UNIVERSITY OF GEORGIA

CSCI/ARTI 4530/6530: Introduction to Robotics

Guest Lecture: An Overview of Machine Learning

Sheng Li

Assistant Professor

Department of Computer Science

University of Georgia

Oct. 15, 2018

What is Machine Learning?

- One formal definition of *machine learning algorithms*:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, **improves** with experience E."

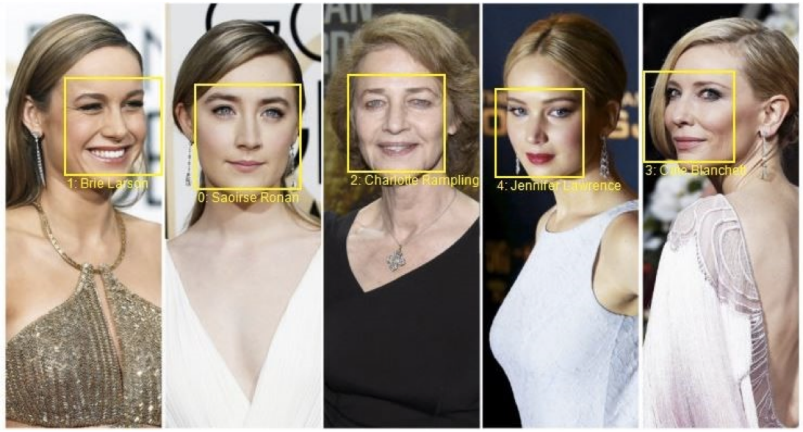
----Tom M. Mitchell

- *Experience E*: data, labels, etc.
- *Tasks T*: image classification, activity prediction, etc.
- *Performance measure P*: accuracy, speed, etc.

Machine Learning Applications

- **Classification**—*predict discrete class labels*

Face Recognition



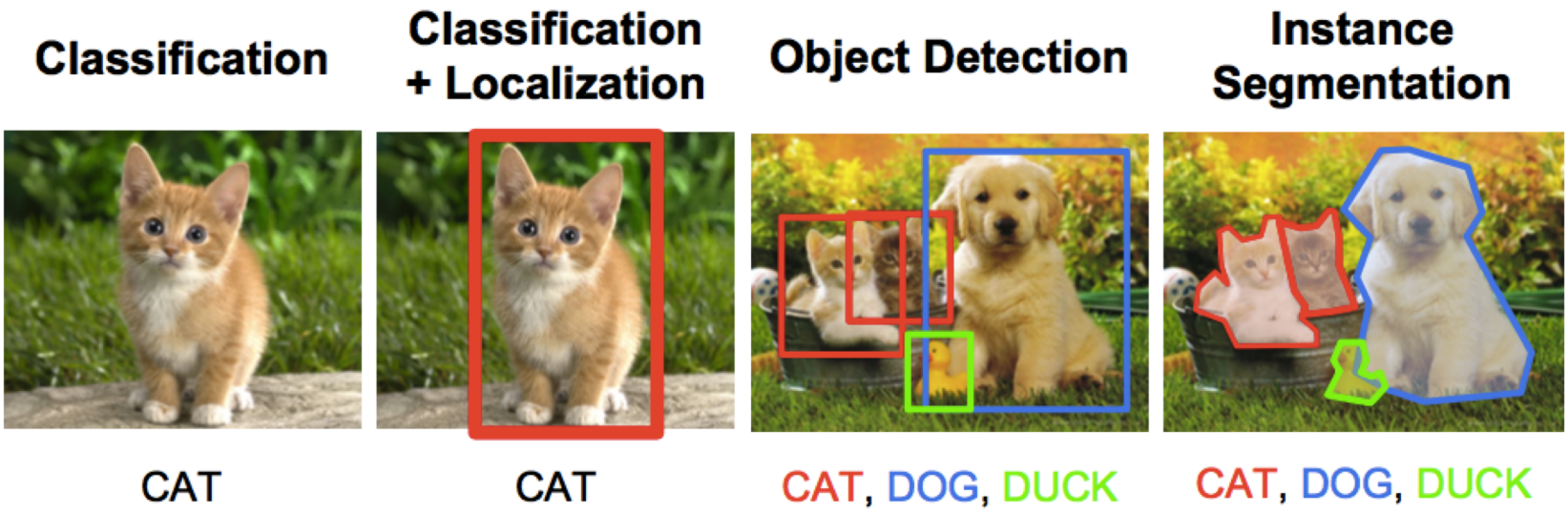
MS-Celeb-1M Challenge (2016)

Sentiment Classification

Positive and Negative Online Reviews

Machine Learning Applications

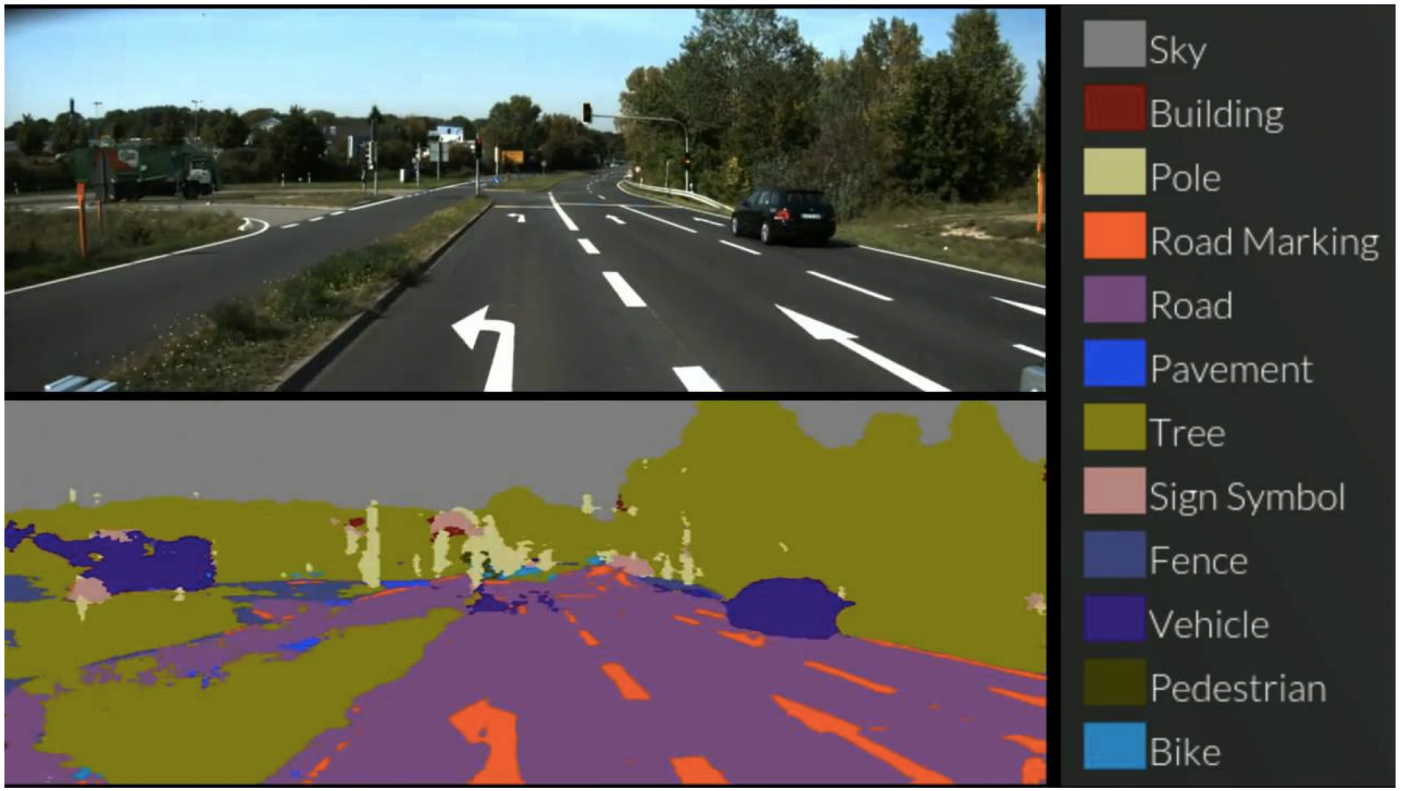
- **Classification**—*predict discrete class labels*



