MACHINE LEARNING

CHAPTER 1: INTRODUCTION

Contact Information

Instructor: Qi Hao

E-mail: hao.q@sustc.edu.cn

Office: Nanshan Zhiyuan A7 Room 906

Office Hours: M 2:00-4:00pm

Available other times by appointment or the open door policy

Office Phone: (0755) 8801-8537

QQ: 463715202 机器学习2017

Web: *http://hqlab.sustc.science/teaching/*

Class Schedule

- **Lectures**: T 8:00 am - 9:50 am Teaching Building 1 Room 405
- **Computer Lab**: W 10:20 am – 12:10 pm LiYuan Park Room 406
- **Grading policy**

In-class test :	20%	Quiz (5~10 times):	20%
Assignment (8~10 times):	20%		
Projects (2 per group):	20%	Final Projects:	20%

90~93: A-	94~97: A	98~100: A+
80~82: B-	83~86: B	87~89: B+
70~72: C-	73~76: C	77~79: C+
60~62: D-	63~66: D	67~69: D+

Textbook and Lecture Notes

Textbooks:

[1] Pattern Recognition and Machine Learning, by Christopher M. Bishop, 2006 Springer

Other books:

[1] Deep Learning, by Bengio

[2] Reinforcement Learning: An Introduction, by Richard S. Sutton

[3] The Elements of Statistical Learning Data Mining, Inference, and Prediction

Paper reading:

[1] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature, 2015, 521(7553):452-9.

[2] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553):436-44.

[3] Littman M L. Reinforcement learning improves behaviour from evaluative feedback. Nature, 2015, 521(7553):445-51.

Lecture notes:

http://hqlab.sustc.science/teaching/

Other Resources

Assignment platform: sakai.sustc.edu.cn

Textbook resource: https://www.microsoft.com/en-us/research/people/cmbishop/#prml-book

Matlab implementation: http://prml.github.io/

Teaching Objectives

- Fundamental knowledge about machine learning and pattern recognition, from Bayesian approaches to deep learning frameworks through lectures, quizzes and assignments
- Machine learning system development methods with Matlab/Python through labs and projects
- Model-based and data-driven machine learning system design and integration skills through the final project, literature surveys and reports

Schedule

Section 0	Course Introduction
Section 1	Probability Distributions
Section 2	Linear Models for Regression and Classification
Section 3	Neural Networks
Section 4	Kernel Methods and Sparse Kernel Machine
Section 5	Graphical Models
Section 6	Mixture Models and EM learning
Section 7	Approximate Inference
Section 8	Sequential Data
Section 9	Bayesian Networks
Section 10	Manifold Learning
Section 11	Deep Learning
Section 12	Reinforcement Learning

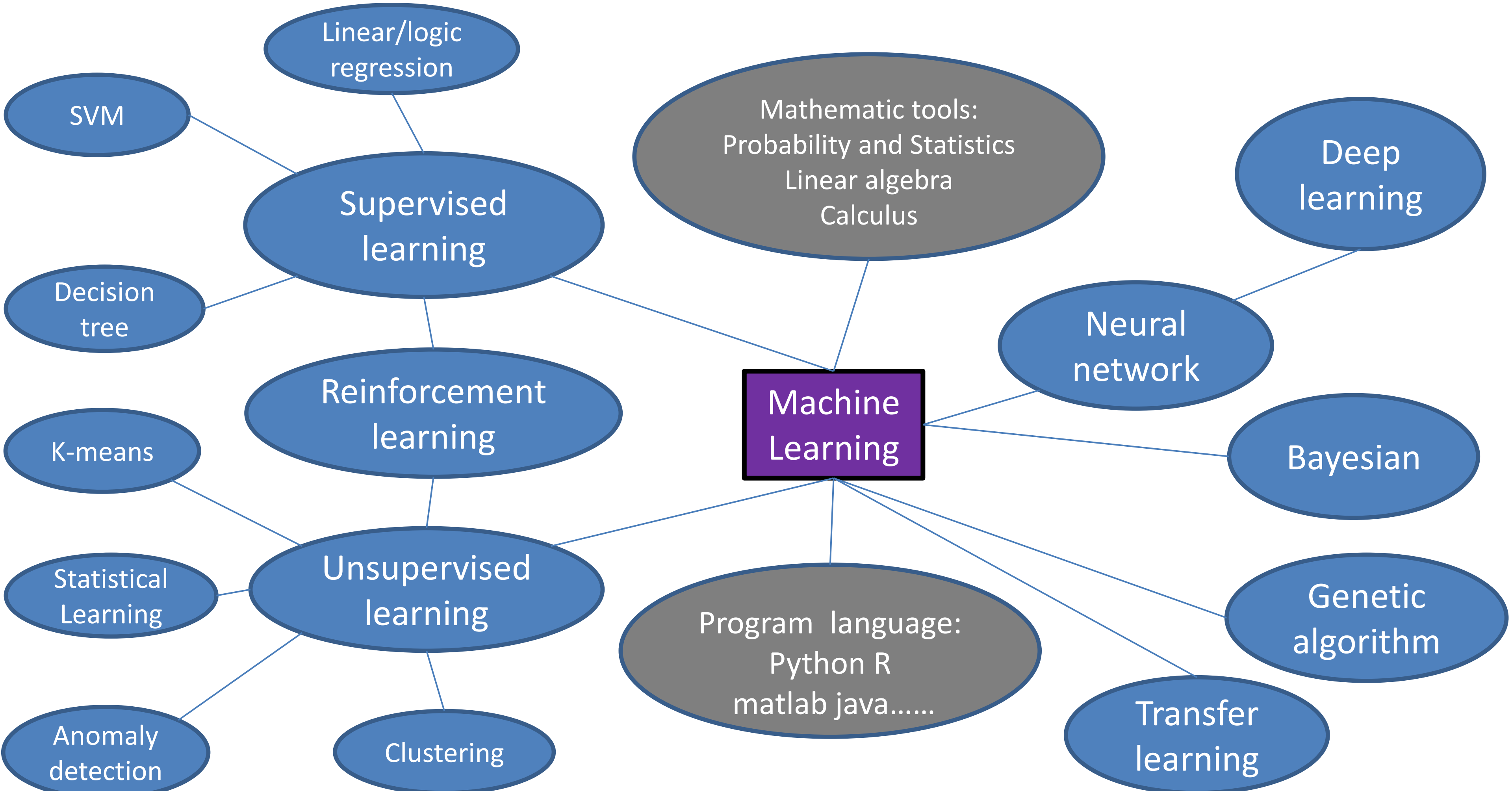
Outlines

- Framework
- Problem Statement
- Related Areas
- History
- Optimization for Machine Learning
- Algorithms
- Examples

Outlines

- Framework
- Problem Statement
- Related Areas
- History
- Optimization for Machine Learning
- Algorithms
- Examples

Framework


Linear/logic regression
SVM
Supervised learning
Decision tree
Mathematic tools:
Probability and Statistics
Linear algebra
Calculus
Deep learning
Neural network
Machine Learning
Reinforcement learning
K-means
Bayesian
Unsupervised learning
Statistical Learning
Program language:
Python R
matlab java……
Genetic algorithm
Transfer learning
Anomaly detection
Clustering

Outlines

- Framework
- Problem Statement
- Related Areas
- History
- Optimization for Machine Learning
- Algorithms
- Examples

Problem Statement

- Problem： Predict the label $\hat{Y}$ and data $\hat{X}$ with training set（X,Y）？

