Introduction to Machine Learning

Speaker: Harry Chao

Advisor: J.J. Ding

Date: 1/27/2011

Outline

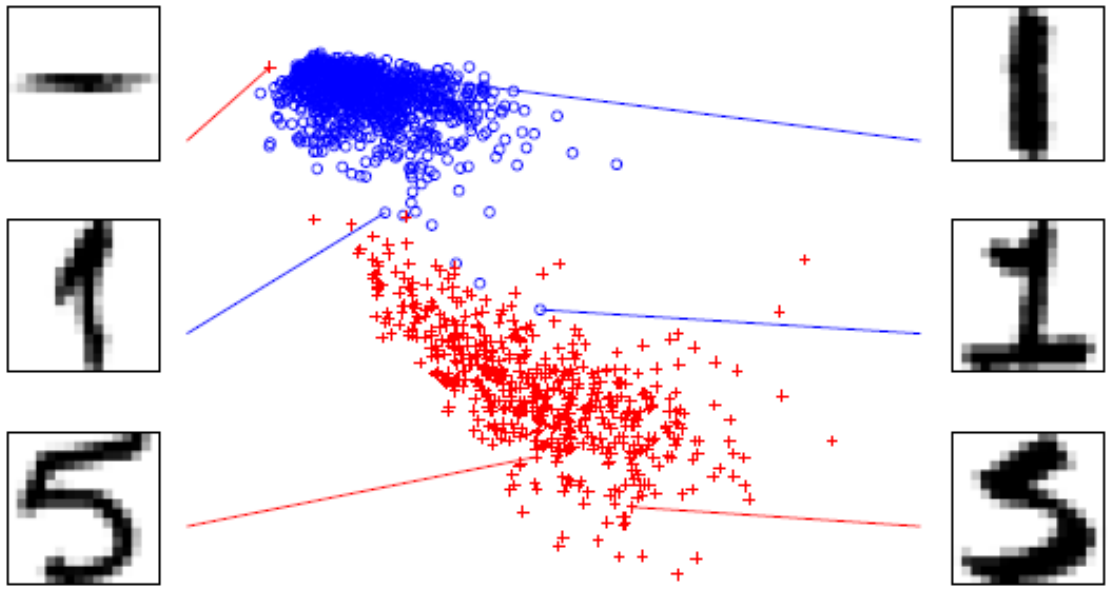
- 1. What is machine learning?
- 2. The basic of machine learning
- 3. Principles and effects of machine learning
- 4. Different machine learning techniques (hypotheses)
  - Linear classifier (numerical functions)
  - Non-parametric (Instance-based functions)
  - Non-metric (Symbolic functions)
  - Parametric (Probabilistic functions)
- 5. Practical usage and application: Pattern recognition
- 6. Conclusion and discussion

1. What is machine learning?

Description:

1. Optimize a performance criterion using example data or past experience
2. Learning and discovering some rules or properties from the given data set

2-dimensional Feature space



1. What is machine learning

Notation:

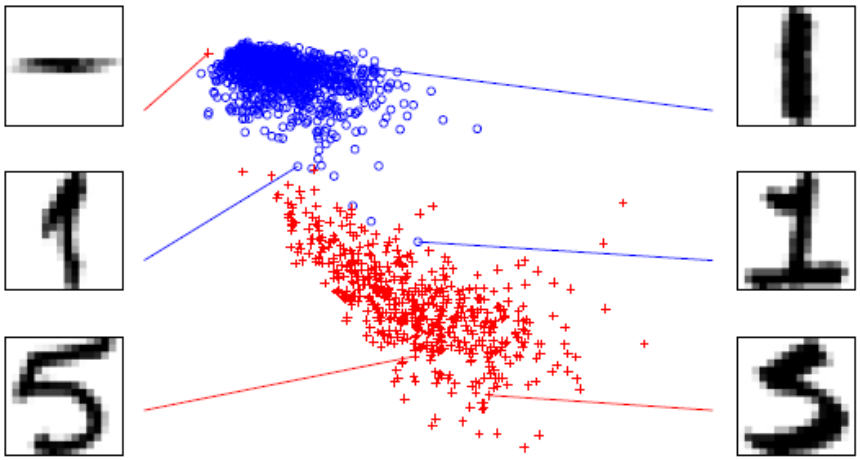
(1) $X = \{x_n \in R^d\}_{n=1}^N$: (training) data set, which contains *N* samples, and each sample is an *d*-dim vector $x_n = [x_{n1}, x_{n2}, \dots\dots, x_{nd}]$

