Probabistic Models and Machine Learning

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;**
    - **Self-study on some advanced machine learning topics.**
- **Evaluation:**
  - **Class Participation (10%)**
  - **Two assignments (30%)**
  - **Two lab projects (40%)**
  - **In-class presentation (20%)**

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
  - Generative models: Gaussian, GMM, Markov Chain, HMM, Graphical models
  - Discriminative models: SVM, Neural networks (NN) and beyond

- **Part III: Advanced Topics (6 hours)**
  - Self-select and self-study
  - Presentation

Reference Materials

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

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

  [2] *Pattern Classification* (2nd Edition) 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)*
- **Prerequisite:**
  - Probability and statistics
  - Linear algebra and/or matrix theory
  - C/C++/Java; matlab; perl/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
  - ✓ …

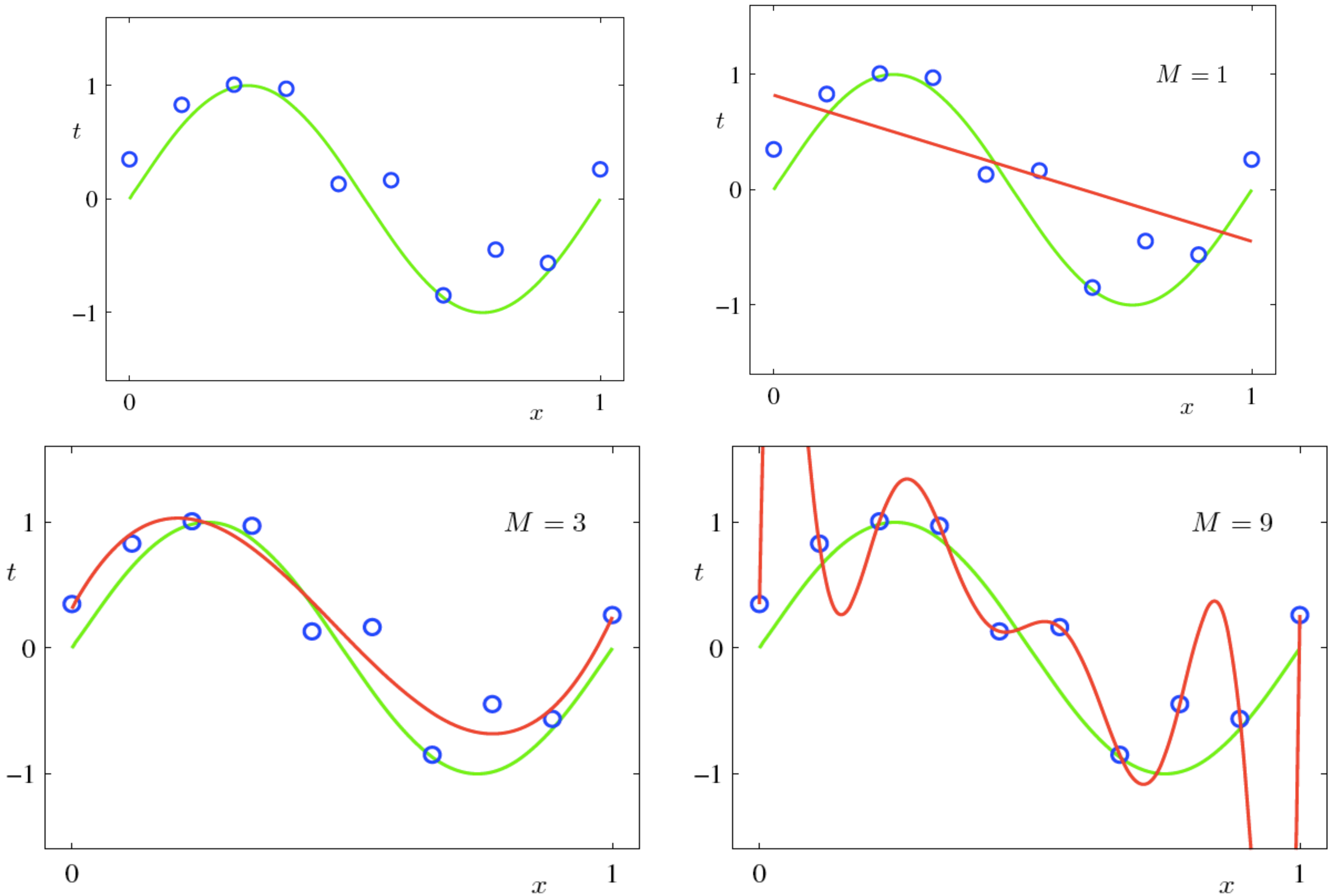
Artificial Intelligence (AI): Paradigm Shift