Unknown model

Training label　　Training data

Training： $Y = \underline{A} \circ X$

Predicting： $\hat{Y} = A \circ \hat{X}$

Predicting label　　Predicting data

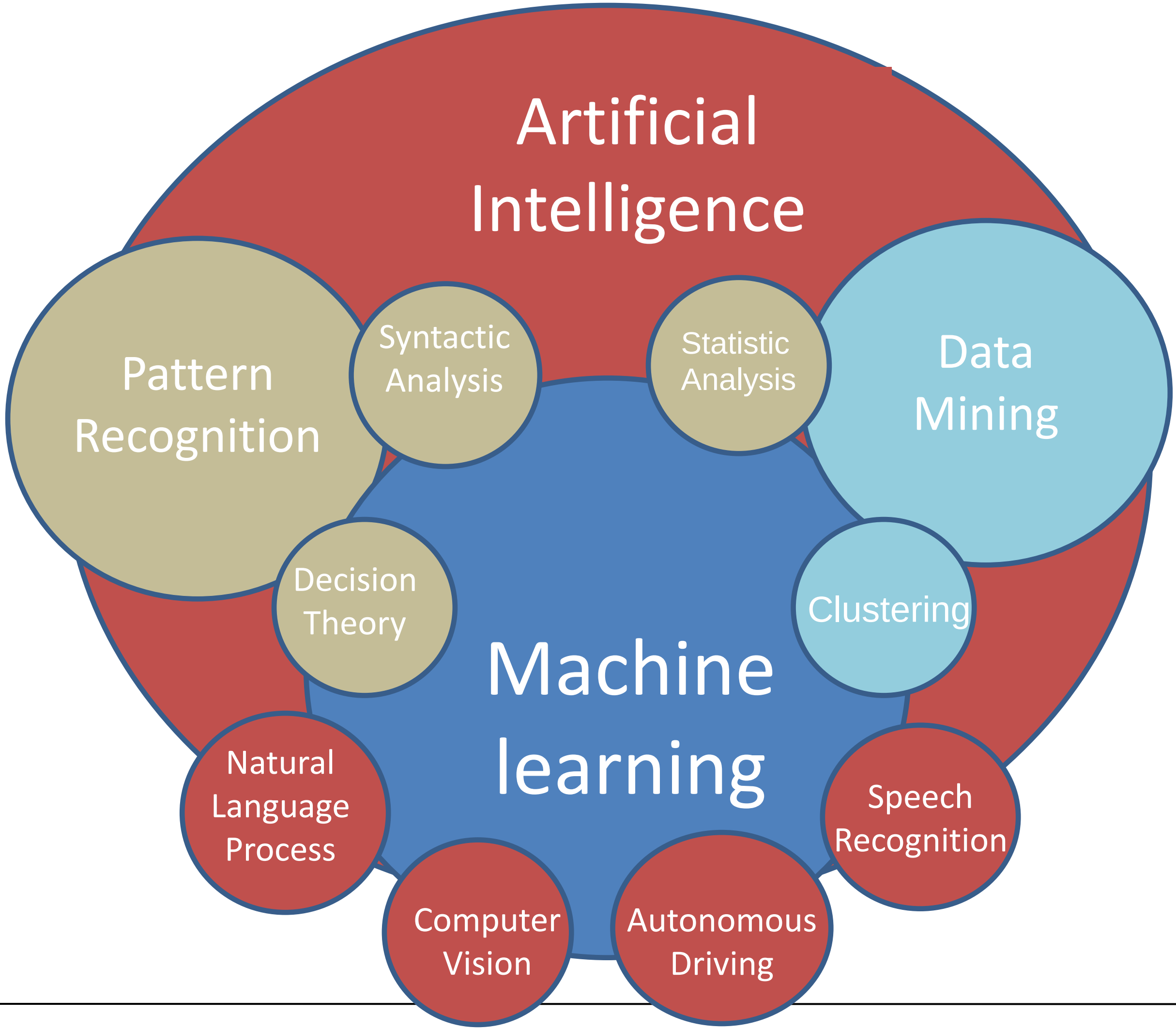
$$\begin{cases} Y\ or\ X\ is\ known\text{： supervised learning} \\ Y\ or\ X\ is\ unknown\text{： unsupervised learning} \end{cases}$$

$$\begin{cases} Y、\hat{Y}\ are\ continuous\text{： Regression} \\ Y、\hat{Y}\ are\ discrete\text{： classification} \end{cases}$$

Outlines

- Framework
- Problem Statement
- Related Areas
- History
- Optimization for Machine Learning
- Algorithms
- Examples

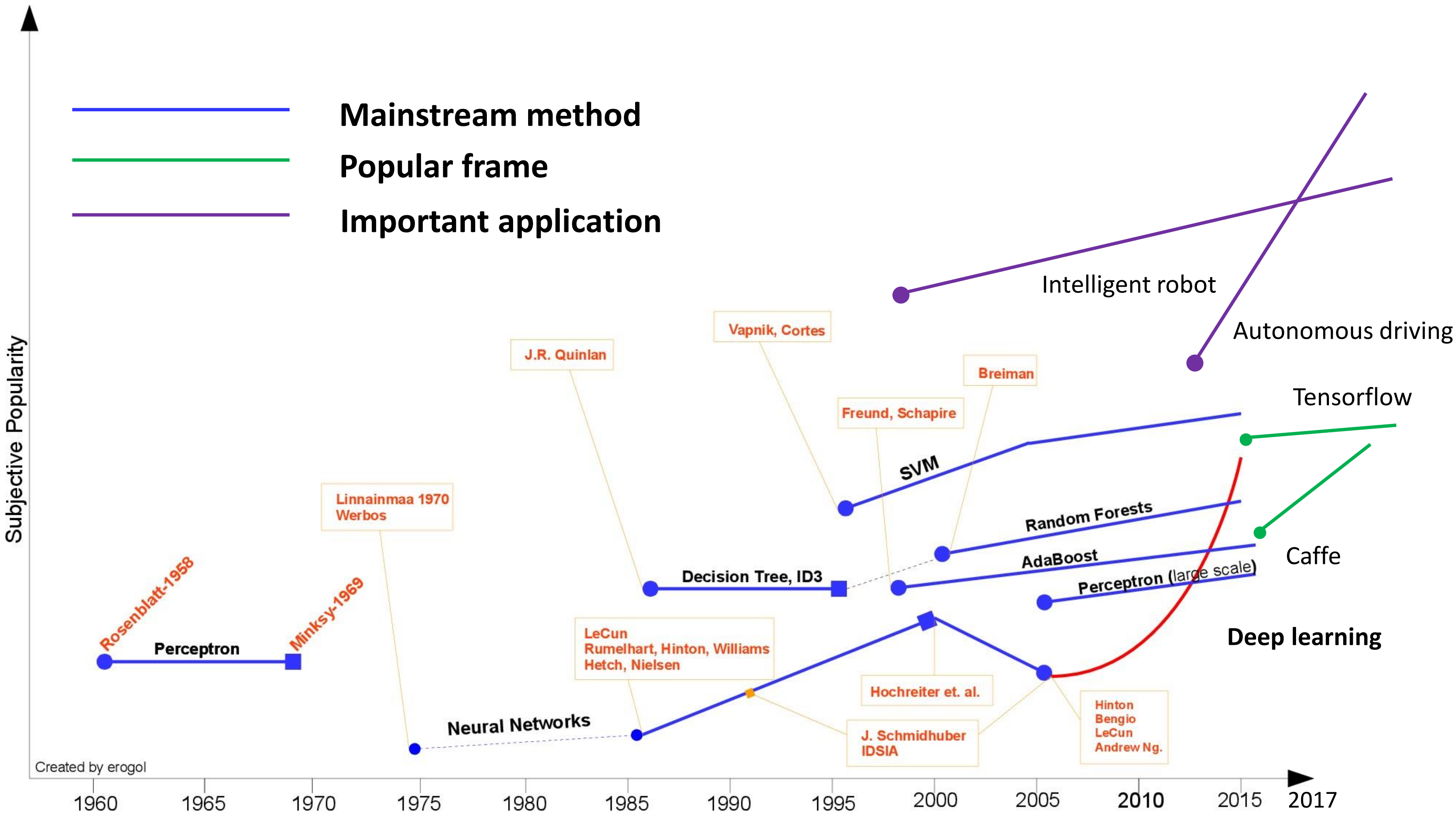
The Whole Picture



Outlines

- Framework
- Problem Statement
- Related Areas
- History
- Optimization for Machine Learning
- Algorithms
- Examples

History



Mainstream method
Popular frame
Important application
Subjective Popularity
Rosenblatt-1958
Perceptron
Minksy-1969
Linnainmaa 1970
Werbos
Neural Networks
J.R. Quinlan
Decision Tree, ID3
LeCun
Rumelhart, Hinton, Williams
Hetch, Nielsen
Vapnik, Cortes
SVM
Freund, Schapire
Breiman
Random Forests
AdaBoost
Perceptron (large scale)
Hochreiter et. al.
J. Schmidhuber
IDSIA
Hinton
Bengio
LeCun
Andrew Ng.
Deep learning
Intelligent robot
Autonomous driving
Tensorflow
Caffe
Created by erogol
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2017

Outlines

- Framework
- Problem Statement
- Related Areas
- History
- Optimization for Machine Learning
- Algorithms
- Examples

Machine learning and Optimization

- **Machine Learning—many machine learning problems are formulated as minimization of some loss function.**
- **Optimization algorithms can minimize the loss.**

What is optimization?

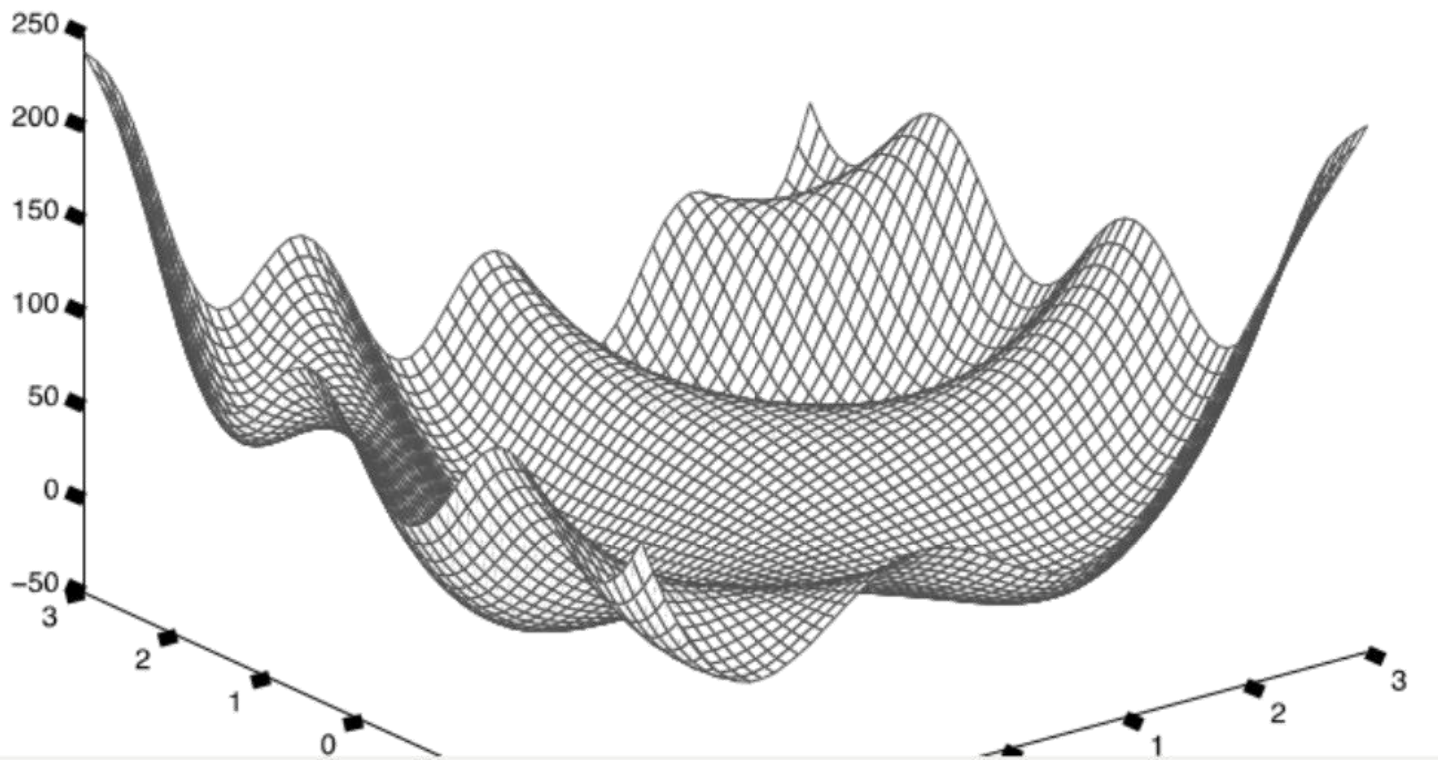
- Finding (one or more) minimizer of a function subject to constraints

