No. 1

Introduction

Hui Jiang

Department of Electrical Engineering and Computer Science

Lassonde School of Engineering

York University, Toronto, Canada

Course Info (tentative)

- **Instructor:**

  Hui Jiang (hj@cse.yorku.ca)
- **Course web site:**

  https://www.eecs.yorku.ca/course/6327/
- **Course Format:**
  - **Lectures (40 hours):**
    - **Covers basic probabilistic models, pattern classification theory, machine learning algorithms;**
    - **Selected coverage on advanced machine learning topics.**
- **Evaluation:**
  - **Two assignments (25%)**
  - **Two lab projects (50%)**
  - **Exam or In-class presentation (25%)**

Course Outline

- Part I: Introduction (6 hours)
  - Machine Learning: basic concepts
  - Math foundation: review

- Part II: Basic theory of pattern classification and machine learning (24 hours)
  - Bayesian decision rule; Model Estimation
  - Discriminative models: SVM, Neural networks (NN) and beyond
  - Generative models: Gaussian, GMM, Markov Chain, HMM, Graphical models

- Part III: Two lab projects (6 hours)
  - Write lab report as a conference paper
  - In-class presentation

Reference Materials

- **Lecture notes**
- **Assigned reading materials throughout the course**
- **Reference books:**

[1] Pattern Recognition and Machine Learning by C. M. Bishop.
(Springer, ISBN 0-387-31073-8)

[2] *Pattern Classification* by R. O. Duda, P. Hart and D. Stork.
(John Wiley & Sons, Inc., ISBN 0-471-05669-3)

[3] Machine Learning: A Probabilistic Perspectives by *K. P. Murphy.*
(The MIT Press, ISBN 978-0-262-01802-9)

[4] Deep Learning by I Goodfellow, Y. Bengio and A. Courville
(The MIT Press, ISBN 9780262035613)

- **Prerequisite:**
  - Calculus, probability and statistics
  - Linear algebra and/or matrix theory
  - C/C++/Java/python; matlab; python/shell (plus)

Relevant AI Research Topics

- Theory
  - ✓ Knowledge Representation and Inference
  - ✓ Machine Learning
  - ✓ Pattern Recognition
  - ✓ Statistical Signal Processing
- Applications
  - ✓ Speech Processing
  - ✓ Natural Language Processing
  - ✓ Computer Vision
  - ✓ Data Mining
  - ✓ Robotics
  - ✓ …

