Machine Learning in BioMedical Informatics

: Support Vector Machines & Semi-Supervised Learning

Hyunjung (Helen) Shin

Dept. of Industrial & Information Systems Engineering
Ajou University, Korea

shin@ajou.ac.kr
http://www.alphaminers.net
http://www.kyb.tuebingen.mpg.de/~shin

Abstract: Machine Learning in BioMedical Informatics

The emergence of the fields of **bio-** and **medical- informatics** has alleviated the burden of solving the related biological or medical domain problems, saving the **time and cost** required for **wet-lab experiments** or **patient treatments**, and it also has provided **predictions** that guide the next experiments or treatments.

In **biomedical informatics**, **machine learning** has played a central role in dealing with the flood of data. The **goal of this talk** is to raise **awareness** and **comprehension** of **machine learning** so that the domain specialists can properly match the task at hand to the corresponding analytical approach.

Abstract: Machine Learning in BioMedical Informatics

We start by **categorizing** the types of the tasks with given data and introduce the **general machine learning schemes** that fit best to each of the categories. We then explore **representative models** from traditional statistical models to the most recent ones such as **support vector machines (SVM)** and **semi-supervised learning (SSL)**

To exemplify how biological or medical questions can benefit from **machine learning**, we present several **up-to-date research projects**: **protein functional class** prediction based on genome and proteome data, **breast cancer** survivability prediction based on clinical data of patients, and clinical outcome prediction of **brain- and ovarian cancer** based on genome to phenome data.

Outline

Preliminaries	Data & Tasks
Machine Learning	Machine Learning Fields of Machine Learning Machine Learning Models
Support Vector Machines (SVM)	Kernel Method Basic Idea of SVM
Semi-Supervised Learning (SSL)	Fundamentals Problem Setting Family of Algorithm Graph-based SSL
Case Examples (SVM vs. SSL)	[Intra-Relation] Single Data Source [Integration] Multiple Data Sources
Closing Remarks	Future Work

Data & Tasks

- **Prediction**
  - Classification
  - Regression
- **Description**
  - Clustering
  - Feature Description
  - Outlier Detection
- **Dimensionality Reduction**
  - Feature Selection
  - Feature Extraction
- **Data Reduction (Sample Selection)**
- **Data Integration**

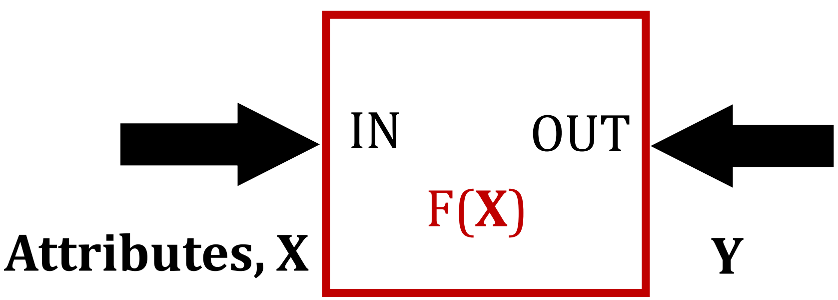
Machine Learning

Machine Learning : Algorithms or Techniques that allow computers to "learn" using "data"

1

"TRAINING"

Develop Model
With **Known Data**

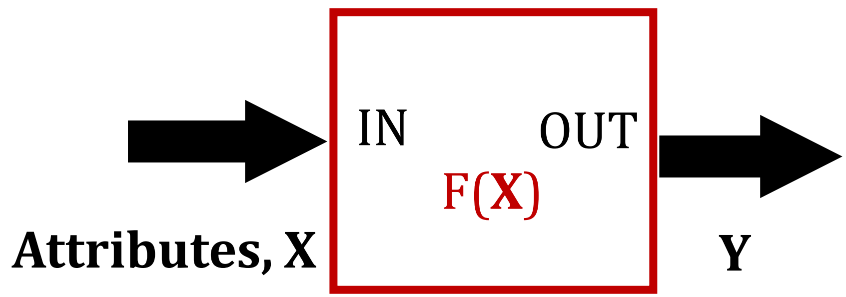


Determine F(X)

2

"TEST/EVALUATION"

Use Model
For **New Data**



Y = F(X)

The Fields of Machine Learning

The **field of machine learning** has traditionally been divided into **three sub-fields**:

Supervised Learning

Unsupervised Learning

Reinforcement Learning

The Fields of Machine Learning: Supervised Learning

The **field of machine learning** has traditionally been divided into **three sub-fields**:

Supervised Learning

The learning system observes a **labeled training set** consisting of **(attribute, label) pairs**, denoted by $\{(x_1, y_1), \ldots, (x_n, y_n)\}$.