$$\arg\min_{x} \ f_0(x)$$
$$s.t.\, f_i(x) \le 0, i = \{1, \ldots, k\}$$
$$f_j(x) = 0, i = \{1, \ldots, l\}$$

- Most of the machine learning problems are, in the end, optimization problems

General Problem

- Minimize f(x)



Optimization

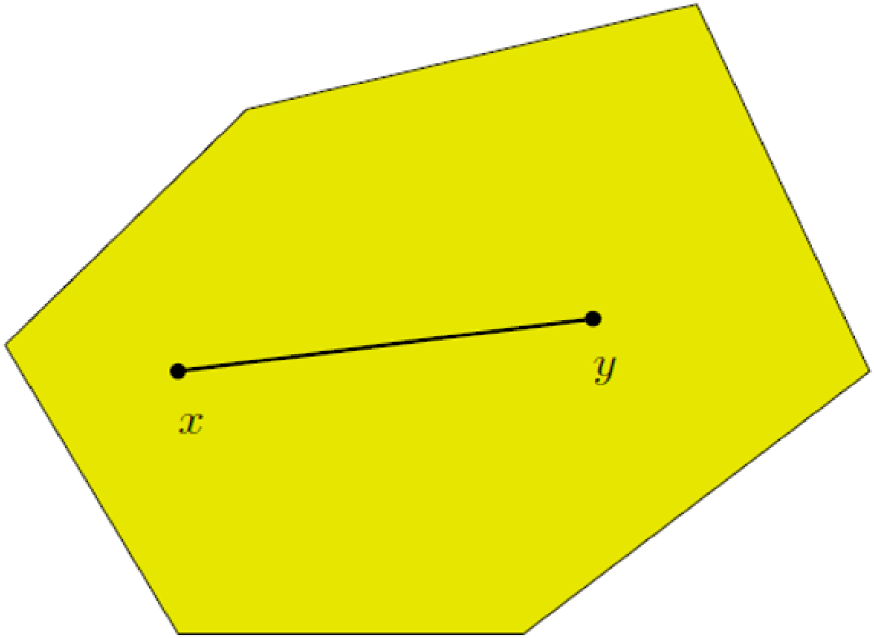
- Convex
  - Unconstrained optimization
  - Constrained optimization
- Non-convex
  - Heuristic algorithms
  - Neural networks

What is Convex?

- Convex sets

Def: A set $C \subseteq R$ is convex if for $x,y \in C$; $a \in [0,1]$

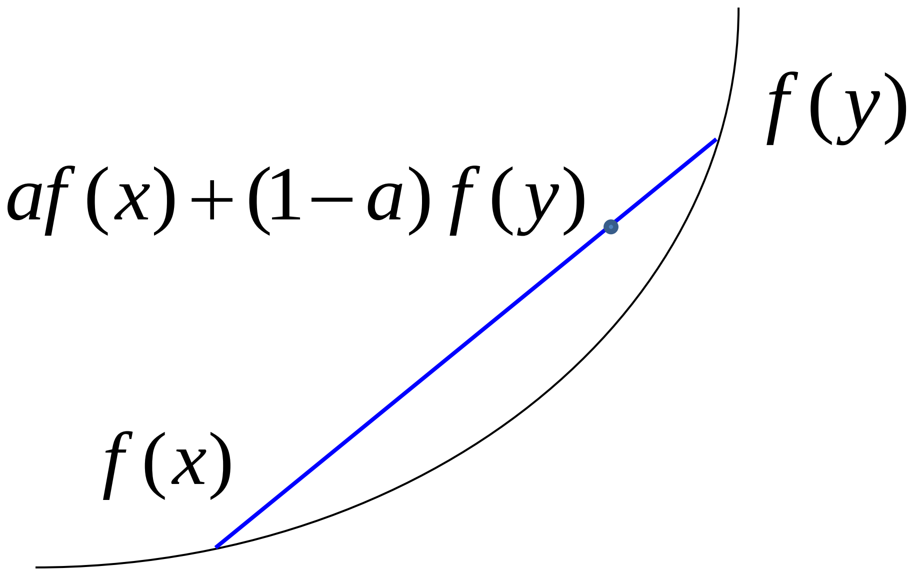
$$ax + (1-a)y \in C$$



What is Convex?