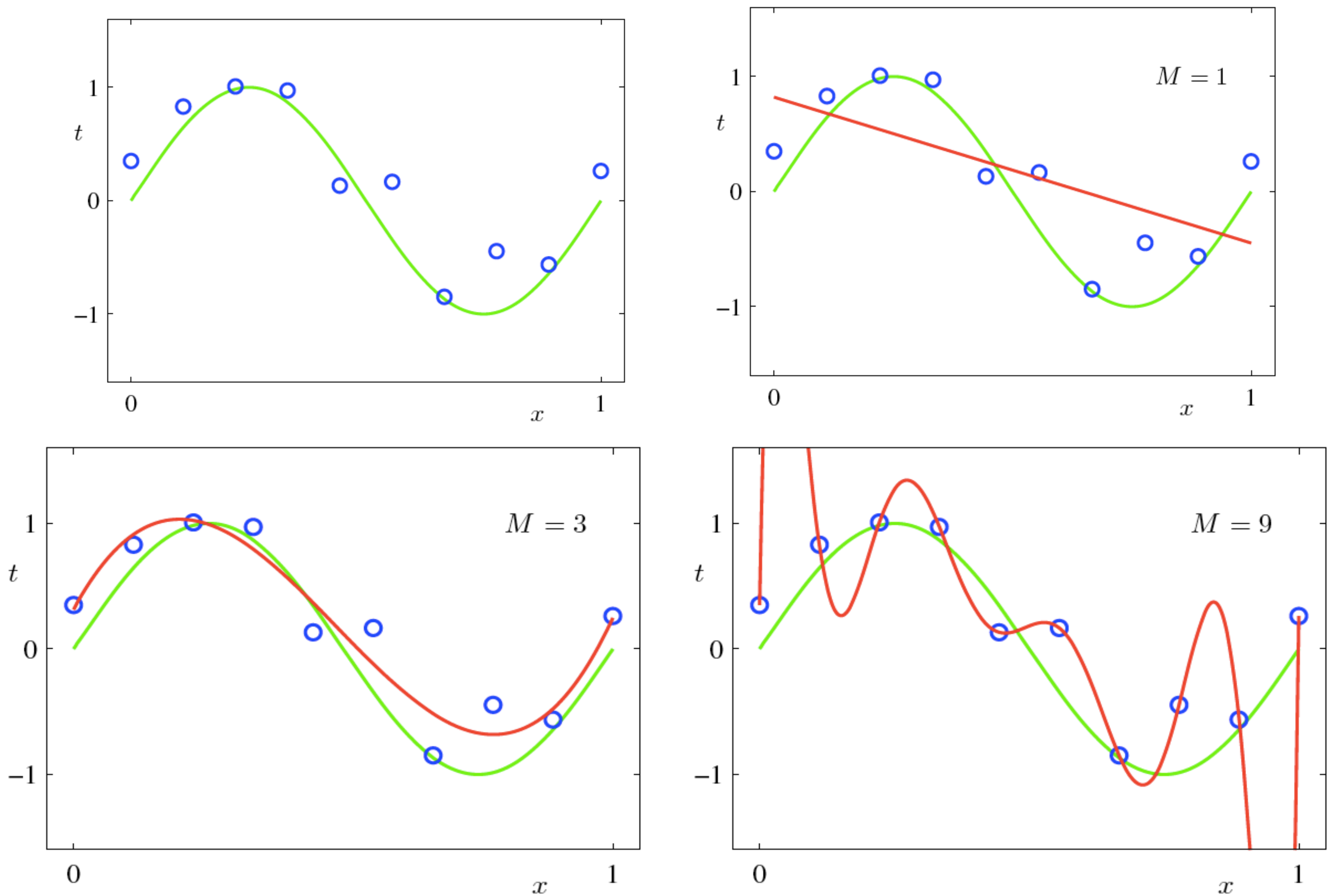
Artificial Intelligence (AI): Paradigm Shift

- Knowledge based ➔ KR
  - Reply on expert(s); Small data samples
  - Simple toy problems
- Data-Driven ➔ ML
  - Large data samples
  - Statistical models; machine learning algorithms
- Big Data + Big Model Era
  - Massive real-world data samples ➔ powerful models
  - Data intensive computing ➔ computation power
  - Parallel/distributed platform: e.g. GPU, map-reduce