Computer Vision Example: Object Classification, Detection, and Segmentation

Machine Learning Applications

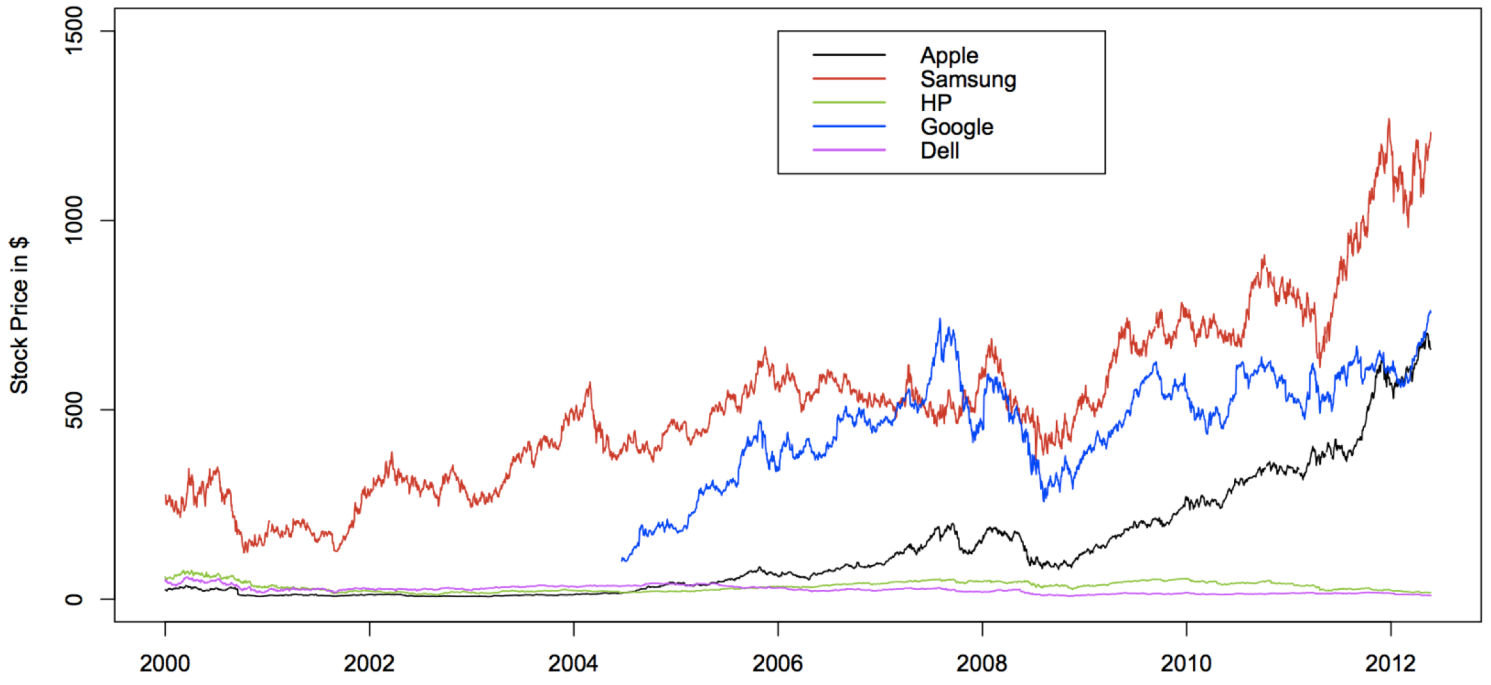
- **Classification**—*predict discrete class labels*



