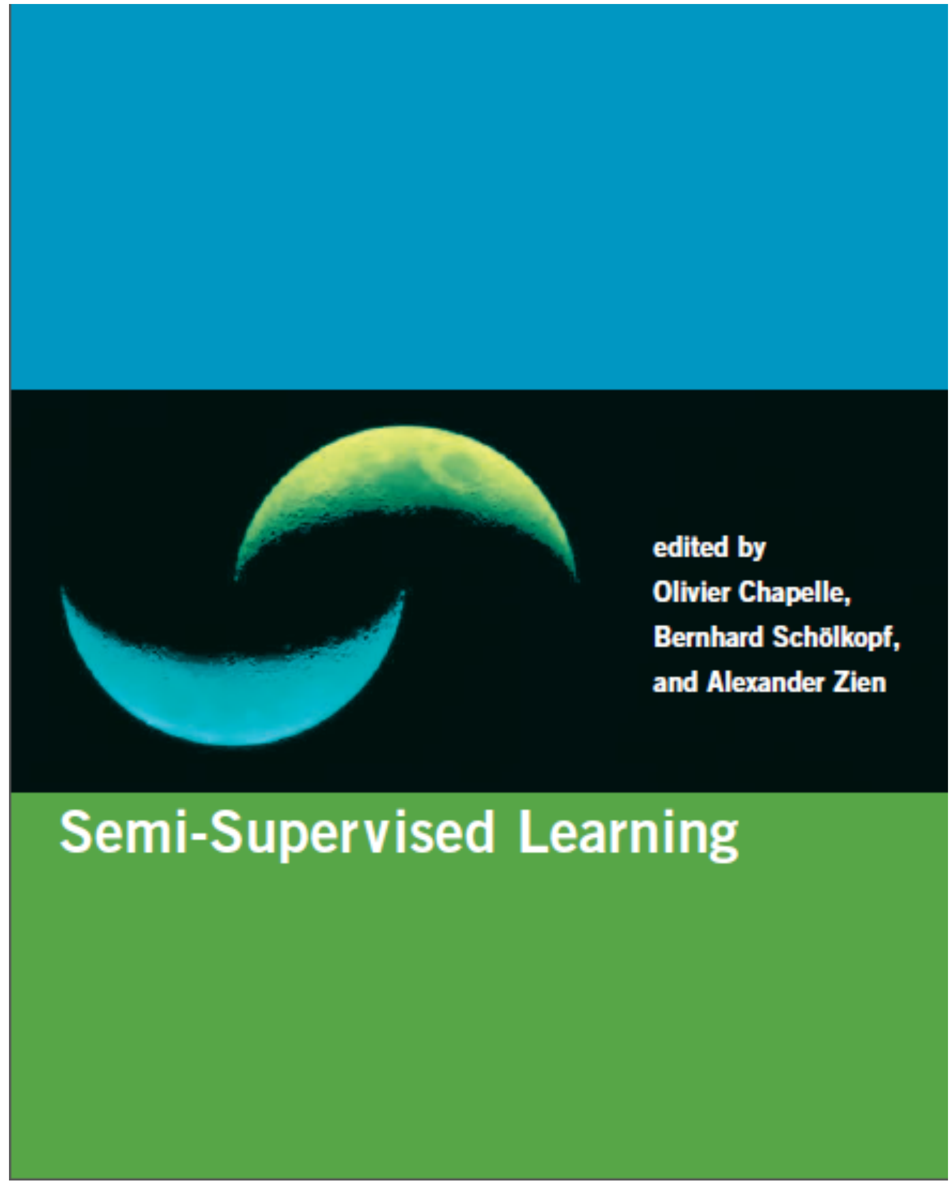
Class 1

Class 2



Using all the data to learn a classifier as usual

Reference



http://olivier.chapelle.cc/ssl-book/

Questions