The ***goal*** is to predict the **label y** for any new input with feature x.

A supervised learning task is called ***regression*** **when y ∈ R**, and ***classification*** when **y** takes a set of **discrete values**.

Unsupervised Learning

Reinforcement Learning

The Fields of Machine Learning: Unsupervised Learning

The **field of machine learning** has traditionally been divided into **three sub-fields**:

Supervised Learning

Unsupervised Learning

The learning system observes an **unlabeled set of items**, represented by their attributes $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$.

The ***goal*** is to **organize the items:** Typical unsupervised learning tasks includes ***clustering*** that groups items into clusters**;** ***outlier detection*** which determines if a new item **x** is significantly different from items seen so far; ***dimensionality reduction*** which maps **x** into a low dimensional space, while preserving certain properties of the dataset.

Reinforcement Learning

The Fields of Machine Learning: Reinforcement Learning

The **field of machine learning** has traditionally been divided into **three sub-fields**:

Supervised Learning

Unsupervised Learning

Reinforcement Learning

The learning system repeatedly **observes** the ***environment "x,"*** **performs** an ***action "a,"*** and **receives** a ***reward "r."***

The ***goal*** is to choose the actions that ***maximize the future rewards.***

SUPPORT VECTOR MACHINES (SVM)

SVM: Kernel Methods

Why KM?

- Kernel methods can operate on very **general types of data** and can detect very **general types of relations**

- Various tasks-- $\left\{\begin{array}{l}\textbf{PCA,}\\ \textbf{CCA,}\\ \textbf{FA,}\\ \textbf{DA,}\\ \textbf{Clustering}\end{array}\right\}$ can be performed on diverse data $\left\{\begin{array}{l}\textbf{vectors,}\\ \textbf{sequences,}\\ \textbf{text,}\\ \textbf{images,}\\ \textbf{graphs}\end{array}\right\}$

- **Integration of different types of data** is easy and natural

SVM: Wrap-up

$$\min_{0\le \alpha_i \le C} W(\alpha_i, b) = \frac{1}{2}\sum_{i,j=1}^{M} \alpha_i \alpha_j y_i y_j K(\vec{x}_i, \vec{x}_j) - \sum_{i=1}^{M} \alpha_i + b\sum_{i=1}^{M} y_i \alpha_i$$

KKT

$$\frac{\partial W(\alpha_i, b)}{\partial \alpha_i} = \sum_{j=1}^{M} y_i y_j K(\vec{x}_i, \vec{x}_j)\alpha_j + y_i b - 1 = y_i \bar{f}(x_j) - 1$$

$$\frac{\partial W(\alpha_i, b)}{\partial b} = \sum_{j=1}^{M} y_j \alpha_j = 0$$

where $\bar{f}(\vec{x}) = \sum_{\mathrm{i}=1}^{\mathrm{M}} y_{\mathrm{i}} \alpha_{\mathrm{i}} K(\vec{x}_{\mathrm{i}}, \vec{x}) + b$