(2) $Y = \{y_n \in R\}_{n=1}^N$: (training) label set

(3) Data pair: Each x_n corresponds to a y_n

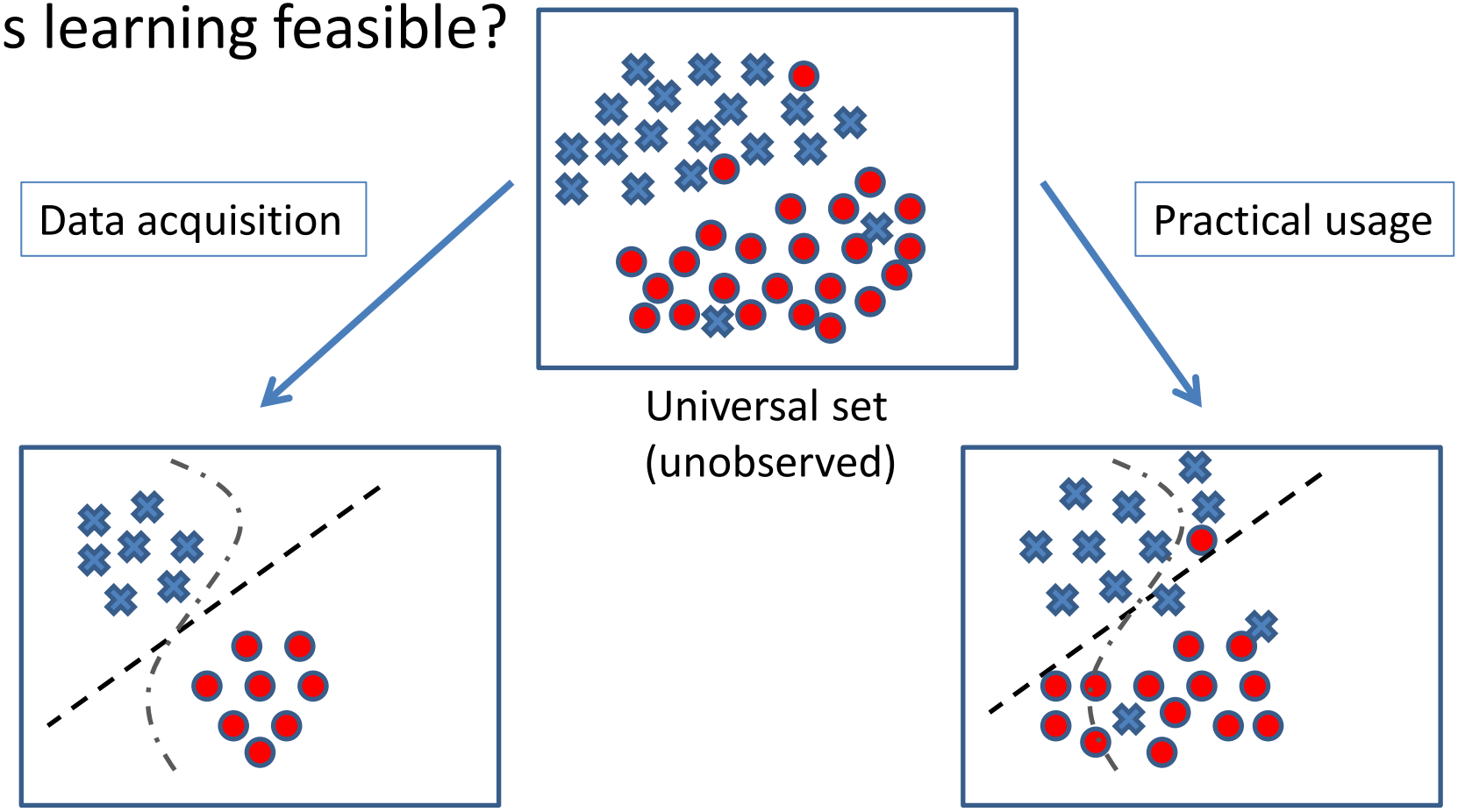
(4) Another form of data set: $\{x_n \in R^d, y_n \in R\}_{n=1}^N$

N = 1000
d = 2



1. What is machine learning?

Is learning feasible?



1. What is machine learning?

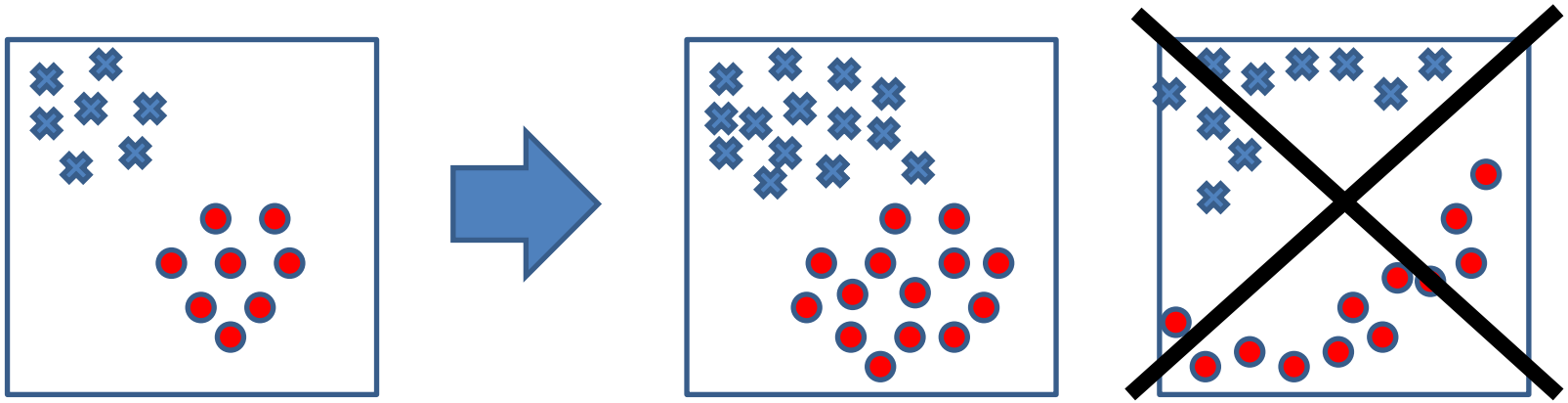
Is learning feasible?