Computer Vision Example: Road Scene Understanding for Self-driving Car

Machine Learning Applications

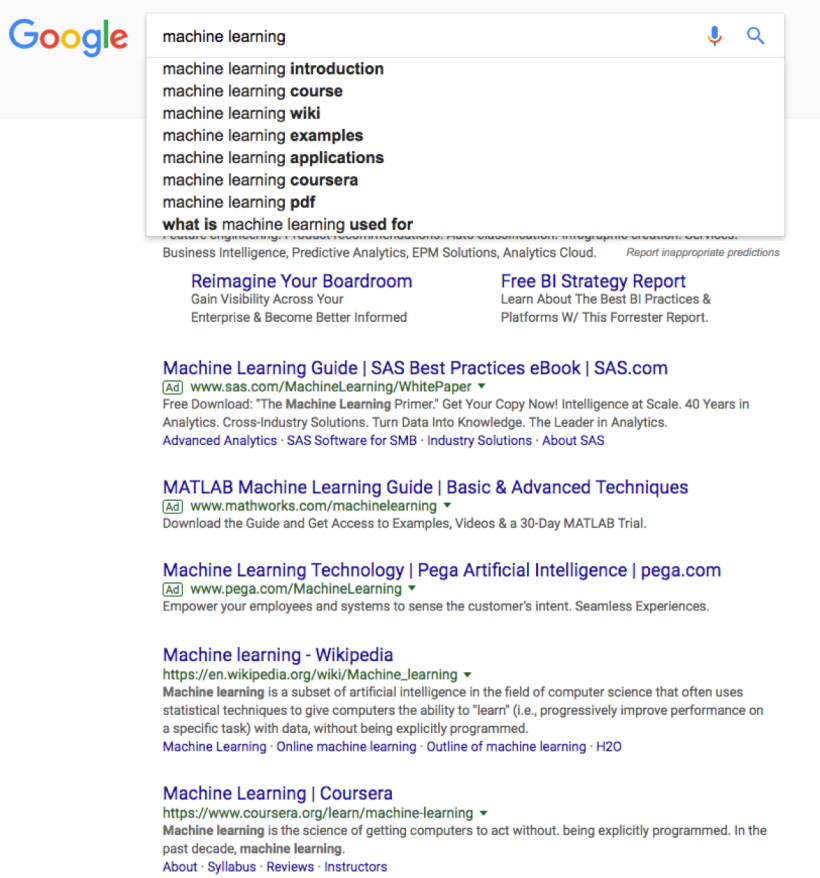
- **Regression**—*predicting numeric values*



Stock Market Analysis

Machine Learning Applications

- **Ranking**—*producing a permutation of items in the given list*



Web Search

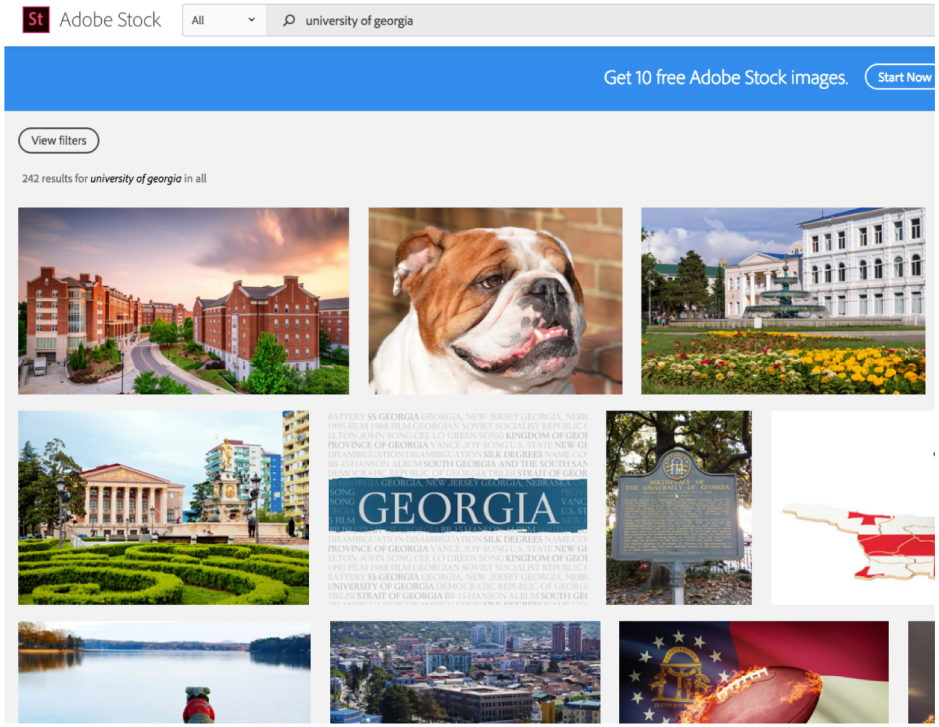
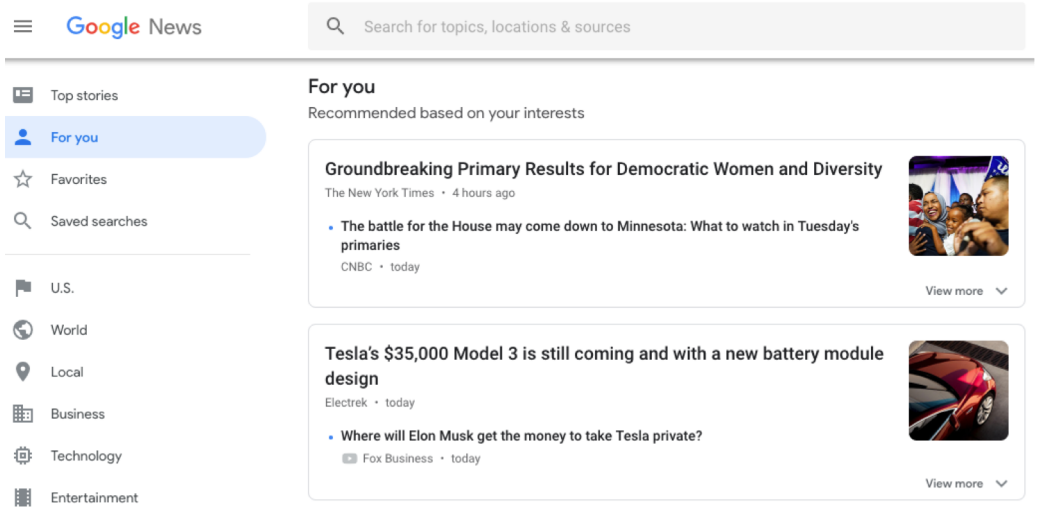