Semi-Supervised Learning: Learning Schemes

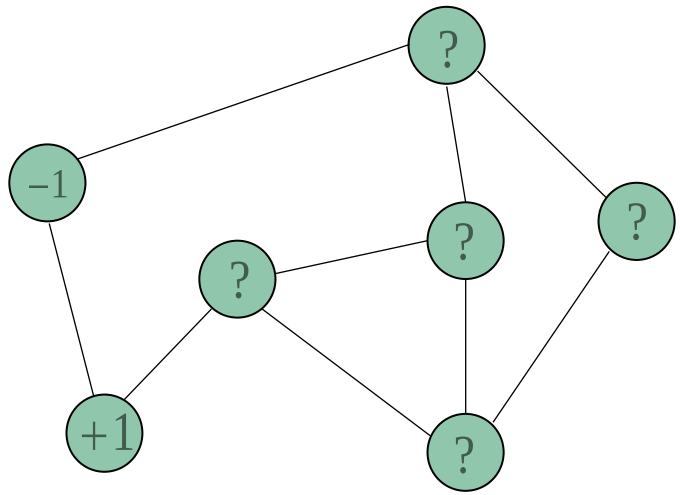
Supervised Learning

"Semi"-Supervised Learning

Unsupervised Learning

Reinforcement Learning

Graph-based SSL: WRAP-UP



Objective Function $$\min_{\boldsymbol{f}} \quad \mu\, \boldsymbol{f}^T L\, \boldsymbol{f} + (\boldsymbol{f}-\boldsymbol{y})^T(\boldsymbol{f}-\boldsymbol{y})$$

Solution $$\boldsymbol{f} = \{\, \boldsymbol{I} + \mu L \,\}^{-1} \boldsymbol{y}$$

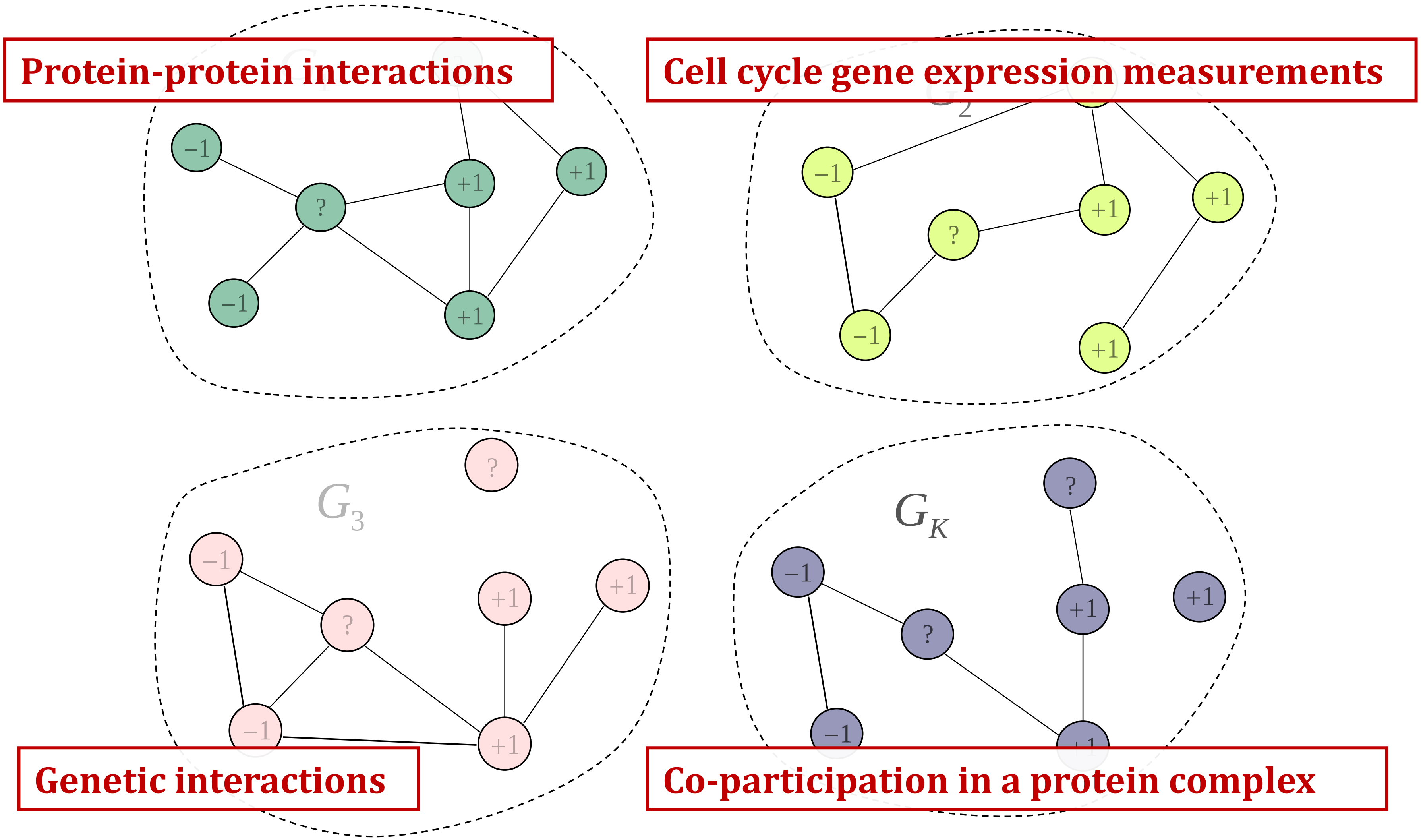
where $L = D - W,\ D = diag(d_i),\ \ d_i = \sum_j w_{ij}$