(1) Yes: in the probabilistic way by "Hoeffding Inequality"

$$P[|\nu - \mu| > \epsilon] \leq 2e^{-2\varepsilon^2 N}$$

(2) No free lunch rule:

- Training set and testing set come from the same distribution
- Need to make some assumptions or bias (ex: classifier type, ……)



1. What is machine learning

Disciplines: Computer science and Statistics

- Statistics: Learning and Inference given samples
- Computer science: Efficient algorithms for optimization, and model representation & evaluation

Two important factors:

- Modeling
- Optimization

Other related disciplines:

- Neural networks, Pattern recognition, Information retrieval, Artificial Intelligence, Data mining, Function approximation, …

Outline

- 1. What is machine learning?
- 2. The basic of machine learning
- 3. Principles and effects of machine learning
- 4. Different machine learning techniques (hypotheses)
  - Linear classifier (numerical functions)
  - Non-parametric (Instance-based functions)
  - Non-metric (Symbolic functions)
  - Parametric (Probabilistic functions)
- 5. Practical usage and application: Pattern recognition
- 6. Conclusion and discussion

2. The basic of machine learning

- Design or learning
- Types of machine learning
- What’re we seeking for?
- Machine learning structure (supervised)
- Optimization criterion
- Supervised learning strategies

2. The basic of machine learning

Design or learning:

- Design: design some rules or strategies (If we know the intrinsic factors!)
- Learning: Learning from the (training) data

Knowledge types:

- Domain knowledge
- Data-driven knowledge
- Example: feature generation or data compression

  (DCT, Wavelets VS. KL transform, PCA)

2. The basic of machine learning

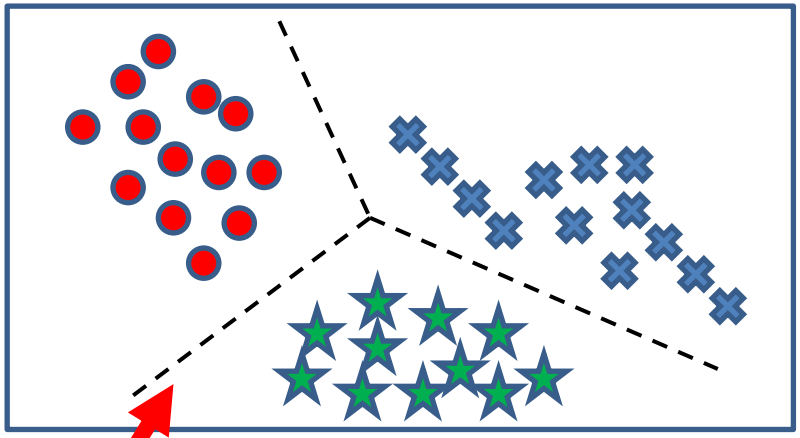
- Design or learning
- Types of machine learning
- What’re we seeking for?
- Machine learning structure (supervised)
- Optimization criterion
- Supervised learning strategies

2. The basic of machine learning

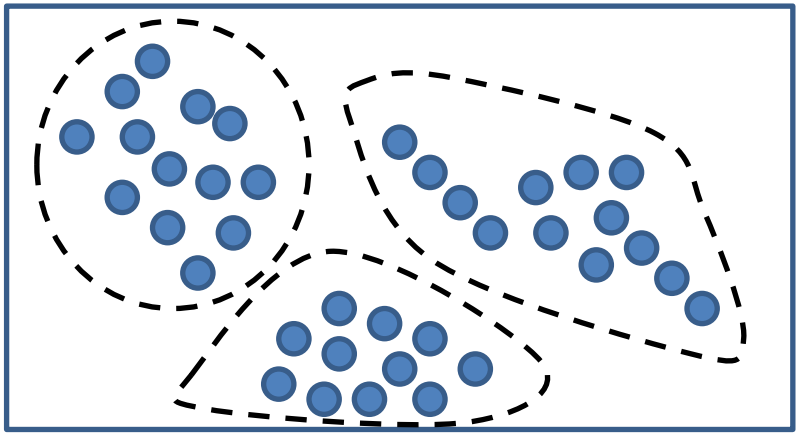
Types of machine learning:

- Supervised learning ($\{x_n \in R^d, y_n \in R\}_{n=1}^N$)
  - (1) Prediction
  - (2) Classification (discrete labels), Regression (real values)
- Unsupervised learning ($\{x_n \in R^d\}_{n=1}^N$)
  - (1) Clustering
  - (2) Probability distribution estimation
  - (3) Finding association (in features)
  - (4) Dimension reduction
- Reinforcement learning
  - (1) Decision making (robot, chess machine)

2. The basic of machine learning

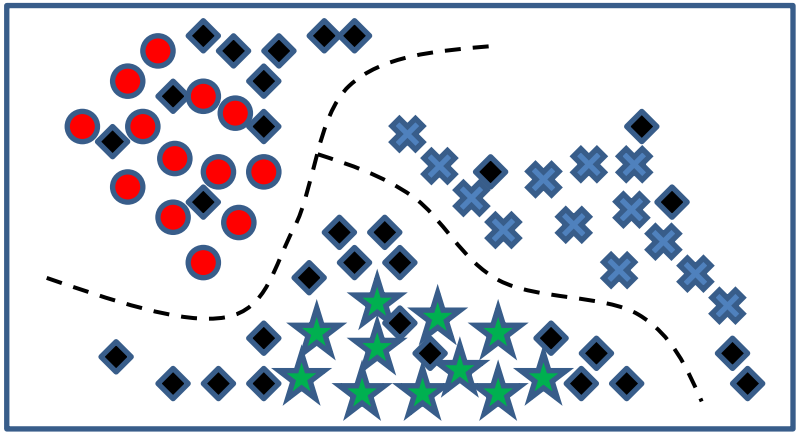


Supervised learning



Unsupervised learning

What we learned



Semi-supervised learning

2. The basic of machine learning

- Design or learning
- Types of machine learning
- What’re we seeking for?
- Machine learning structure (supervised)
- Optimization criterion
- Supervised learning strategies

2. The basic of machine learning

What're we seeking?

(1) Supervised: Low E-out or maximize probabilistic terms

$$error = \frac{1}{N}\sum_{n=1}^{N}[y_n \neq g(x_n)]$$

E-in: for training set

E-out: for testing set

- Hoeffding inequality and VC-bound tell us that minimizing E-in has some correspondences to reach low E-out

With probability δ: $Eout(g) \leq Ein(g) \pm O\left(\sqrt{\frac{d_{VC}}{N}\ln N}\right)$

About Model complexity

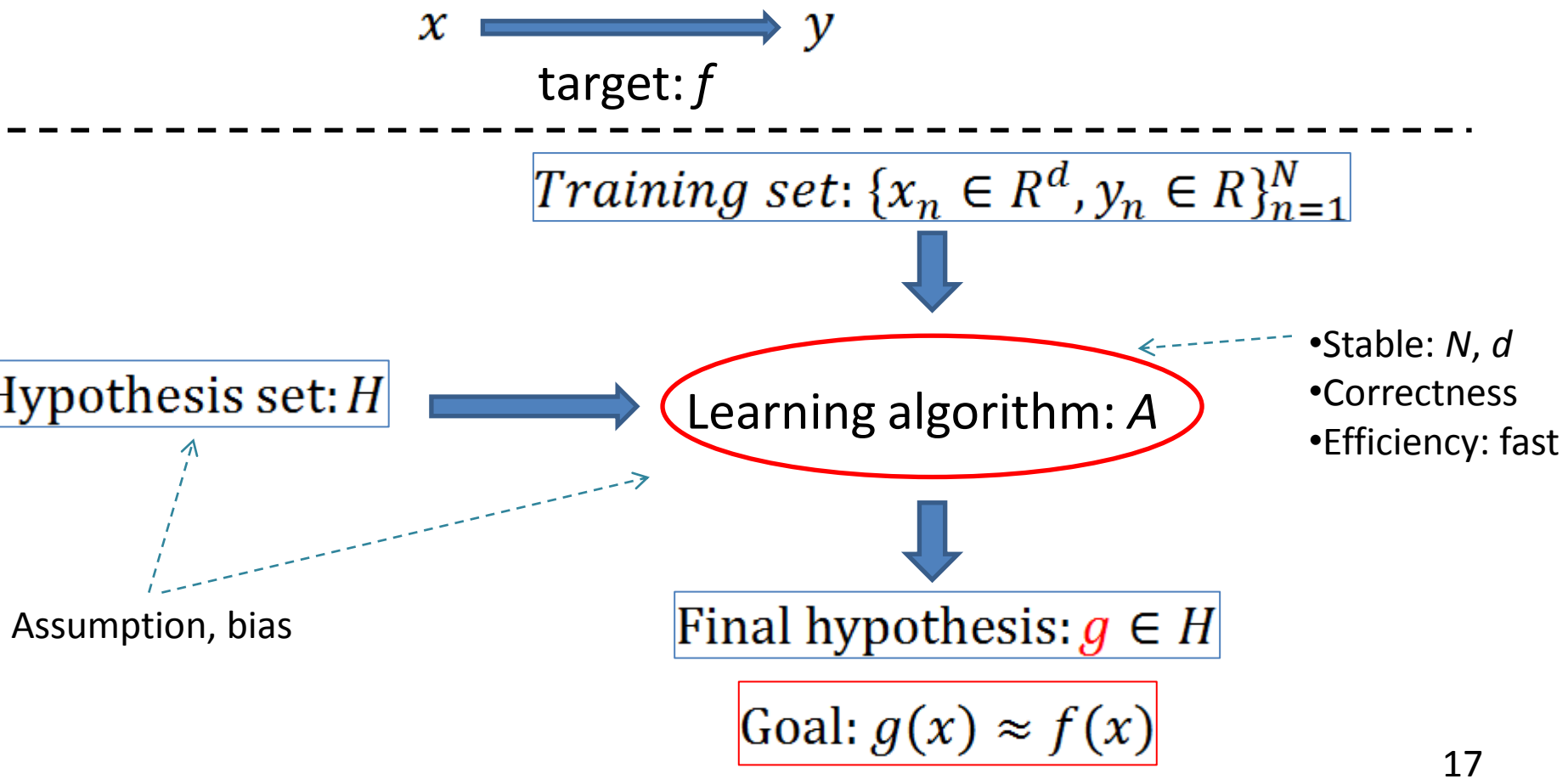
(2) Unsupervised: Minimum quantization error, Minimum distance, MAP, MLE(maximum likelihood estimation)

