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

Service Providing
Improve urban planning, Ease Traffic Congestion, Save Energy, Reduce Air Pollution, ...

Urban Data Analytics
Data Mining, Machine Learning, Visualization

Urban Data Management
Spatio-temporal index, streaming, trajectory, and graph data management,...

Urban Sensing & Data Acquisition
Participatory Sensing, Crowd Sensing, Mobile Sensing

Urban Computing: concepts, methodologies, and applications.
Reference

Semi-supervised Learning

http://olivier.chapelle.cc/ssl-book/
Semi-supervised Learning
Introduction

Labelled data

Unlabelled 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 \( U \gg R \)
  • 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.
  • 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
\( \theta^* \) (network parameter) from labelled data

New target for \( x^u \) is \([1, 0] \)

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

New target for \( x^u \) is \([0.7, 0.3] \)

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

Not similar?  

similar?

\[ \alpha \]  

\[ \beta \]  

\[ \gamma \]
Smoothness Assumption

- Classify astronomy vs. travel articles

<table>
<thead>
<tr>
<th></th>
<th>$d_1$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>asteroid</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bright</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>comet</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>year</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>zodiac</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>airport</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>bike</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>camp</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>yellowstone</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>zion</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$d_1$</th>
<th>$d_3$</th>
<th>$d_4$</th>
<th>$d_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>asteroid</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>bright</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>comet</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>year</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>zodiac</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>airport</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>bike</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>camp</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yellowstone</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>zion</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
Smoothness Assumption

- Classify astronomy vs. travel articles

<table>
<thead>
<tr>
<th></th>
<th>d₁</th>
<th>d₅</th>
<th>d₆</th>
<th>d₇</th>
<th>d₃</th>
<th>d₄</th>
<th>d₈</th>
<th>d₉</th>
<th>d₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>asteroid</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bright</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>comet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zodiac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>airport</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bike</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>camp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yellowstone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cluster and then Label

Using all the data to learn a classifier as usual
Reference

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