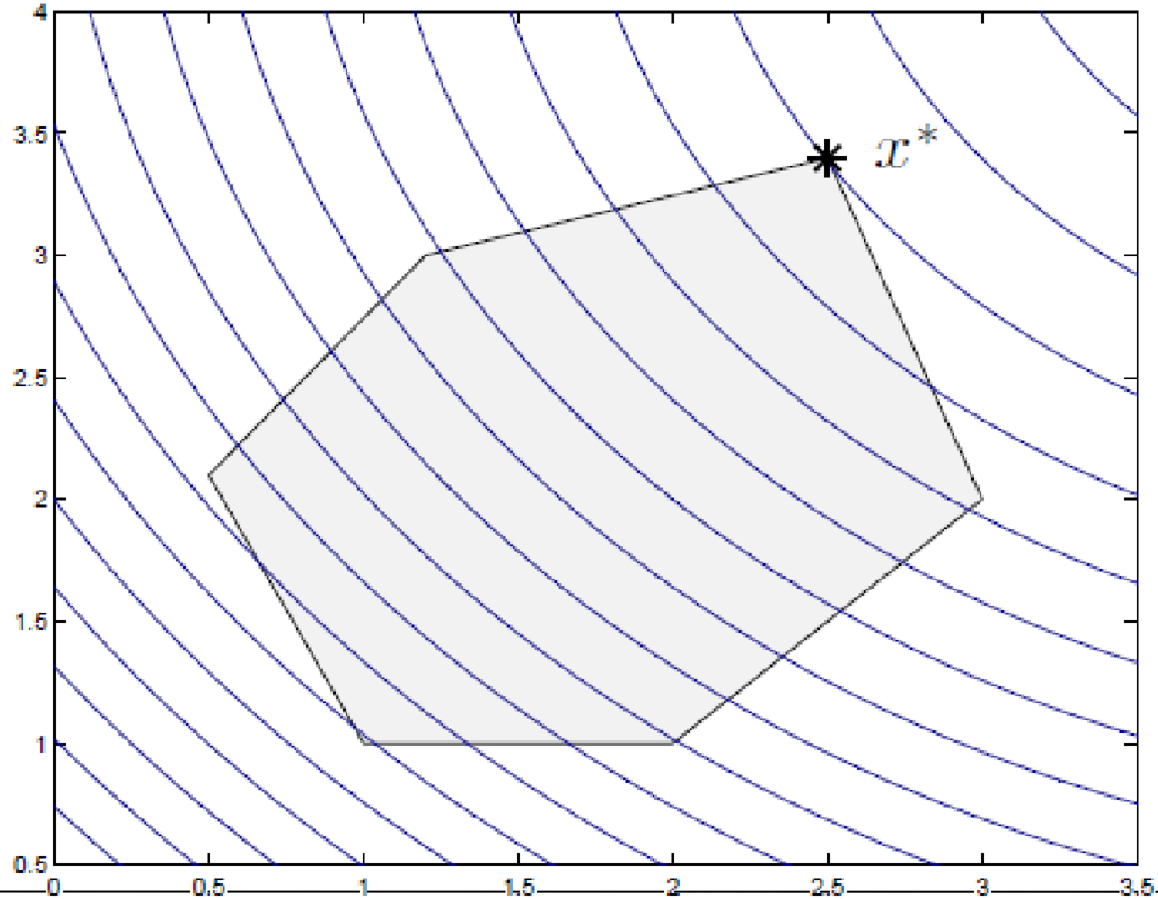
- Convex functions

$$f(ax+(1-a)y) \le af(x)+(1-a)f(y)$$



Convex optimization

- Local minimizer = Global minimizer



Convex optimization

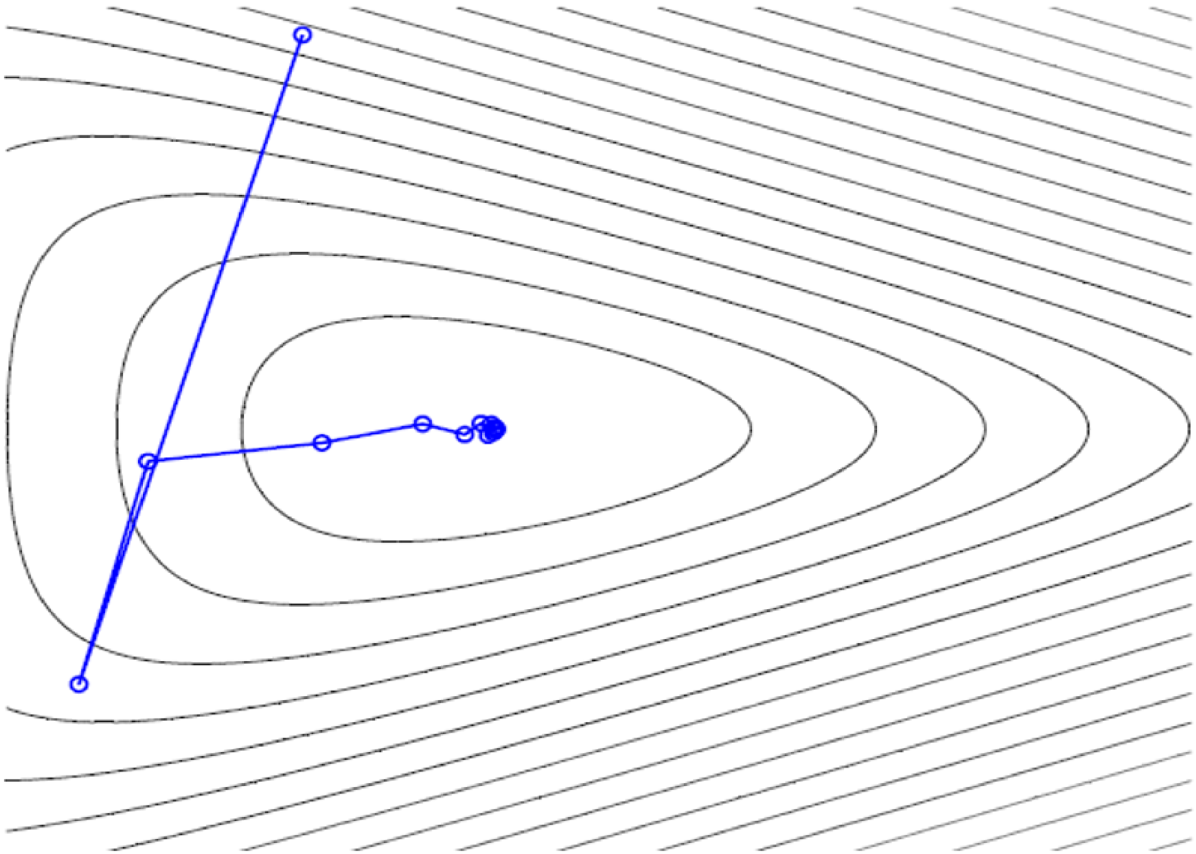
- Unconstrained optimization
  - Gradient descent
  - Newton's method
  - Batch learning
  - Stochastic Gradient Descent
- Constrained optimization
  - Lagrange function
  - General methods

Convex optimization

- Unconstrained optimization
  - Gradient descent
  - Newton’s method
  - Batch learning
  - Stochastic Gradient Descent
- Constrained optimization
  - Lagrange function
  - General methods

Gradient descent

$$f(x_{t+1}) = f(x_t) - \eta \nabla f(x_t)^T (x - x_t)$$



Newton's method

- Idea: use a second-order approximation to function

$$f(x+\Delta x) \approx f(x) + \nabla f(x)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x) \Delta x$$

- Choose Δx to minimize above:

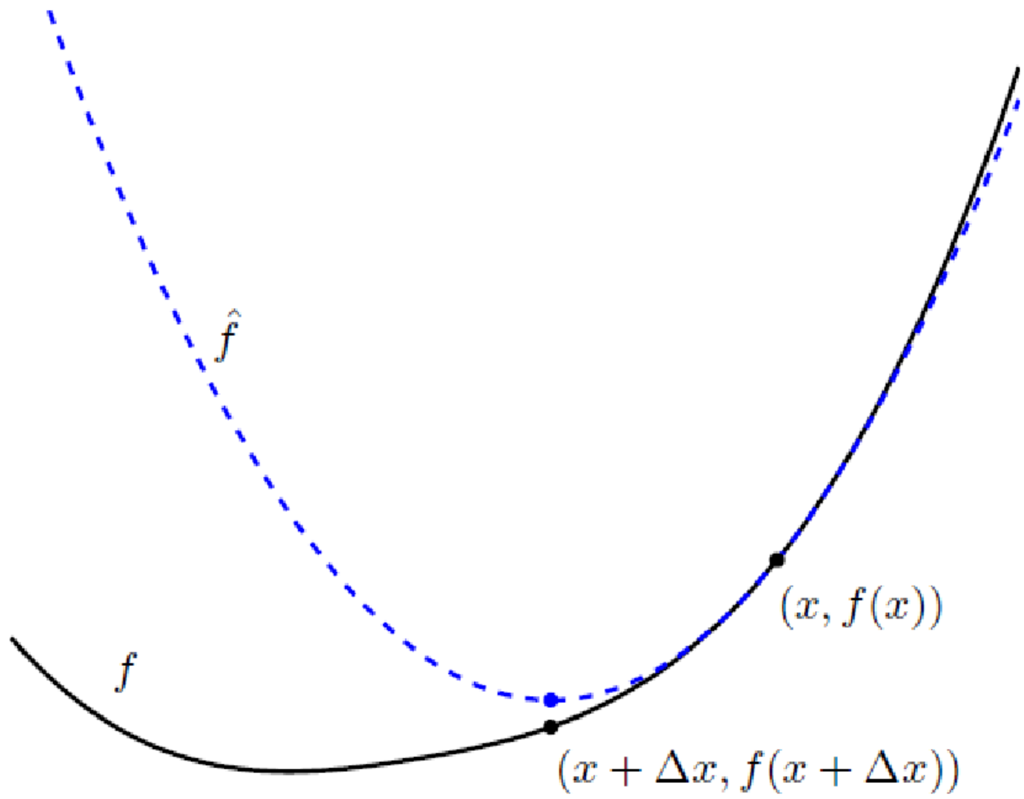
$$\Delta x = -[\nabla^2 f(x)]^{-1} \nabla f(x)$$

- This is descent direction:

$$\nabla f(x)^T \Delta x = -\nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x) < 0$$

Newton's method

$\hat{f}$ is 2-order approximation, f is true function.



Batch gradient descent

- Minimize empirical loss, assuming it's convex and unconstrained
  - Gradient descent on the empirical loss
  - At each step:

$$w^{(k+1)} \leftarrow w^{(k)} - \eta_t \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial L(w, x_i, y_i)}{\partial w}\right)$$

  - Note: ate each step, gradient is the average of the gradient for all samples (i=1,…n)
  - Very slow when n is very large

Stochastic gradient descent

- Alternative: compute gradient from just one (or a few samples)
- Known as stochastic gradient descent:
  - At each step,

$$w^{(k+1)} \leftarrow w^{(k)} - \eta_t \frac{\partial L(w, x_i, y_i)}{\partial w}$$

(choose one sample i and compute gradient for that sample only)

Convex optimization

- Unconstrained optimization
  - Gradient descent
  - Newton's method
  - Batch learning
  - Stochastic Gradient Descent
- Constrained optimization
  - Lagrange function
  - General methods

Lagrange function

- Start with optimization problem:

$$\arg\min_x f_0(x)$$

$$s.t.\, f_i(x) \le 0, i = \{1,\dots,k\}$$

$$f_j(x) = 0, i = \{1,\dots,l\}$$

- Is equivalent to min-max optimization:

$$\arg\min_x [\sup_{\lambda > 0, \nu} (f_0(x) + \sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{l} \nu_j h_j(x))]$$