2. The basic of machine learning

- Design or learning
- Types of machine learning
- What’re we seeking for?
- Machine learning structure (supervised)
- Optimization criterion
- Supervised learning strategies

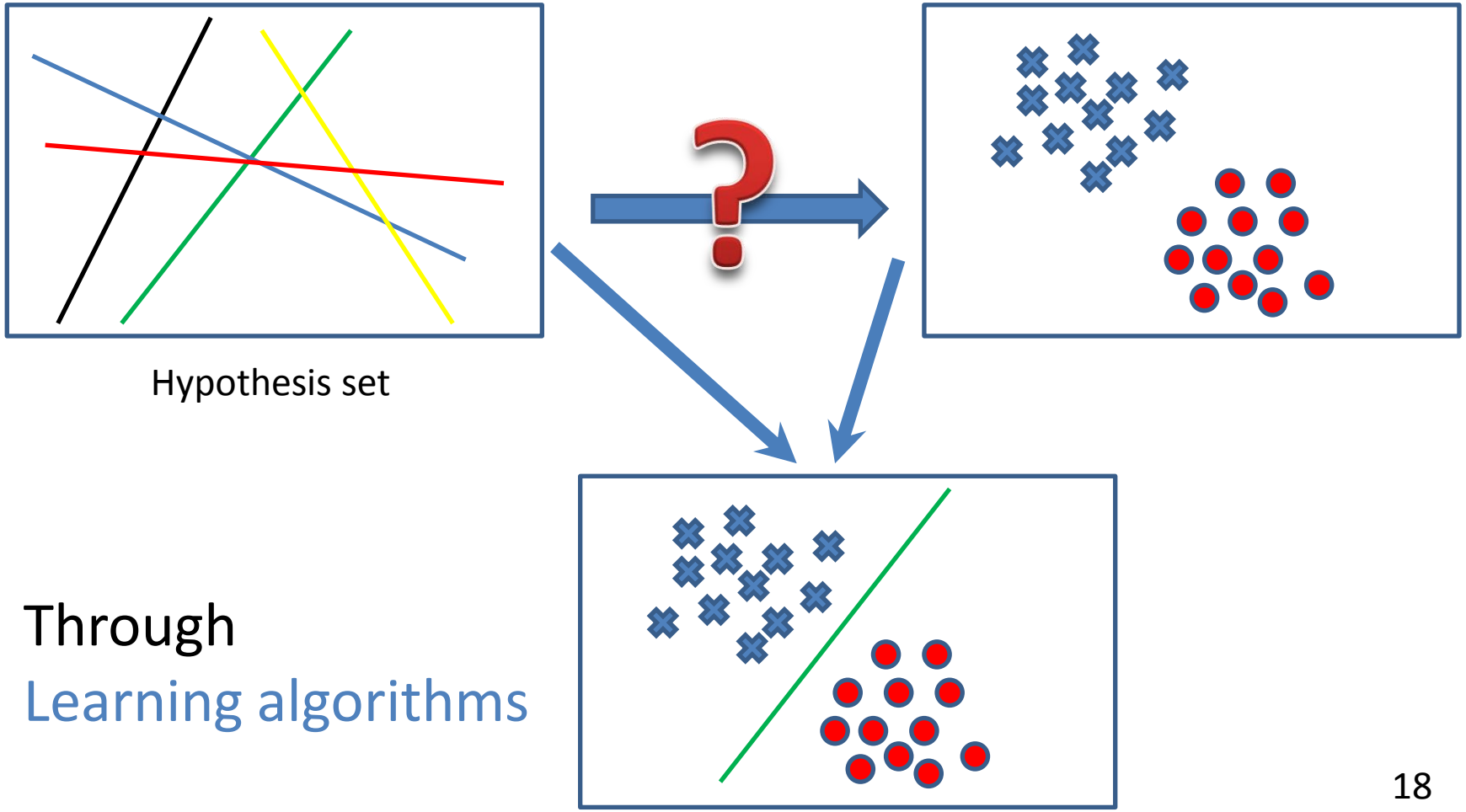
2. The basic of machine learning

Machine learning structure: (Supervised)