- 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 Era
  - Massive real-world data samples ➔ powerful models
  - Data intensive computing ➔ computation power
  - Parallel/distributed platform: e.g. GPU, map-reduce

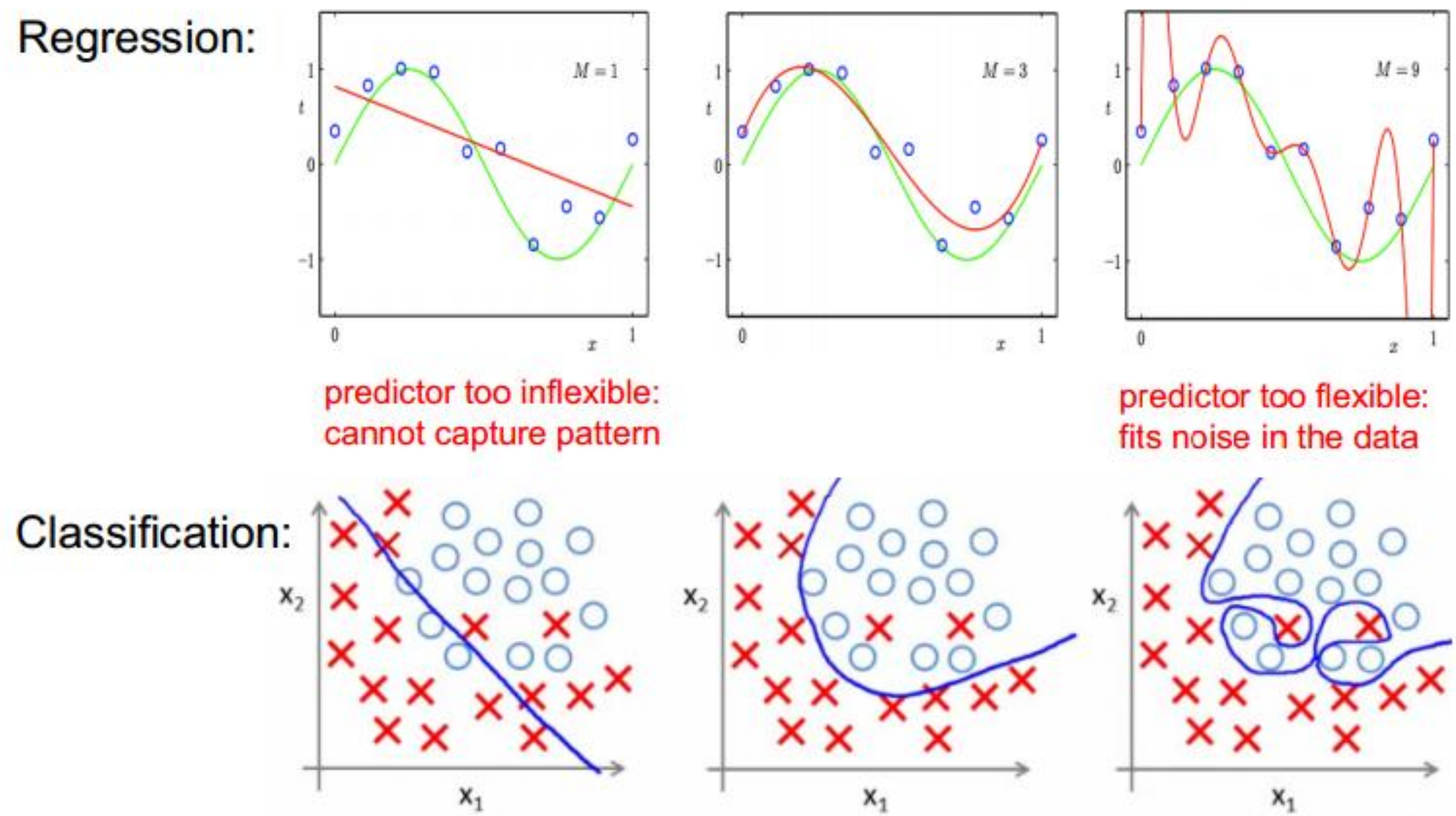
Some Machine Learning Concepts

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