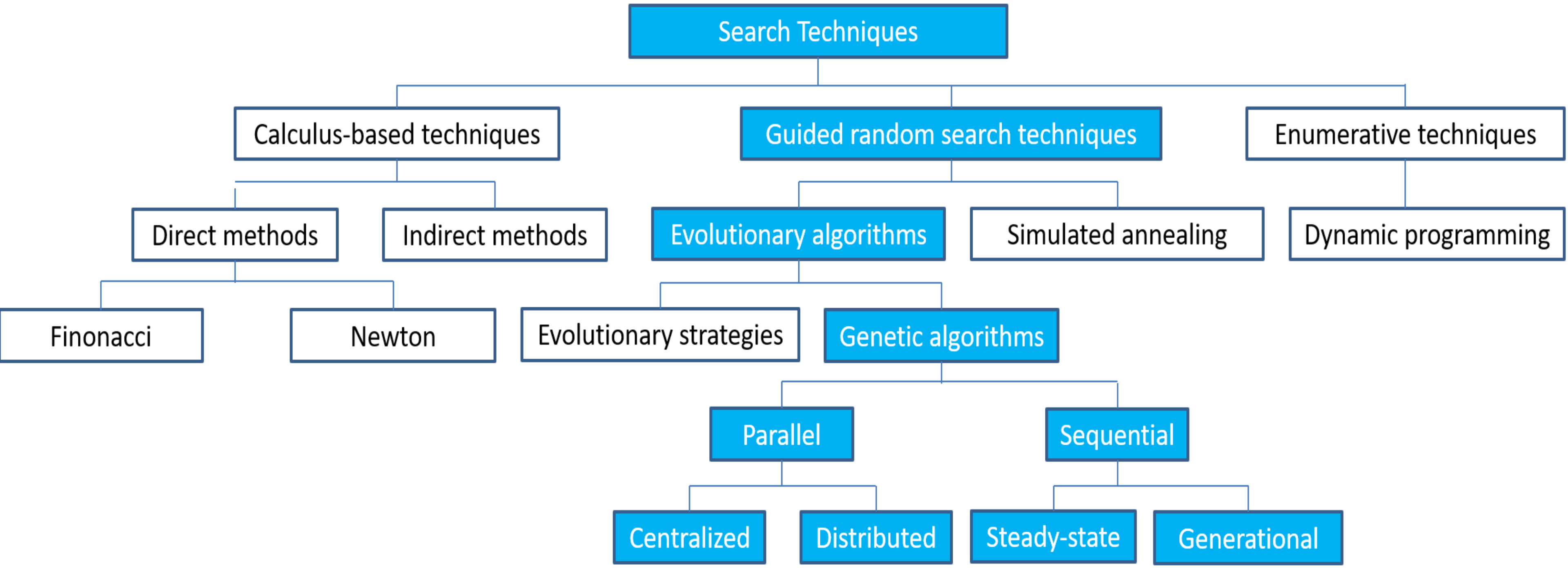
Constrained optimization

- SVM
- Bayesian models:
  - EM
  - Variational methods
  - Graph optimization

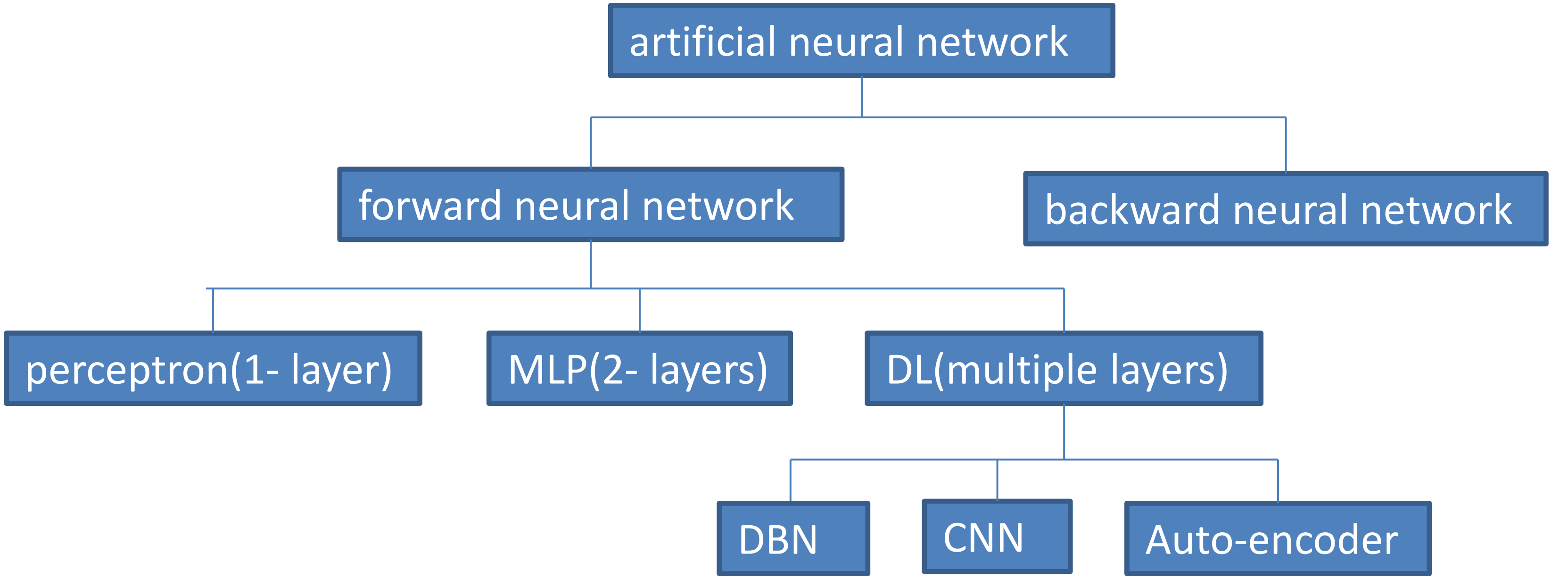
Optimization

- Convex
  - Unconstrained optimization
  - Constrained optimization
- Non-convex
  - Heuristic algorithms
  - Neural networks

Heuristic algorithms



Neural Networks



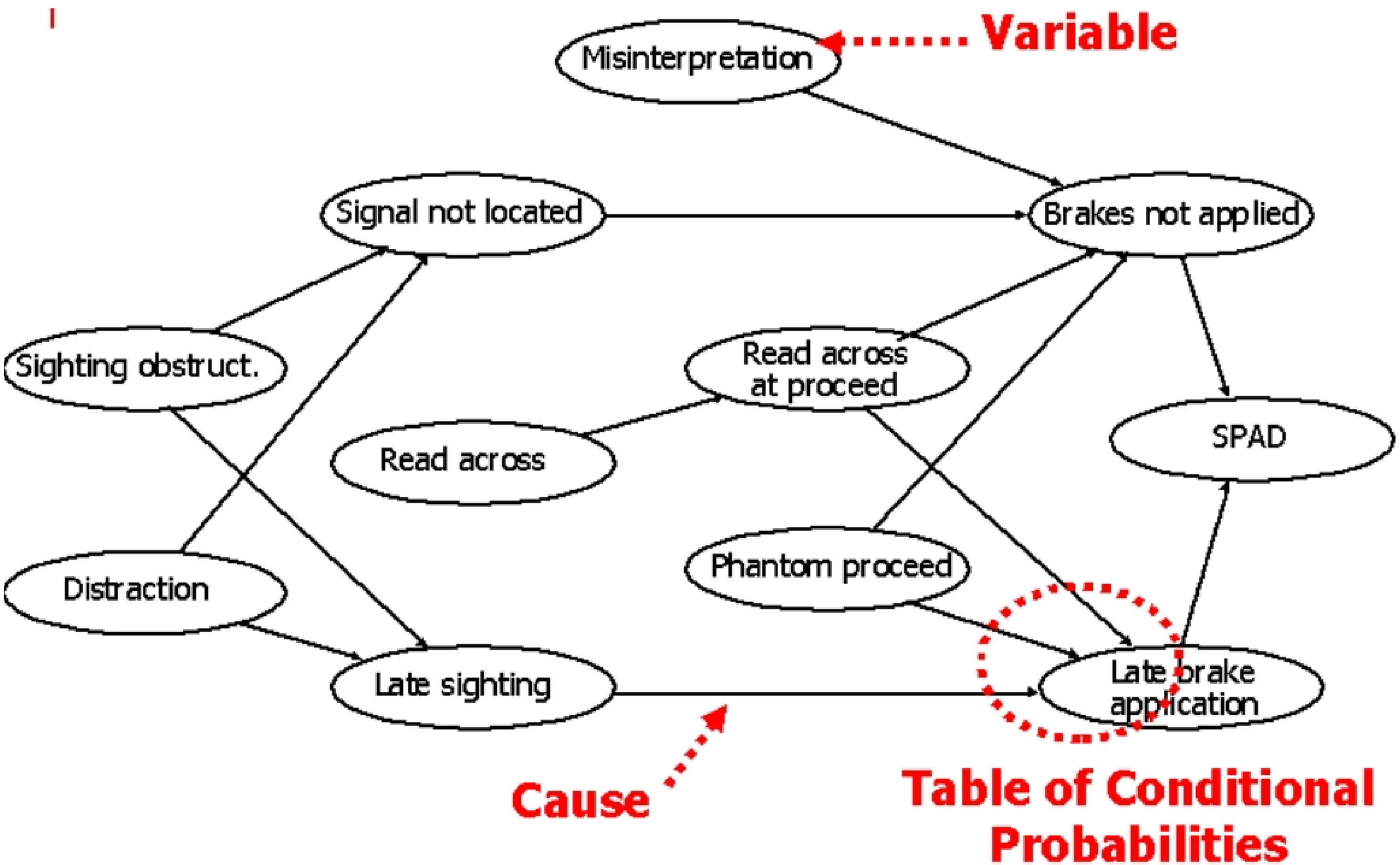
Outlines

- Framework
- Problem Statement
- Related Areas
- History
- Optimization for Machine Learning
- Algorithms
- Examples