2. The basic of machine learning

Hypothesis set: (Linear classifier)



Hypothesis set

Through

Learning algorithms

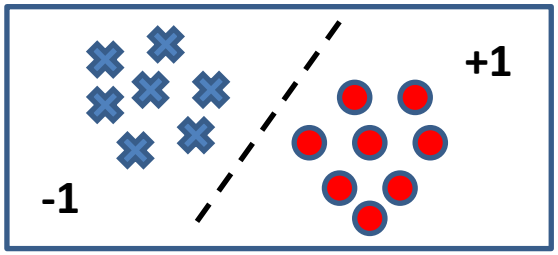
2. The basic of machine learning

- Design or learning
- Types of machine learning
- What’re we seeking for?
- Machine learning structure (supervised)
- Optimization criterion
- Supervised learning strategies

2. The basic of machine learning

Optimization criterion (for supervised linear classifier)

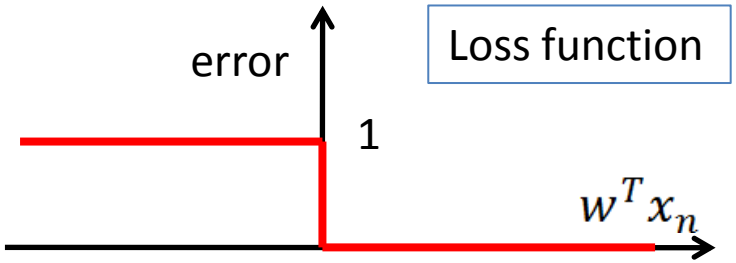
Assume $g(x_n) = sign(w^T x_n)$, where *w* is an *d*-dim vector (learned)



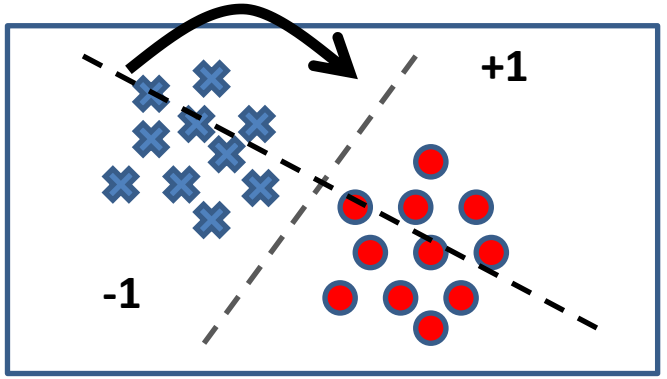
$$error = \frac{1}{N}\sum_{n=1}^{N}[y_n \neq g(x_n)]$$

How to optimize with non-continuous error functions?

For a sample with label 1:



Loss function



2. The basic of machine learning

Optimization is a big issue in machine learning:

- Partial differential (Ex: convex optimization……)
- Genetic algorithm, Neural evolution……

Make differential available:

- Non-continuous → (Piecewise) differentiable

Approximation: (loss functions)