An Example: Curve fitting

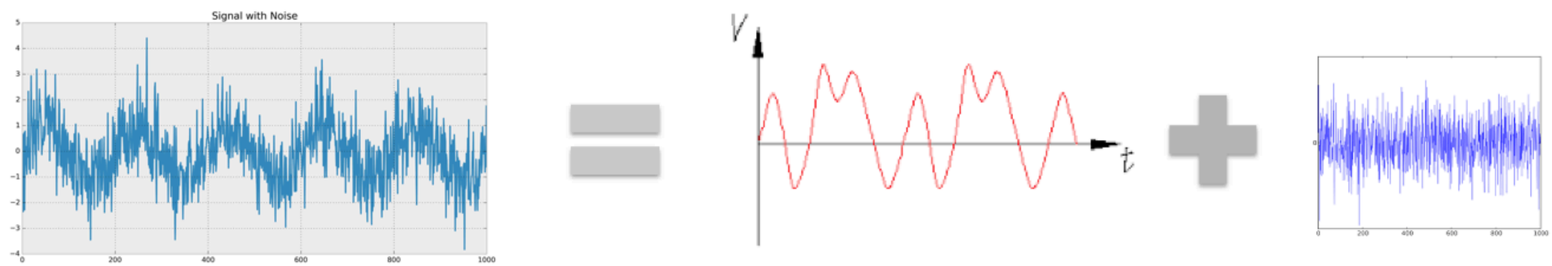


Under-fitting vs. Overfillting (Regularization)



Under-fitting vs. Overfillting (Regularization)

data = signal + noise



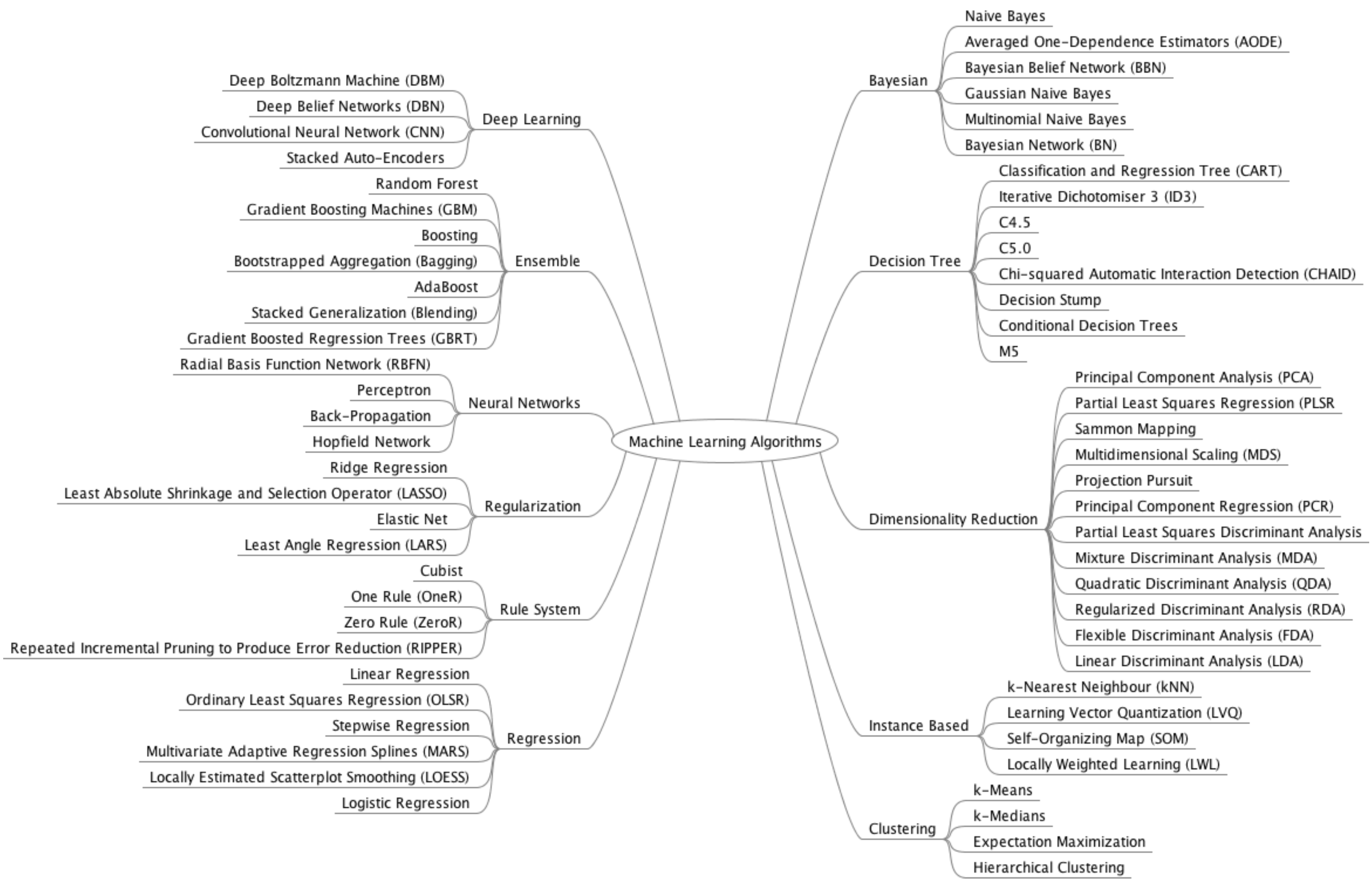
- **Weak models ➔ under-fitting**
- **Too strong models ➔ over-fitting (why?)**