Algorithms

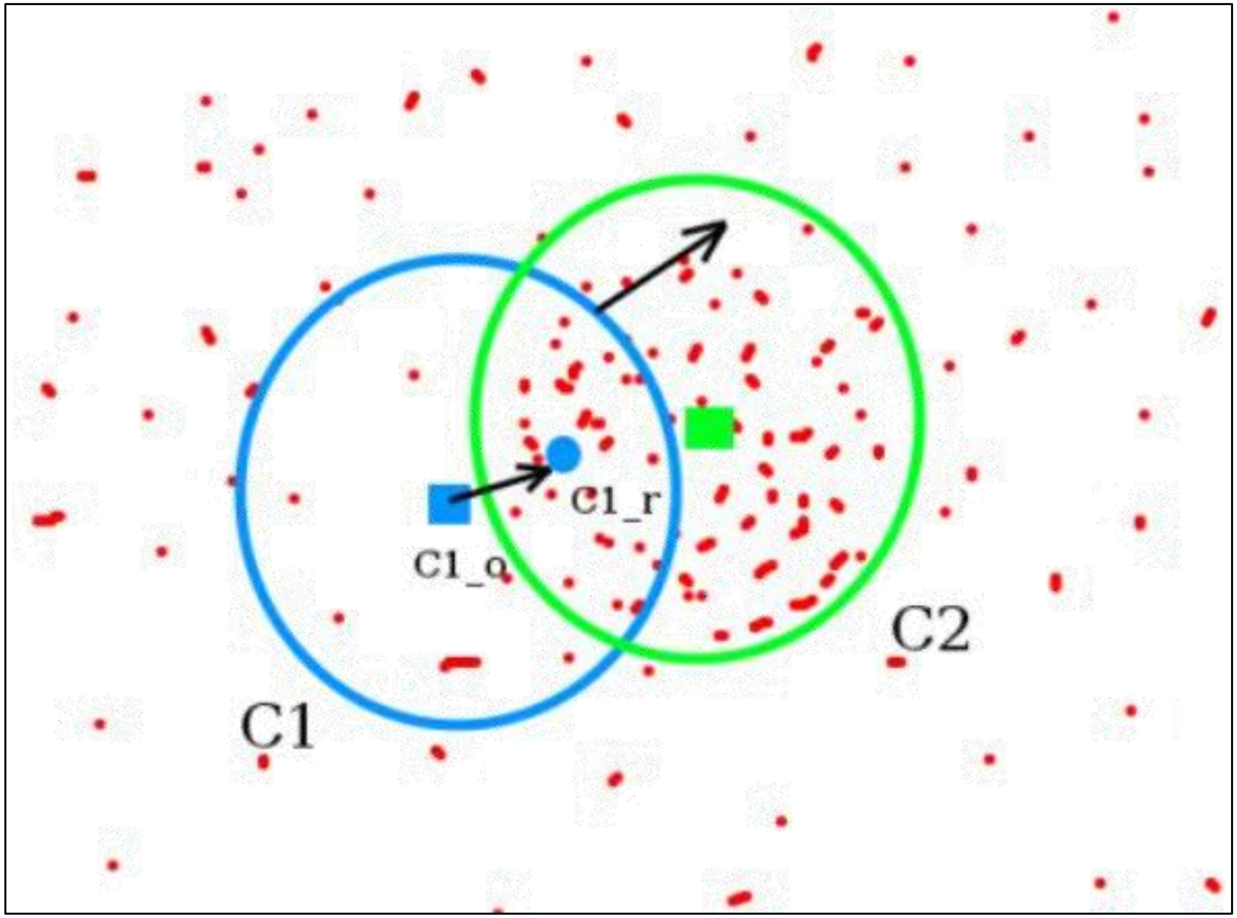
- Bayesian
- K-means
- Decision tree
- Support Vector Machine
- Linear Statistical Learning (PCA, ICA, NMF)
- Nonlinear Statistical Learning (Manifold learning)
- Deep Neural Network
- Generative Adversarial Networks
- Reinforcement learning

Bayesian


Variable
Misinterpretation
Signal not located
Brakes not applied
Sighting obstruct.
Read across
at proceed
SPAD
Read across
Phantom proceed
Distraction
Late sighting
Late brake
application
Cause
Table of Conditional
Probabilities