Image Search

Machine Learning Applications

- **Collaborative Filtering**—*completing matrix with missing values*

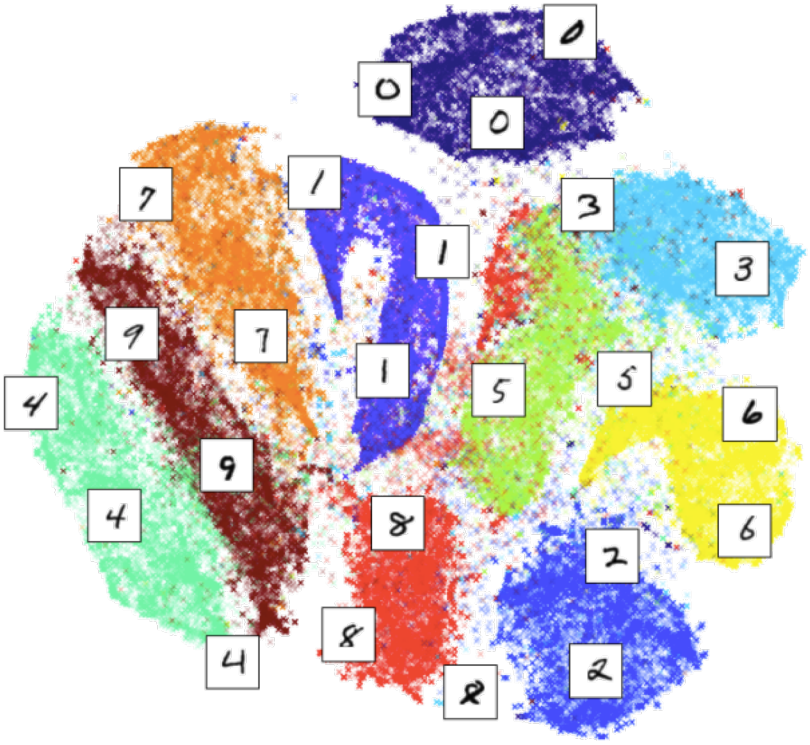
Movie Recommendation



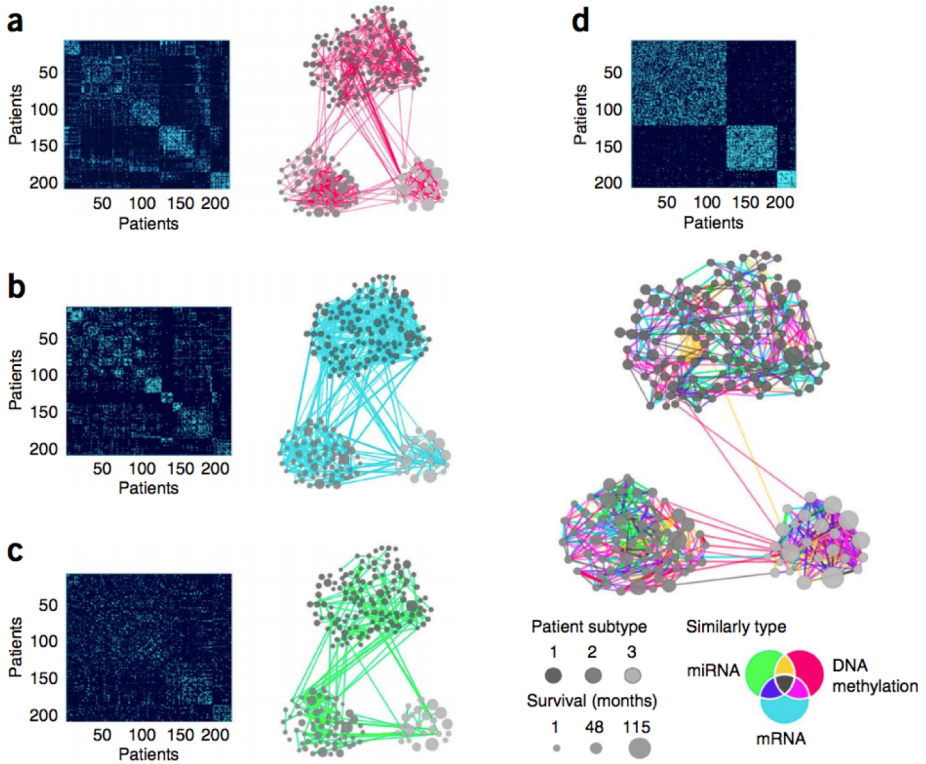
Google News Personalization

Machine Learning Applications

- **Clustering**—*discovering the underlying structure in data*



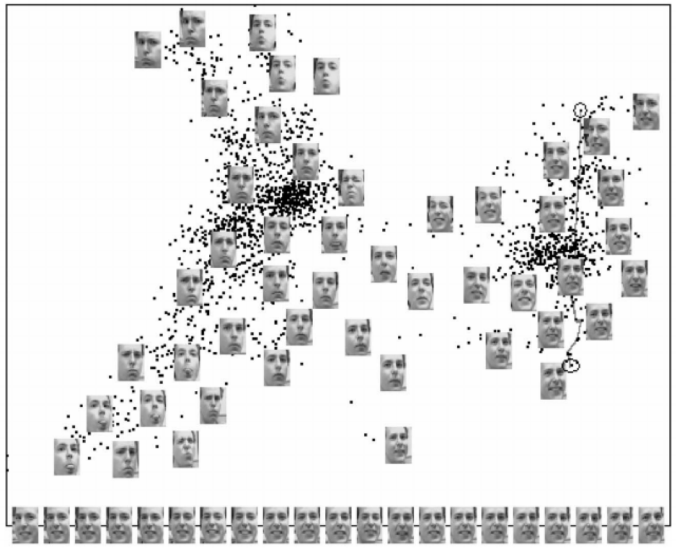
Clustering MNIST Digits



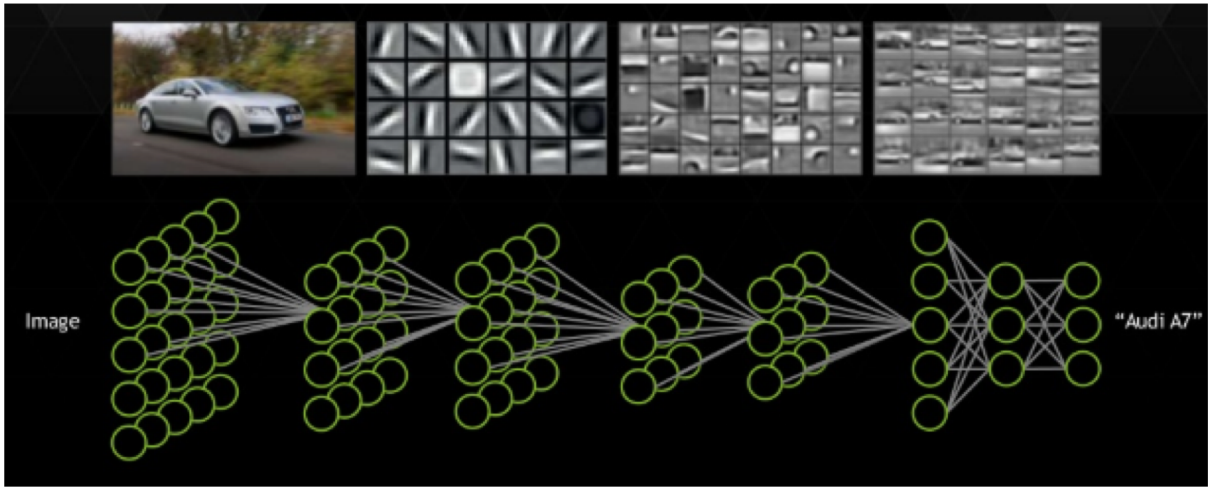
Patients Clustering (*Nature Methods*, 2014)