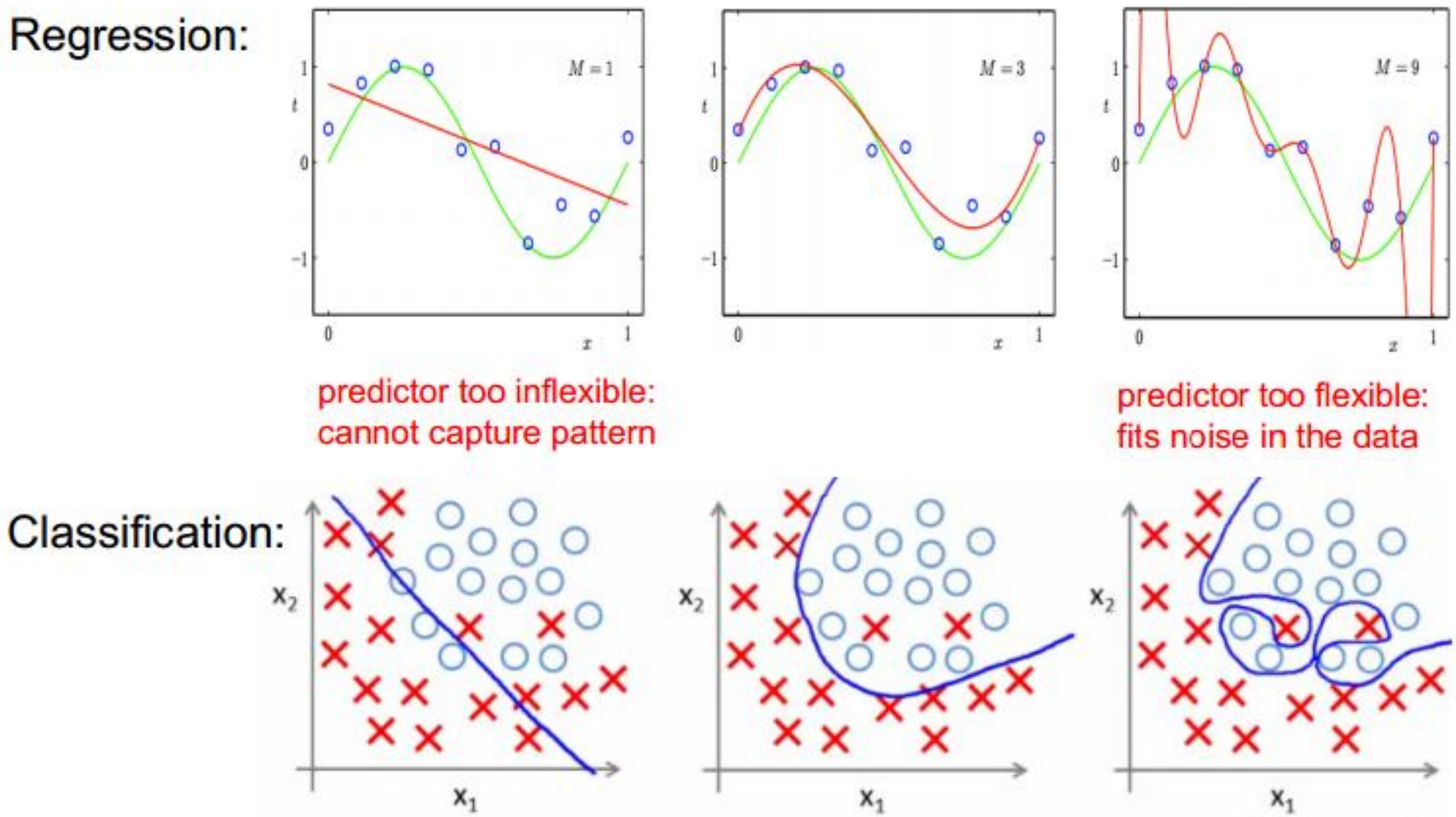
Some Machine Learning Concepts

- Classification vs. Regression
- Supervised vs. Unsupervised (Clustering)
- Linear (simple) vs. Nonlinear (complex) models
- Underfitting vs. Overfitting (Regularization)
- Parametric vs. Non-parametric
- Frequentist vs. Bayesian
- Statistic models vs. Rule-based (ML vs. AI)

An Example: Curve fitting

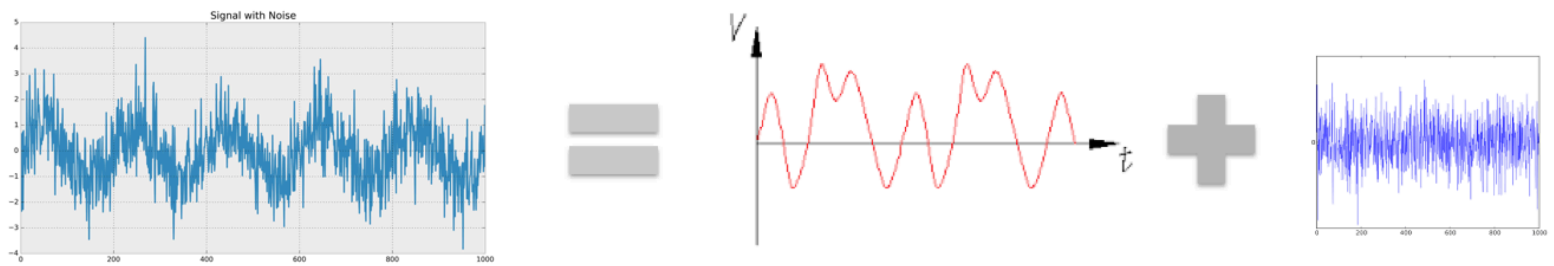


Under-fitting vs. Overfillting (Regularization)