$$Eapp(g) \geq Ein(g)$$

$$\min_{H} Eapp(h) \rightarrow \min_{H} Ein(h)$$

VC bound

$$\text{low } Eout(g)$$

Hypothesis + Loss function

$$Eout(g) \leq Ein(g) \pm O\left(\sqrt{\frac{d_{VC}}{N} \ln N}\right)$$

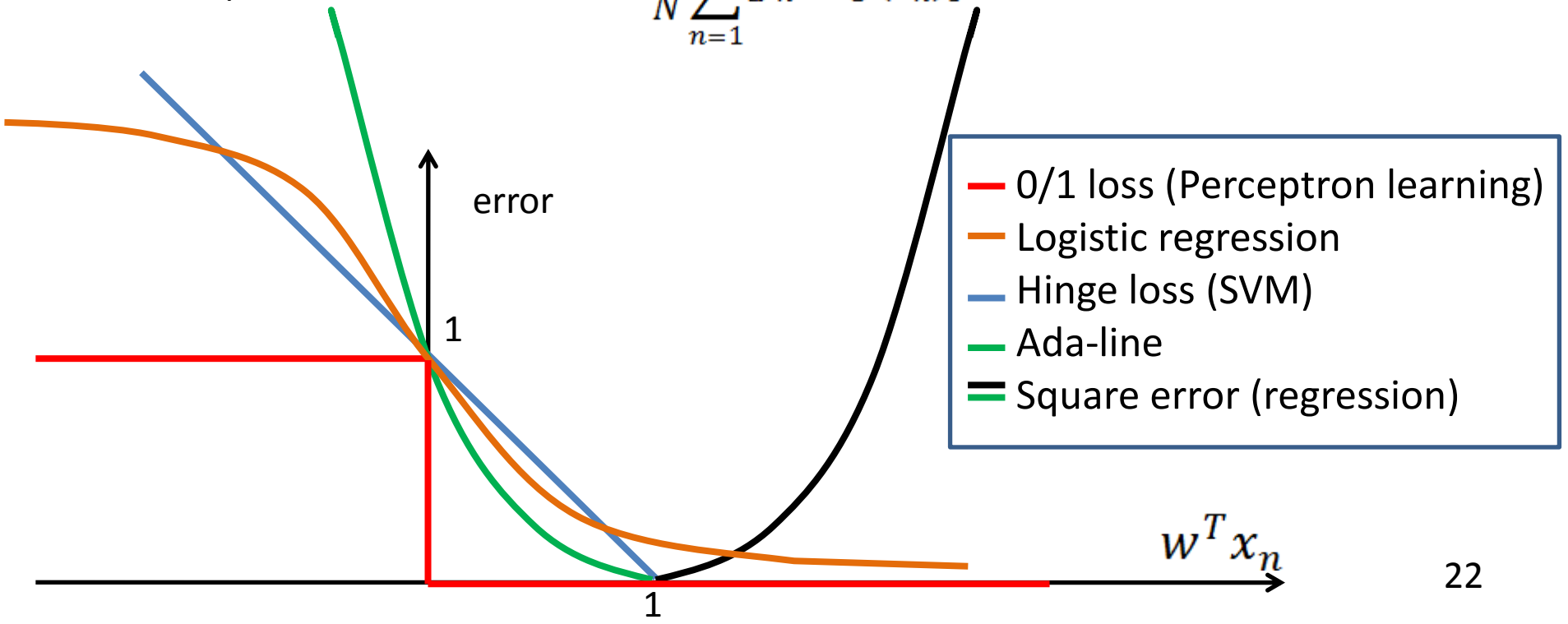
VC bound or generalization error

2. The basic of machine learning

Approximation: Loss (objective) functions

- Different loss functions may find different final hypothesis

For a sample with label 1: $error = \frac{1}{N}\sum_{n=1}^{N}[y_n \neq g(x_n)]$



2. The basic of machine learning

Optimization criterion:

(1) Hypothesis type: linear classifier, decision tree,……

(2) Loss (objective) functions: Hinge loss, square error,……

(3) Optimization algorithms: (stable, efficient, correct)

- Gradient descent: back-propagation, general learning
- Convex optimization: primal & dual problems
- Dynamic programming: HMM
- Divide and Conquer: Decision tree induction,……

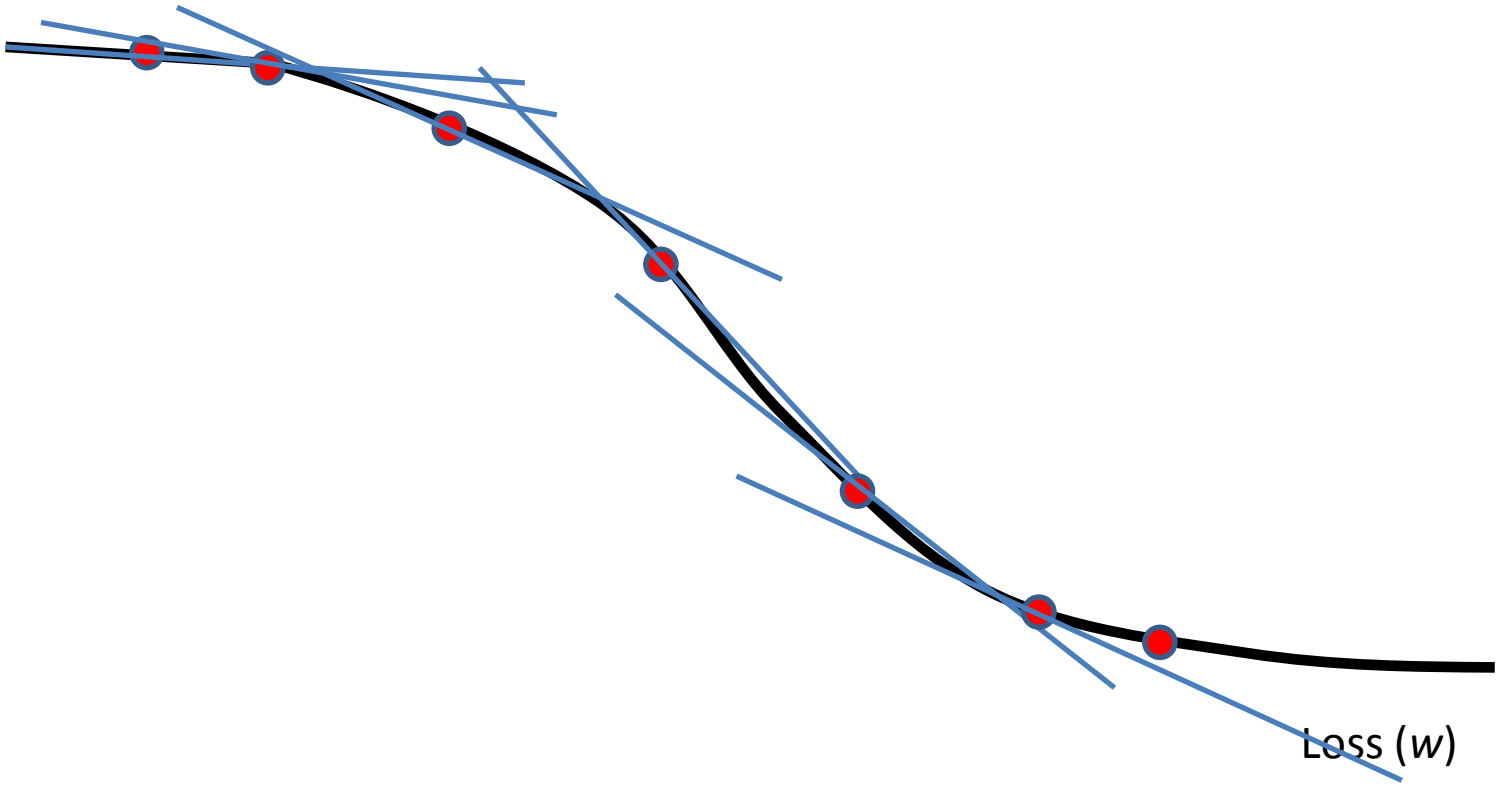
** Hypothesis type affects model complexity