Machine Learning Applications

- **Representation Learning**—*extracting low-dimensional features from high-dimensional data*



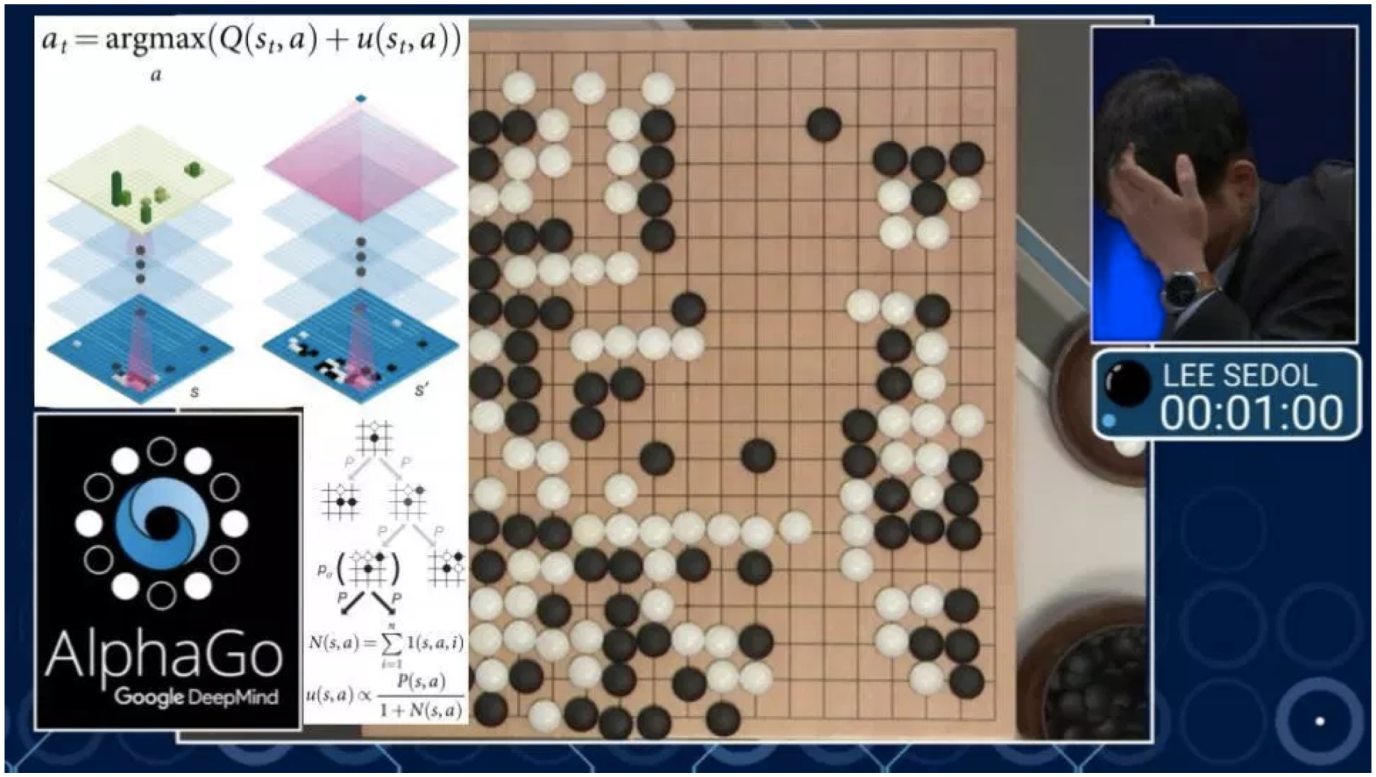
Manifold Embedding of Faces



Deep Representation Learning

Machine Learning Applications

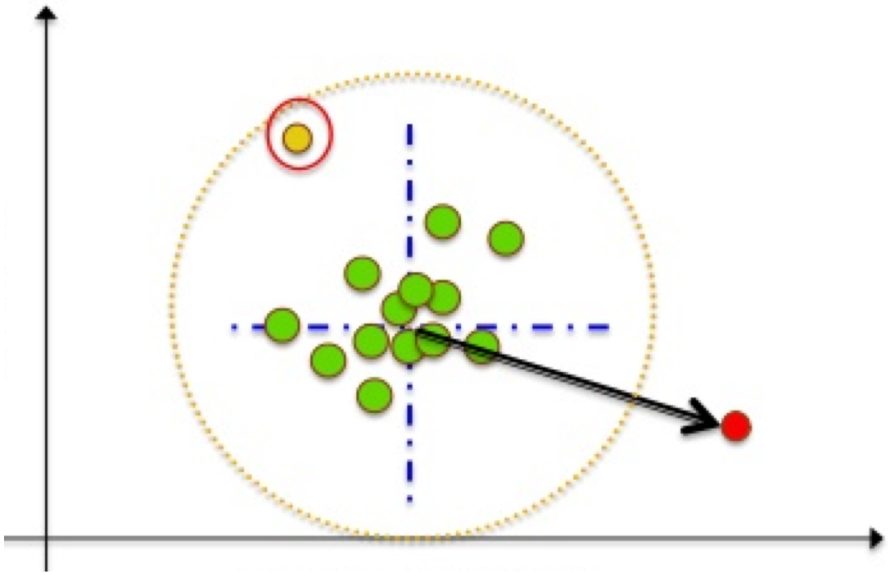
- **Reinforcement Learning**—*optimizing actions in an environment to maximize the reward*