Under-fitting vs. Overfillting (Regularization)

data = signal + noise



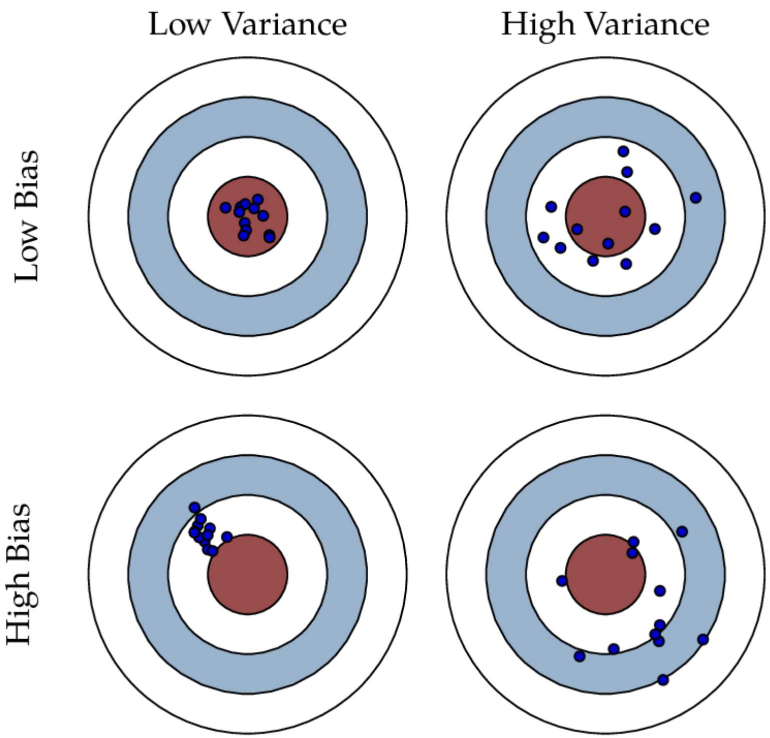
- **Weak models ➔ under-fitting**
- **Too complex models ➔ over-fitting (why?)**

Bias-Variance Trade-off

- Simple model ➔ under-fitting ➔ high bias
- Complex models ➔ over-fitting ➔ high variance
- Expected error = (bias)^2 + variance

$$\text{true model: } y = f(x)$$
$$\text{learned model: } y = \hat{f}(x)$$

$$\text{Expected error: } E[(f-\hat{f})^2] = \underbrace{\left(f - E(\hat{f})\right)^2}_{bias} + \underbrace{E\left[\left(\hat{f} - E(\hat{f})\right)^2\right]}_{variance}$$



Some General Principles in Machine Learning

- Bias-variance tradeoff
- Curse of dimensionality
- No free lunch theorem
- Local constancy prior