Abstract: Data Integration

Data Integration is concerned with the **integration of different or heterogeneous data sources** in order to **enhance the total information** about the problem at hand**.**

Each of data sources contains **partly independent** and **partly complementary** pieces of information about the problem**...**

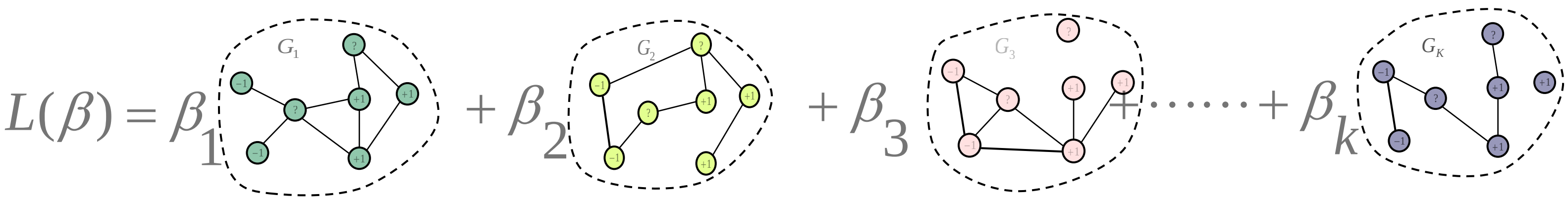
If Multiple Graphs are Given?



Graph Integration using SSL

Linear Combination of Laplacians

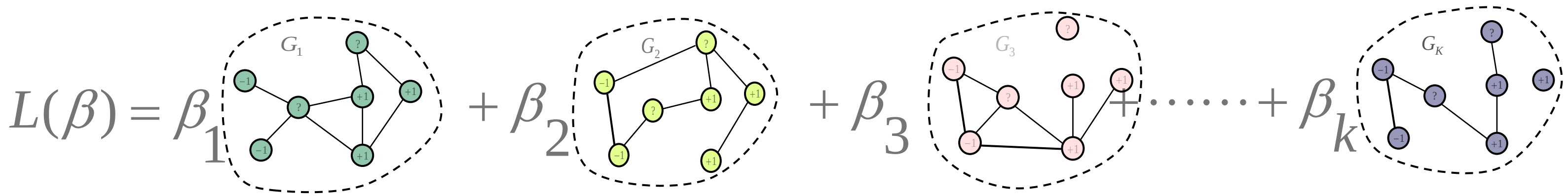
$$L(\beta) = \sum_{k=1}^{K} \beta_k L_k$$



How to Find Combining Weights ?

Graph Integration using SSL

Mutiple Graph (Data) Integration



Shin, H., Lisewski, A.M. and Lichtarge, O. (2007)
Graph sharpening plus graph integration: a synergy that improves protein functional classification, ***Bioinformatics***, 23, 3217-3224.

Shin, H. and Tsuda, K. (2006)
Prediction of Protein Function from Networks, *in Book:* ***Semi-Supervised Learning, MIT press***, Chapter 20, 339-352.

Tsuda, K., Shin, H. and Scholkopf, B. (2005)
Fast protein classification with multiple networks, ***Bioinformatics*****,** 21 Suppl 2, ii59-65.