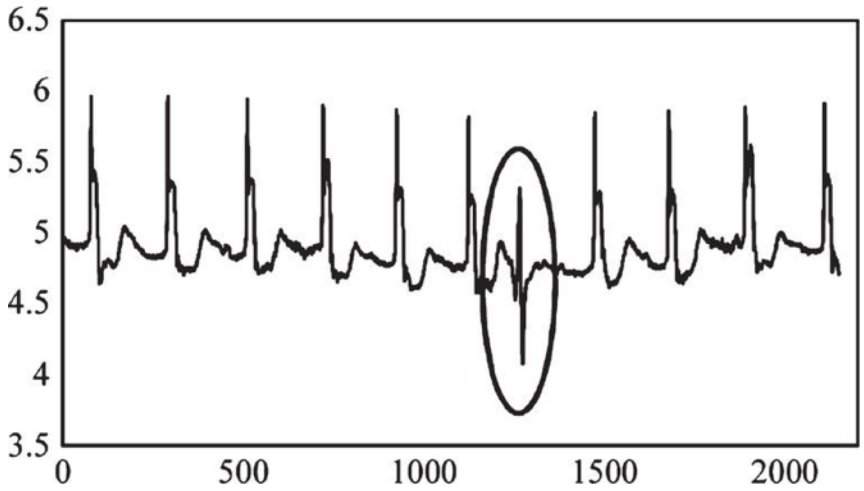


AlphaGo by Google DeepMind

Machine Learning Applications

- **Anomaly Detection**—*identifying outliers in a given dataset*





Machine Learning: Classical Topics

- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
  - Clustering
- Semi-supervised Learning
- Reinforcement Learning

Supervised Learning

- Input: Training set $\{(x_i, y_i) | i = 1, \ldots, N\}$
- Output: A good approximation to $f: X \rightarrow Y$

Real-world examples of X and Y

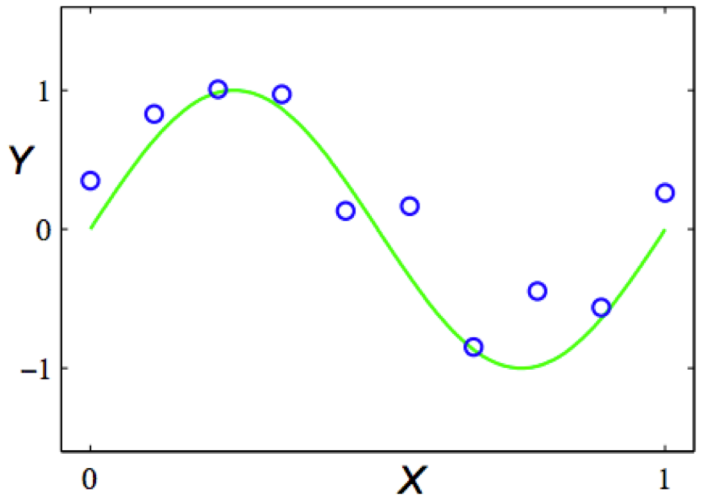
- House Price Prediction (*regression*)
  - X: historic prices; Y: new prices (continuous outputs)
- Spam detection (*binary classification*)
  - X: emails; Y: {Spam, Not Spam}
- Face Recognition (*multi-class classification*)
  - X: face images; Y: identity (a pre-defined set)

Supervised Learning

- Input: Training set $\{(x_i, y_i) | i = 1, \ldots, N\}$
- Output: A good approximation to $f: X \rightarrow Y$

Regression

- Given 10 (x_i, y_i) data points with noise, the goal is to find function f to fit the data

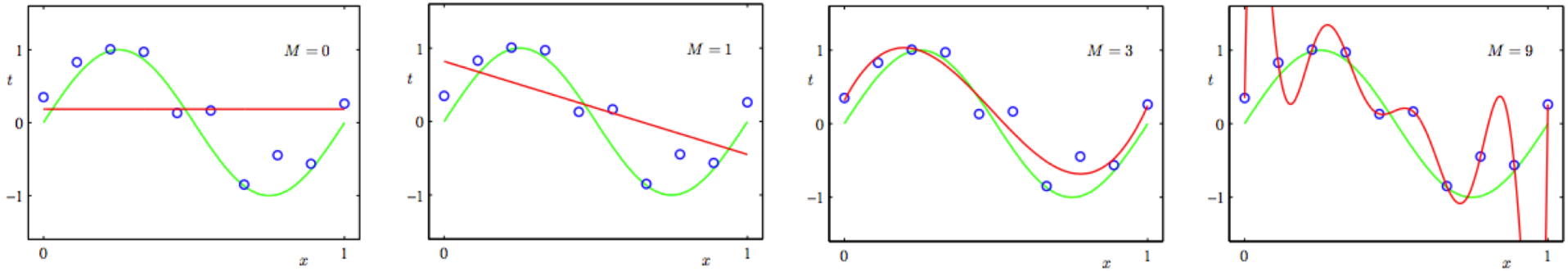


[Bishop]

Supervised Learning

Regression

- Given 10 (x_i, y_i) data points with noise, the goal is to find function f to fit the data



- To evaluate which one is the best, we need to measure error, using a loss function $L(Y, \hat{Y})$
- Squared loss function on the training data:

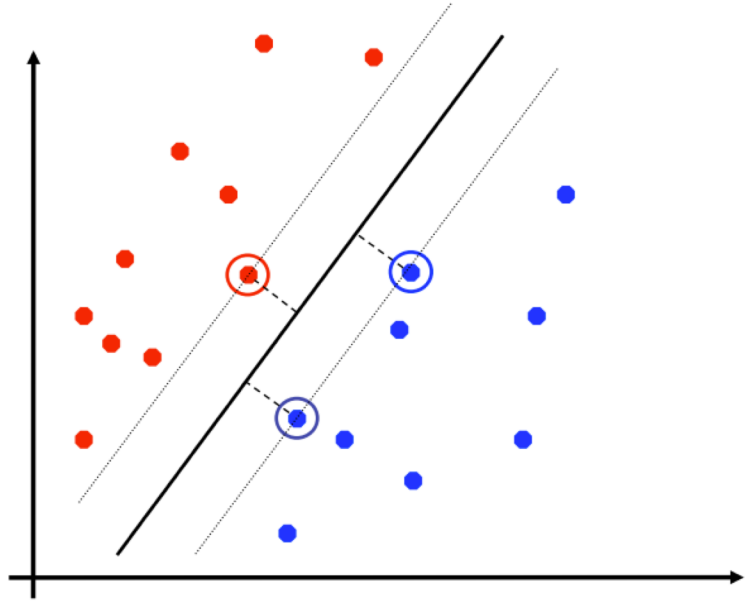
$$L(Y, \hat{Y}) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{y_i}) = \frac{1}{N}\sum_{i=1}^{N} (y_i - f(x_i))^2$$

Supervised Learning

- Input: Training set $\{(x_i, y_i) | i = 1, \ldots, N\}$, y_i are discrete labels
- Output: A good approximation to $f: X \rightarrow Y$