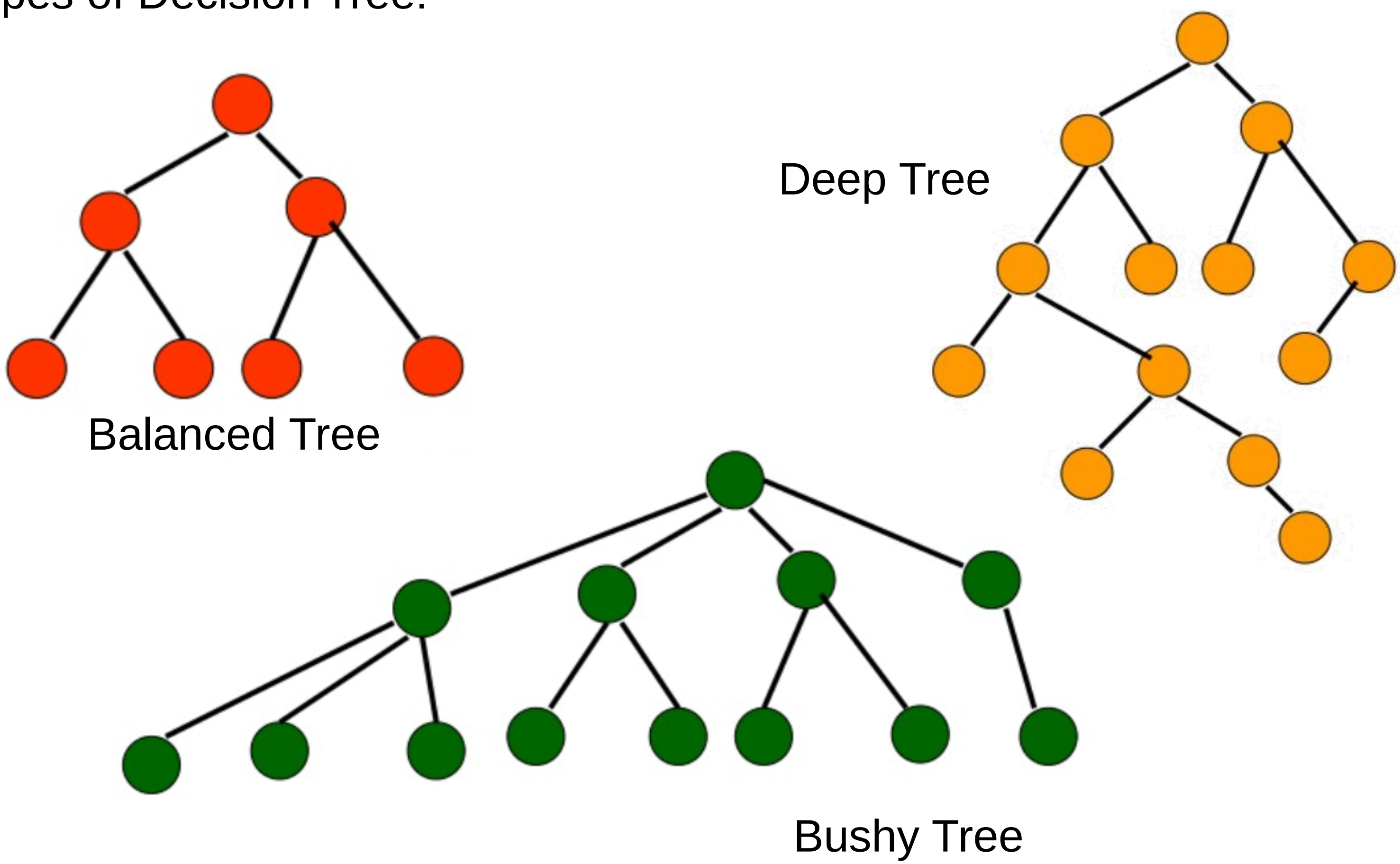
K-Means

- Mean-shift



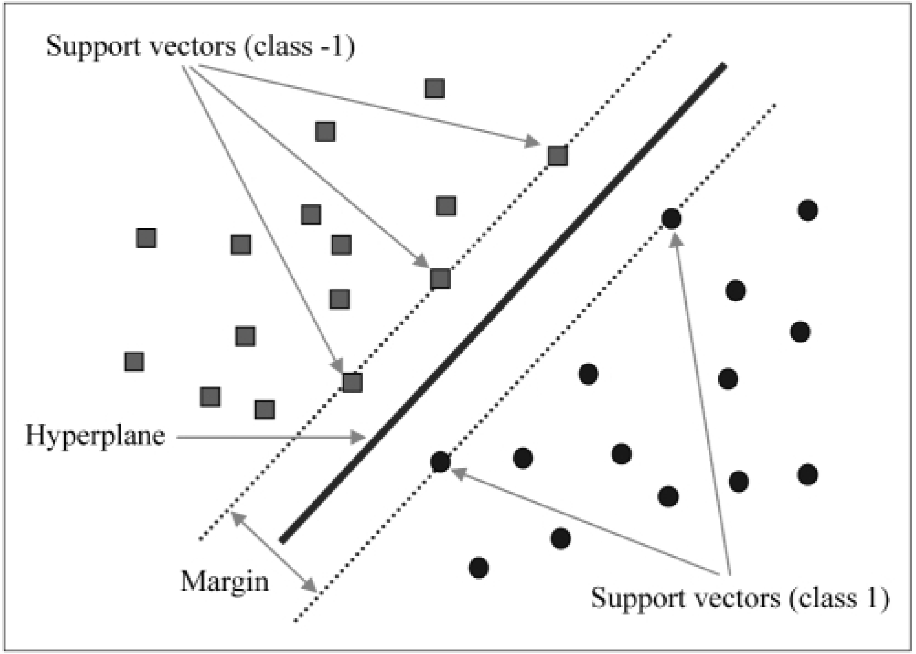
Decision tree

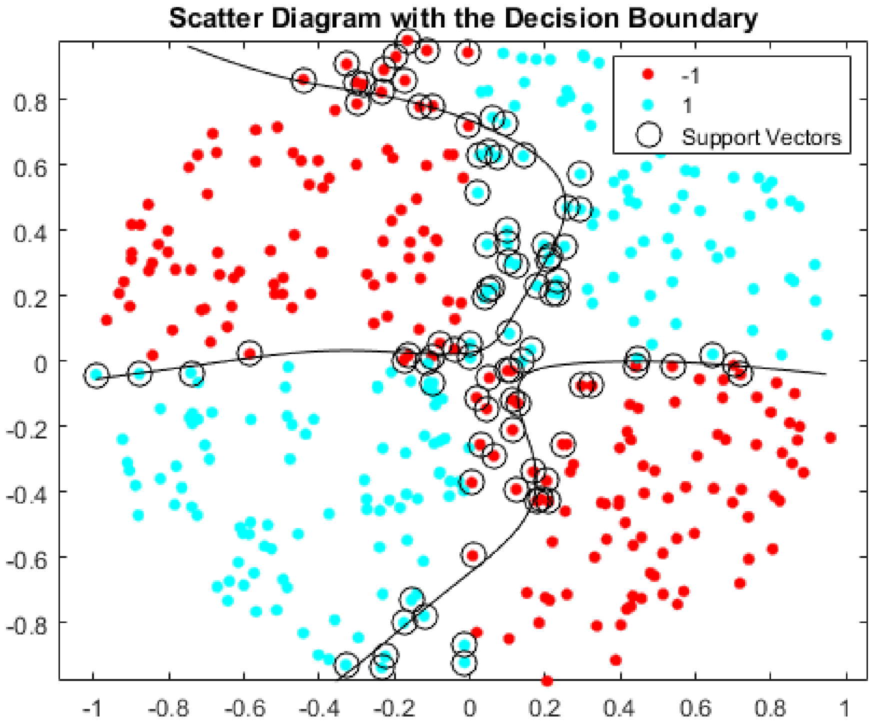
- Types of Decision Tree:



SVM

- eg. Linear SVM: $$\arg\min_{w} \sum_{i=1}^{n} ||w||^2 + C \sum_{i=1}^{n} \xi_i$$
$$\text{s.t. } 1 - y_i x_i^T w \leq \xi_i$$
$$\xi_i \geq 0$$





Linear Statistical Learning

- **PCA**

$$\overset{\text{Data}}{Y} = \overset{\text{Basis}}{A}\,\overset{\text{Coefficients}}{X} \qquad A_i \perp A_j$$

- **ICA**

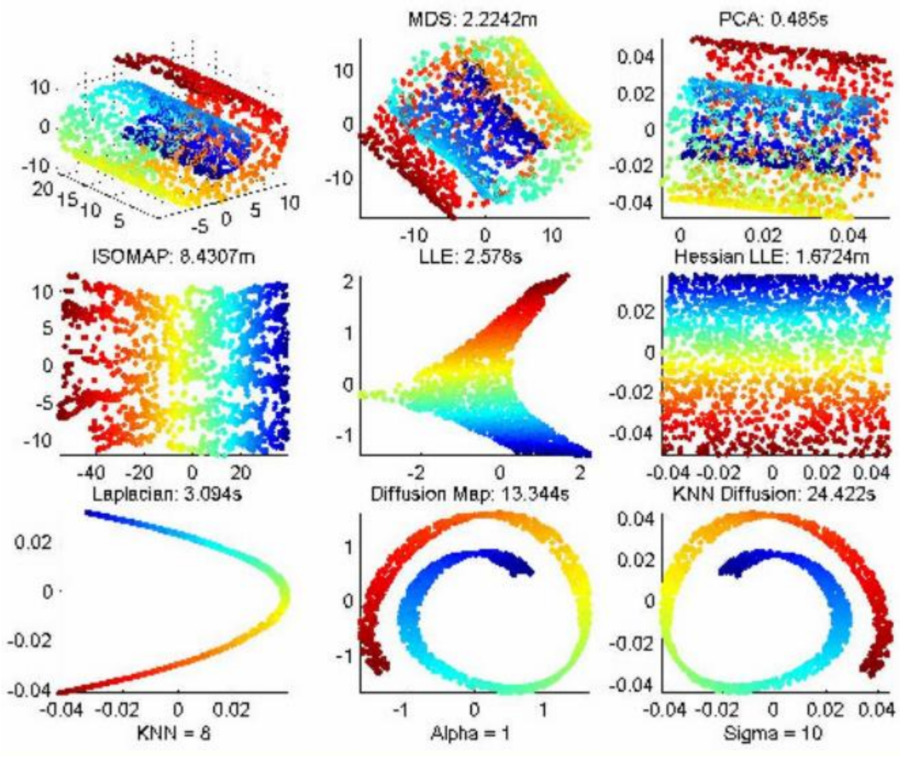
$$\overset{\text{Data}}{Y} = \overset{\text{Mixture Coefficients}}{A}\,\overset{\text{Components}}{X} \qquad \max\; I(X)$$

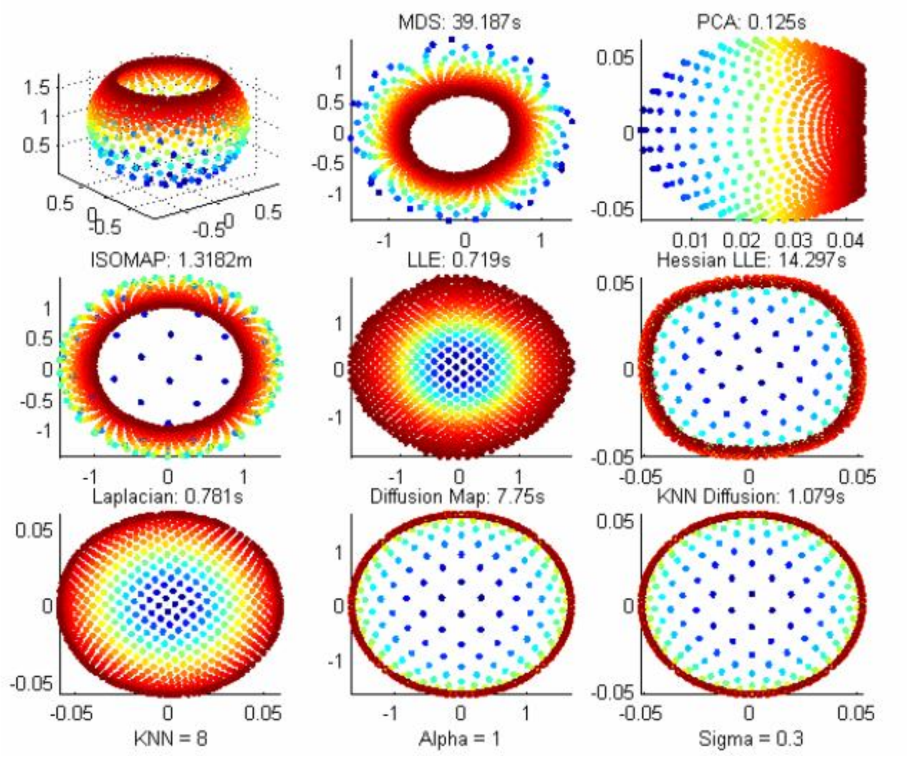
- **NMF**

$$\overset{\text{Data}}{Y} = \overset{\text{Basis}}{A}\,\overset{\text{Coefficients}}{X} \qquad A, X > 0$$