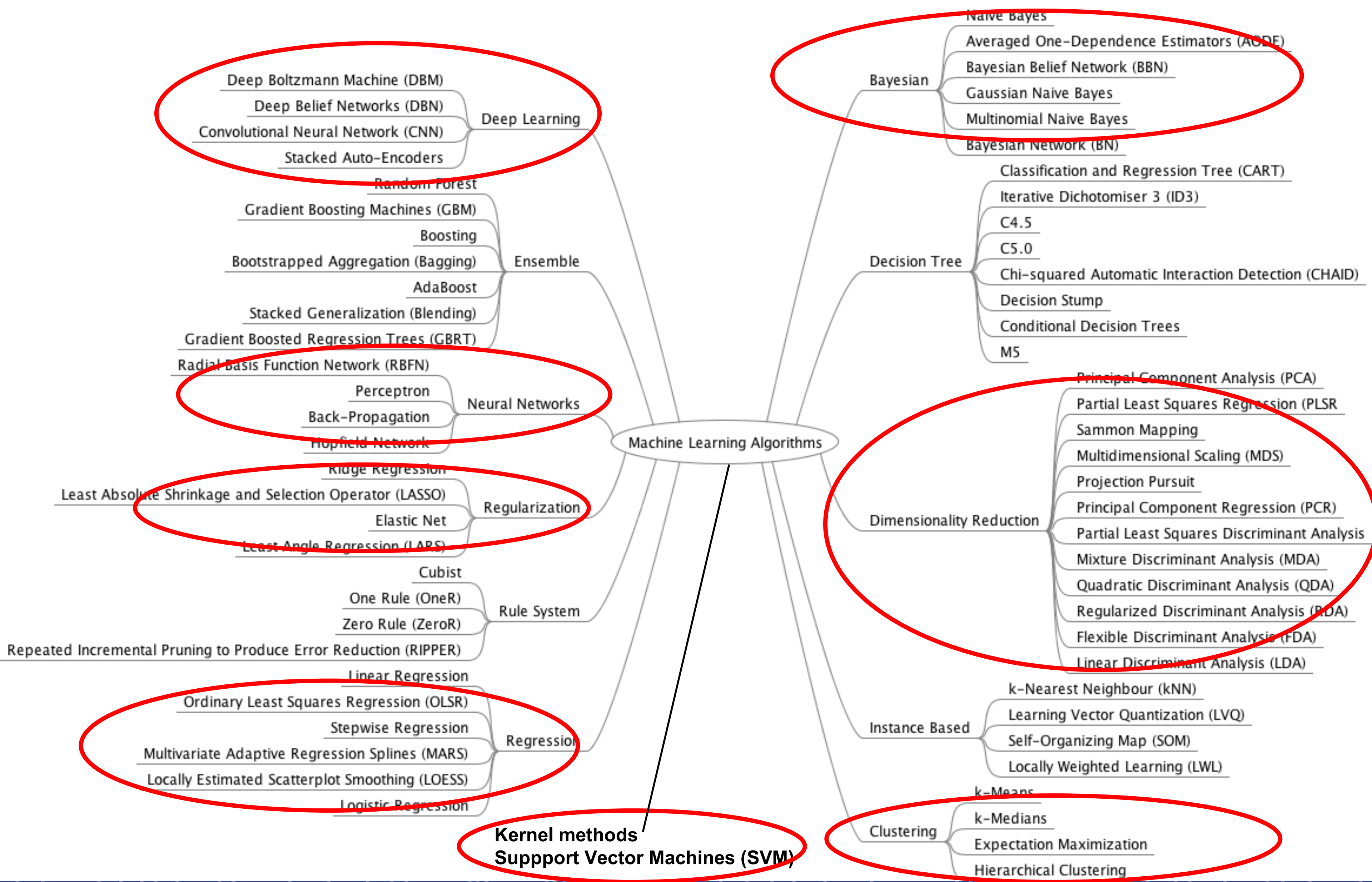
Machine Learning Procedure

- Feature extraction (feature engineering):
  - Need to know objects to extract good features
  - Varies a lot among different applications (speech, audio, text, image, video, gestures, biological sequences, etc)
  - May need reduce dimensionality
- Training: statistical model learning
- Testing:Inference, matching, decision

The basic theories common to various applications

Machine Learning Algorithms

Machine Learning Algorithms



Advanced ML Topics

- Learnability
- On-line Learning
- Reinforcement Learning
- Transfer Learning / Adaptation / One-shot Learning
- Active Learning
- Ensemble Learning
- Imitation Learning
- Gaussian Processes
- Causal Learning

Project One (tentative)

- Project one (20%): machine learning algorithms and models
  - Use a popular data set MNIST (http://yann.lecun.com/exdb/mnist/)
  - Feature extraction, data virtualization
  - Linear regression, logistic regression
  - Linear/nonlinear SVM
  - Neural networks

- Need your own implementation, not just function calls
- Submit all of your codes/scripts and a project report
- Evaluation depends on your implementation, report and performance

Project Two (tentative)

- Project two (30%): machine learning related research
  - Define your own research problem
  - Select your own models (deep learning, graphical models, …)
  - Choose any open source toolkit
  - Link to your advanced study topic
  - Link to your own research areas
- Write me 1-page proposal (500 words) for approval
- Submit codes and a report (as a 8-page conference paper)
- A short presentation (10-15 minutes) in class
- Evaluation: problem, idea, method, experiments, writing and presentation …