Classification

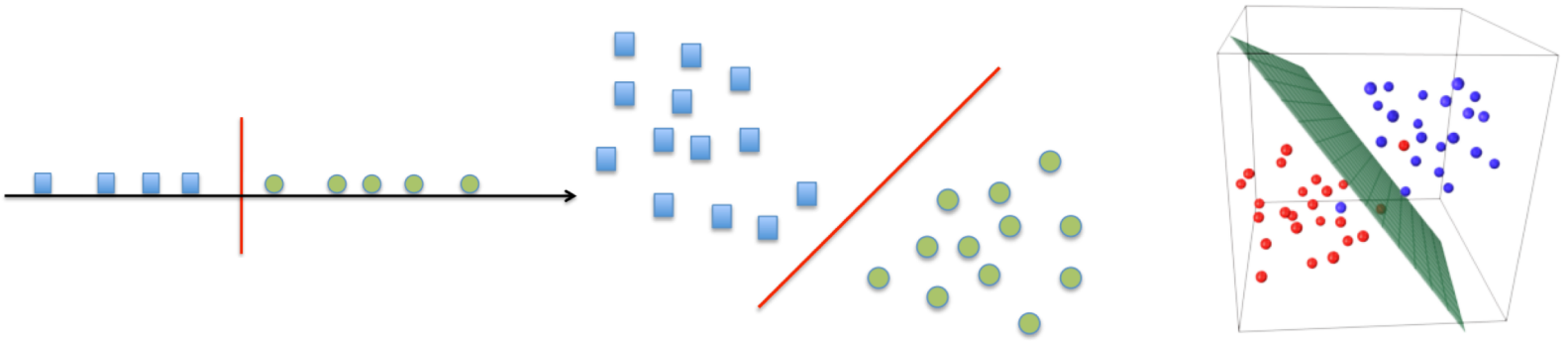
- Logistic Regression
- Support Vector Machines (SVM)
- Neural Networks



Supervised Learning

Classification

- Linear Decision Boundary in 1D, 2D and 3D Spaces



1D Space 2D Space 3D Space

[Fidler]

- Binary classifier: $f(\boldsymbol{x}) = \text{sign}(w_0 + \boldsymbol{w}^T \boldsymbol{x})$
- Loss functions
  - Zero/one loss: $L(f(\boldsymbol{x}), y) = \begin{cases} 0 & \text{if } f(\boldsymbol{x}) = y \\ 1 & \text{if } f(\boldsymbol{x}) \neq y \end{cases}$
  - Squared loss: $L(f(\boldsymbol{x}), y) = (y - f(\boldsymbol{x}))^2$

Unsupervised Learning

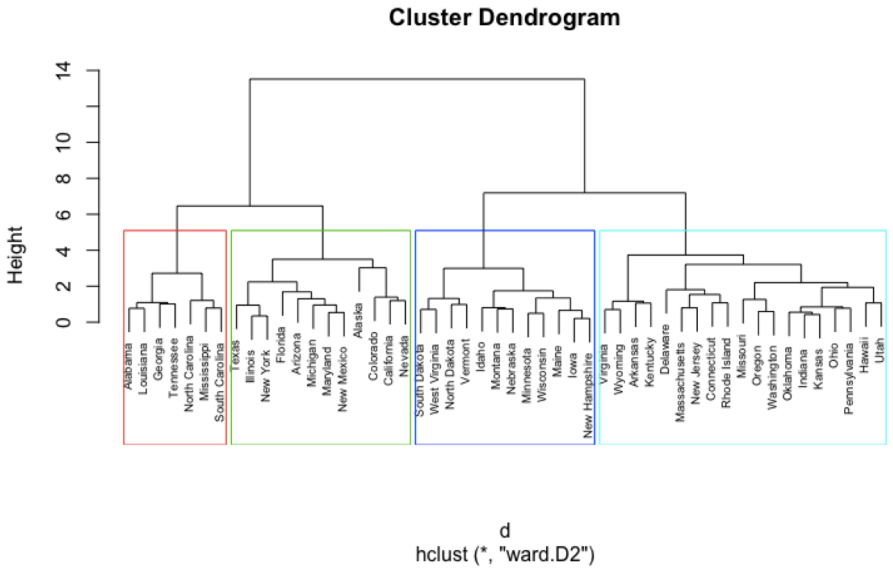
- Input: Training set $\{x_i | i = 1, \ldots, N\}$
- Output: Structure of training set, e.g., clusters

Representative Approaches

- Clustering
  - Represent each sample using a prototype example
- Dimensionality Reduction
  - Represent each sample using a small number of variables (e.g., feature selection, feature extraction)
- Density Estimation
  - Estimate the probability distribution over the data space

Unsupervised Learning

- **Clustering**
  - Grouping N samples into K clusters
  - We assume that the data were generated from a number of different classes (but, how many?)
- **Algorithms**
  - K-Means Clustering
  - Spectral Clustering
  - Hierarchical Clustering
  - Consensus Clustering
  - …