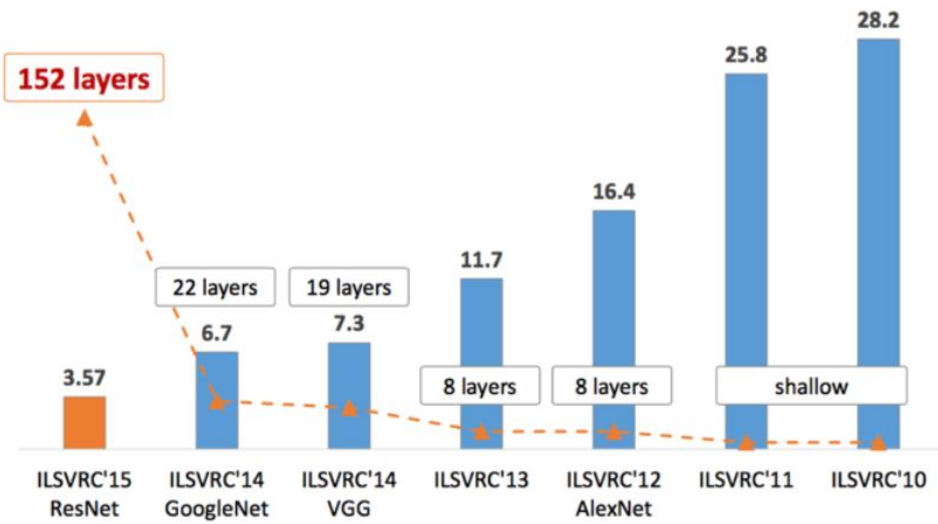
Nonlinear Statistical Learning

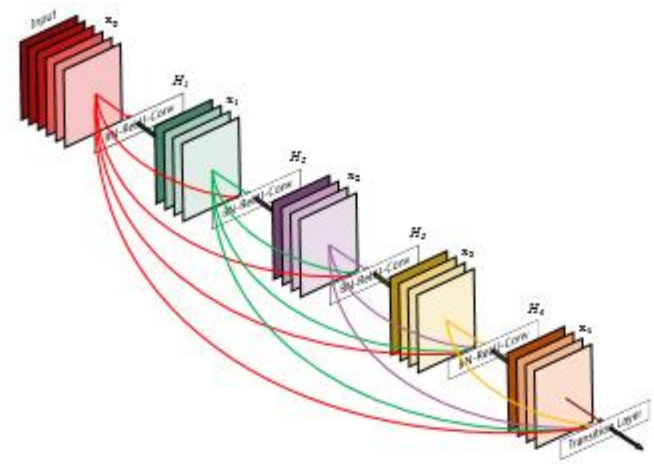
- Manifold learning





DNN

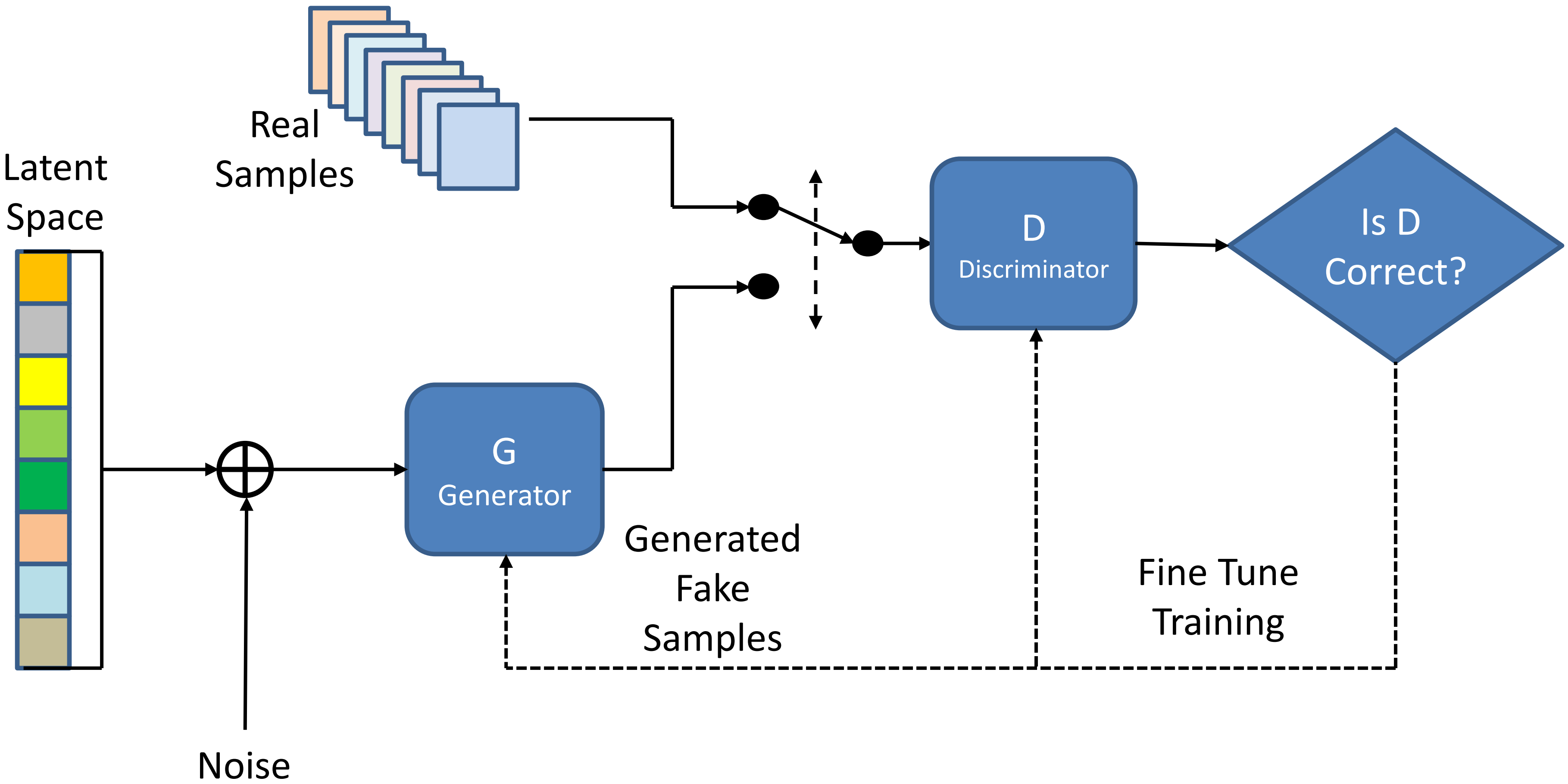




- Task： recognition
- Dataset： ILSVRC

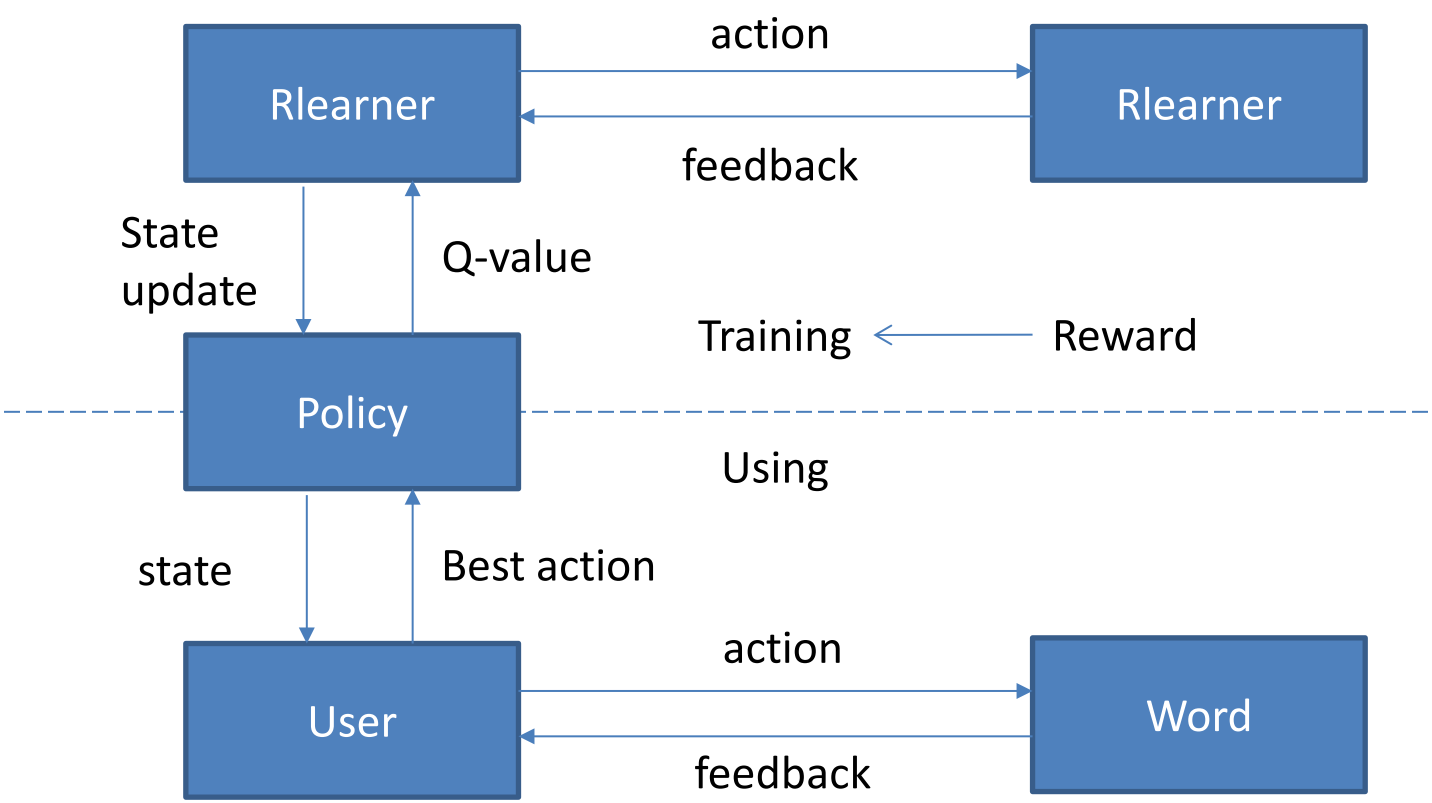
■ Huang G, Liu Z, Weinberger K Q, et al. Densely connected convolutional networks[J]. arXiv preprint arXiv:1608.06993, 2016.

GAN


Real
Samples
Latent
Space
D
Discriminator
Is D
Correct?
G
Generator
Generated
Fake
Samples
Fine Tune
Training
Noise

Reinforcement learning

- Reward, state, action and value



Outlines

- Framework
- Problem Statement
- Related Areas
- History
- Optimization for Machine Learning
- Algorithms
- Examples

Alphago

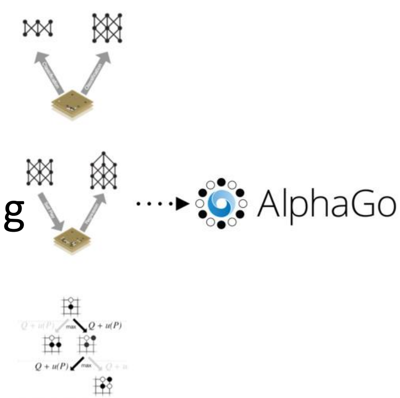
Alphago VS. 李世石

Alphago VS. 柯洁

Supervised learning

Reinforcement learning

Priori knowledge



Better network structure

Enhance the role of Reinforcement learning

Computer Vision-YOLOv2

Future-UAV

■ Dataset

- ➢ Stanford Drone Dataset
- ➢ ASL Dataset
- ➢ EPFL MMSPG drone dataset
- ➢ INRIA Aerial Image Labeling Dataset
- ➢ senseFly samples

Thanks !