References

M. Belkin and P. Niyogi. Semi-supervised learning on Riemannian manifolds. Machine Learning, 56(1):209–239, 2004

M. C. Cirelo F. G. Cozman, I. Cohen. Semi-supervised learning of mixture models. In International Conference on Machine Learning, 2003

T. Joachims. Transductive inference for text classification using support vector machines. In International Conference on Machine Learning, pp. 200–209, 1999

K. Nigam, A. K. McCallum, S. Thrun, and T. M. Mitchel. Text classification from labeled and unlabeled documents using EM. Machine Learning, 39(2/3):103–134, 2000

S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500):2323–2326, 2000

H. Shin and K. Tsuda. Prediction Protein Function from Networks. Chap 20 in book Semi-Supervised Learning edited by Olivier Chapelle, Bernhard Schölkopf and Alexander Zien, pp. 339~352, MIT Press, 2006, Also appears in *Bioinformatics, vol. 21, pp.59–65, Oxford University Press, 2005*

H. Shin, A.M. Lisewski, and O. Lichtarge. Graph Sharpening plus Graph Integration: A Synergy that Improves Protein Functional Classification. Bioinformatics, vol. 23, no. 23, pp. 3217–3224, Oxford University Press, 2007

H.Shin, N.J. Hill, A.M. Lisewski,J. Park. Graph Sharpening. Expert Systems with Applications, vol. 37, no. 12, pp. 7870–7879, 2010

H.Shin, T.Hou, K.Park, C.Park, and S.Choi. Oil Price Prediction Based on Semi-Supervised Learning, Decision Support Systems (in print), 2012

H. Lee, H.Shin, S. Cho, and D. MacLachlan, Semi-Supervised Response Modeling” Journal of Interactive Marketing, vol.24, no. 1, pp.42–54, 2010

D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schoelkopf. Learning with local and global consistency. In Advances in Neural Information Processing Systems, volume 16, 2003

X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In International Conference on Machine Learning, 2003.

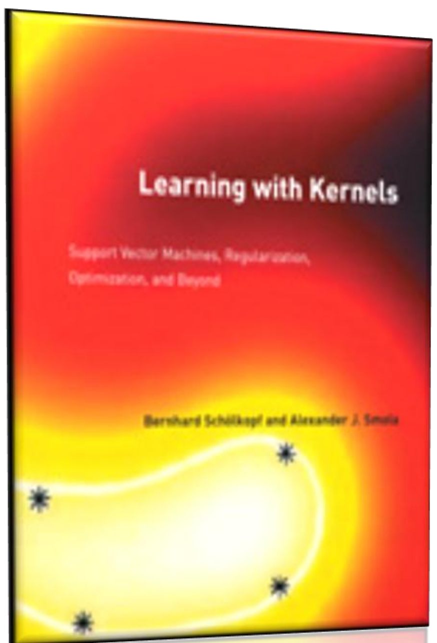
X. Zhu, Semi-Supervised Learning with Graphs (Doctoral Thesis). CMU-LTI-05-192, CMU, 2005

References

Learning with Kernels:

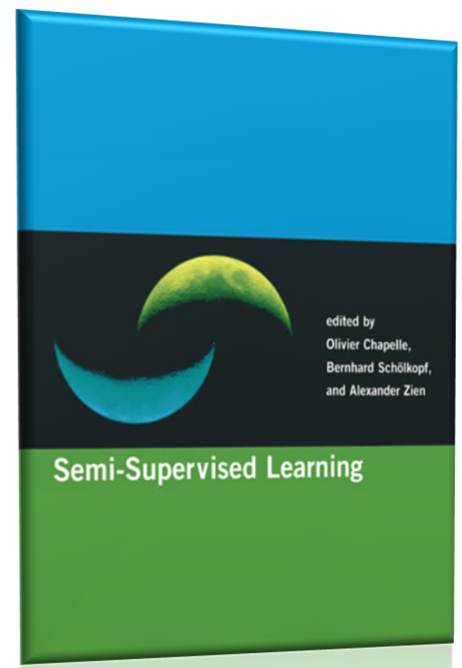
Support Vector Machines, Regularization, Optimization, and Beyond

Bernhard Schölkopf and Alexander J. Smola, MIT Press, 2004



Semi-Supervised Learning

Olivier Chapelle, Bernhard Schölkopf and Alexander Zien, MIT Press, 2006



Acknowledgements

The authors would like to gratefully acknowledge support from Post Brain Korea 21 and the research grant from National Research Foundation of the Korean Government (2009-0065043/2010-0007804/2011-0018257)