Unsupervised Learning

- **K-Means Clustering: Example**

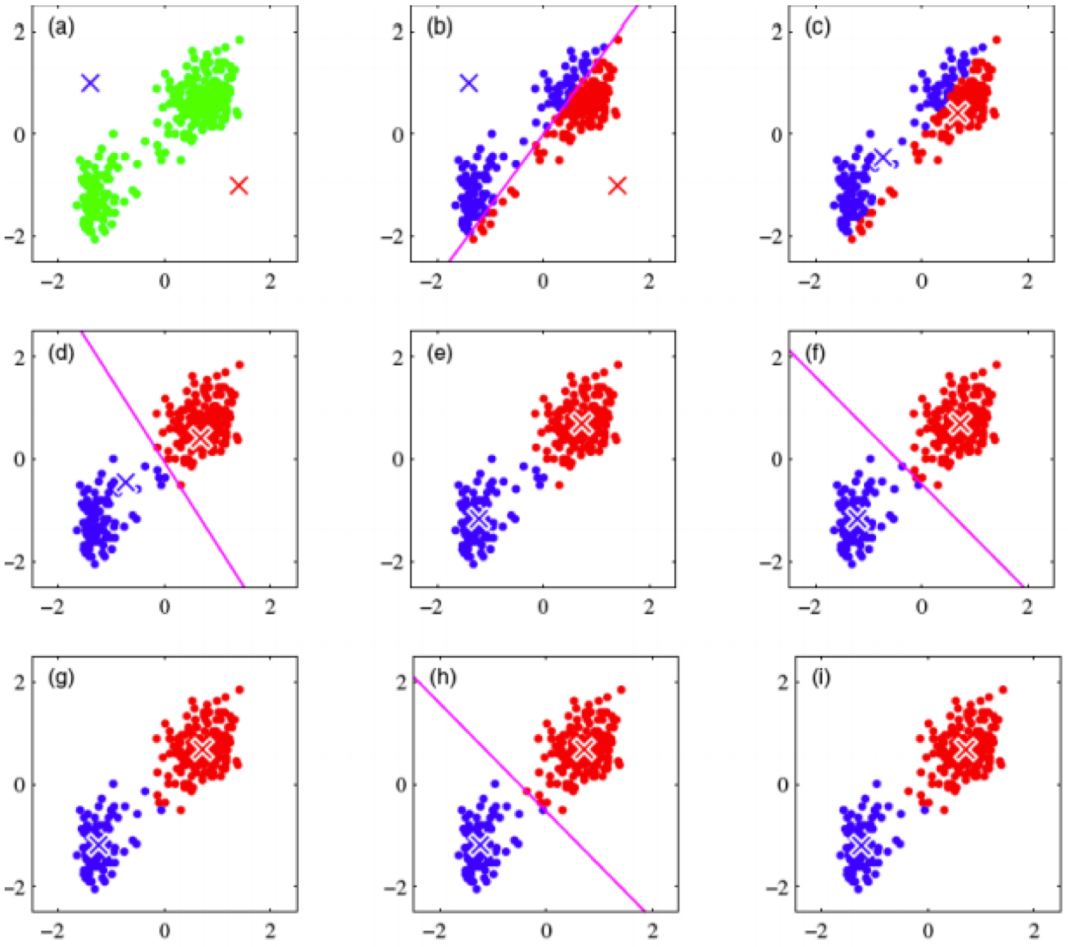


Figure from Bishop

Simple demo: http://syskall.com/kmeans.js/

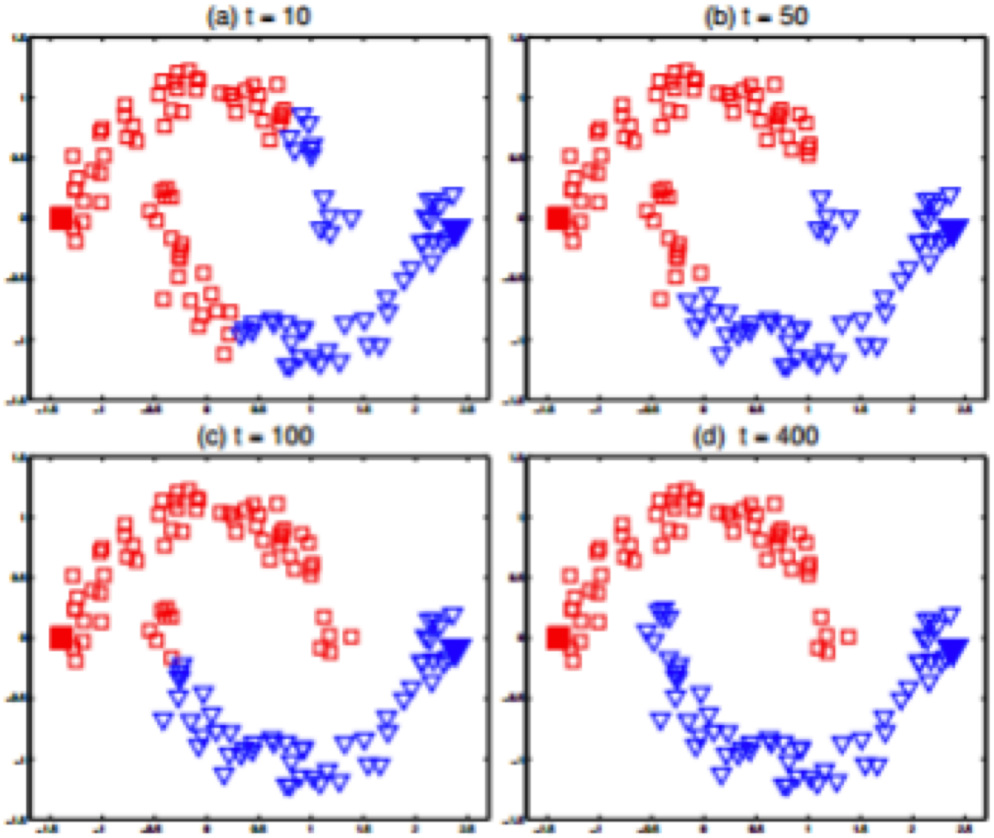
Semi-supervised Learning (SSL)

- Input: Training data include a labeled set, $\{(x_i, y_i)|i = 1, \dots, N_l\}$, and an unlabeled set $\{x_i|i = 1, \dots, N_u\}$
- Output: A good approximation to $f: X \to Y$

- Usually, the unlabeled set is *much larger* than the labeled set
- As labeling data is always expensive and time-consuming, SSL can take advantages of unlabeled data to boost the learning performance
- Approaches
  - Semi-supervised Regression
  - Graph-based Semi-Supervised Learning

Semi-supervised Learning (SSL)

- Graph-based Semi-Supervised Learning: Example



Label Propagation Results at Different Iterations

Model Evaluation

- How to evaluate how good is your model?

Supervised Learning Algorithms (e.g., classification)

- Classification accuracy
- Recall, Precision, ROC Curves
- F1 score, AUC score

Unsupervised Learning Algorithms (e.g., classification)

- Clustering accuracy (*if ground truth are available*)
- Normalized mutual information, NMI (*if ground truth are available*)
- Other internal measures

Other Advanced Topics

- Transfer Learning (supervised/unsupervised domain adaptation, multi-task learning, one-shot learning, zero-shot learning, etc.)
- Multi-view / Multi-source / Multimodal Learning
- Multi-label / Multi-instance Learning
- Online Learning
- Graphical Models, Generative Models
- Representation Learning
- Active Learning
- Reinforcement Learning
- Deep Learning
- …

Top-tier Machine Learning Venues

Conferences

- Annual Conference on Neural Information Processing Systems (NIPS)
- International Conference on Machine Learning (ICML)

Journals

- Journal of Machine Learning Research (JMLR)
- IEEE Trans. Pattern Analysis and Machine Intelligence (T-PAMI)

Top Conferences in Related Areas

Incomplete List

- Artificial Intelligence: IJCAI, AAAI
- ML Conference with a focus on deep learning: ICLR
- Computer Vision: ICCV, CVPR, ECCV
- Natural Language Processing: ACL, EMNLP, NAACL
- Data Mining/Information Retrieval: KDD, SIGIR, WWW, CIKM, WSDM