Machine Learning Procedure

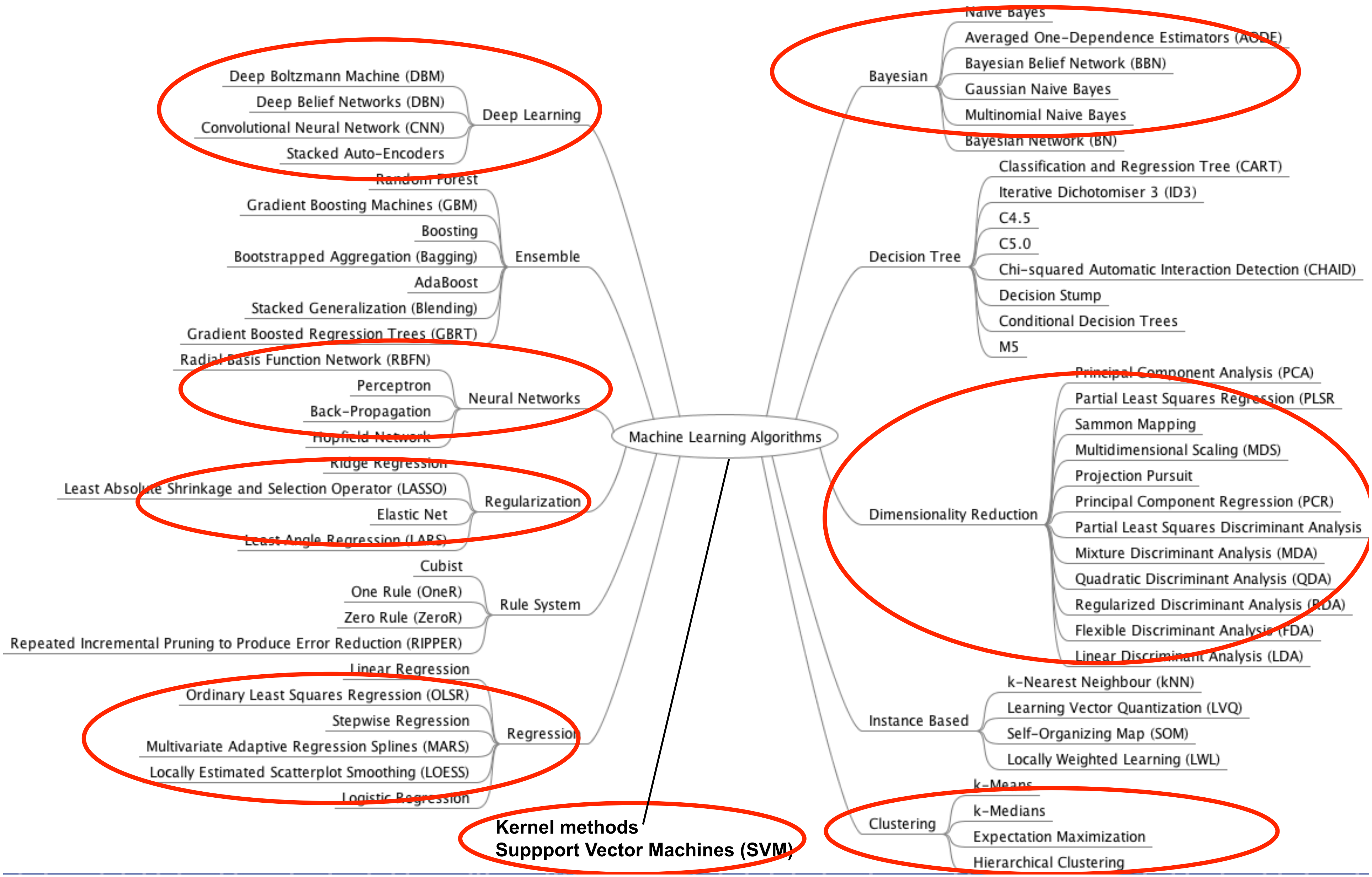
- Feature extraction (feature engineering):
  - Need to know objects to extract good features
  - Varies a lot among different applications (speech, audio, text, image, video, gestures, biological sequences, etc)
  - May need reduce dimensionality
- Statistical model training/learning
- Inference, matching,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