** Loss functions also affects model complexity & final hypothesis

** Algorithms find the desired hypothesis we want

2. The basic of machine learning

Gradient descent: (searching for *w*)



2. The basic of machine learning

- Design or learning
- Types of machine learning
- What’re we seeking for?
- Machine learning structure (supervised)
- Optimization criterion
- Supervised learning strategies

2. The basic of machine learning

Supervised learning strategies: assume $y \in \{1,2,\dots\dots,K\}$

(1) One-shot (Discriminant model)

$$y = f(x)$$

(2) Two-stage (Probabilistic model)

- Discriminative: $P(y = k|x)$
- Generative: $$P(y = k|x) = \frac{P(x|y = k)P(y = k)}{P(x)}$$

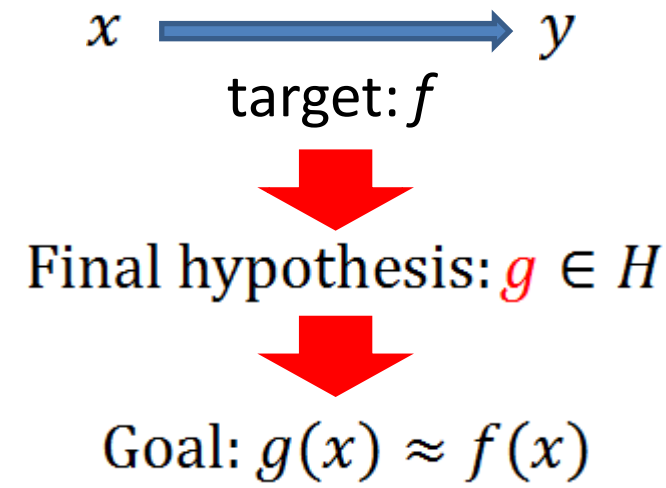
2. The basic of machine learning

What're we seeking? (revised)

(1) One-shot (Discriminant)

- Low E-out by E-in or E-app

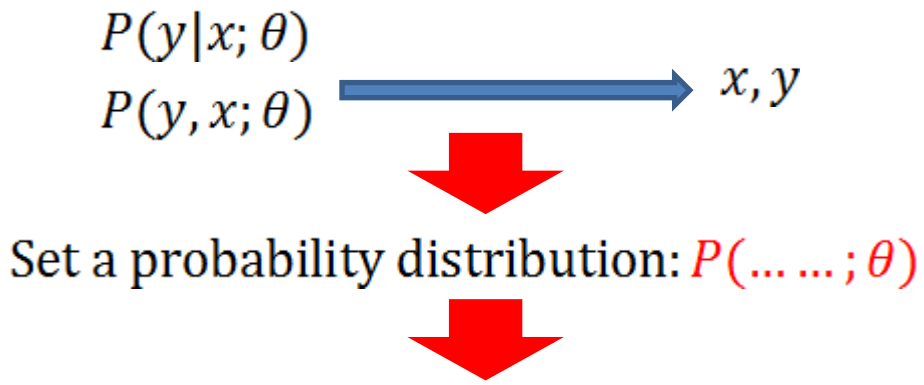
$$error = \frac{1}{N}\sum_{n=1}^{N}[y_n \neq g(x_n)]$$



---

(2) Two-stage (Probabilistic model)

- Discriminative: $\max_{\theta} P(Y|X;\theta)$
- Generative: $\max_{\theta} P(Y,X;\theta)$



2. The basic of machine learning

Supervised learning strategies: Comparison

Category:	One-shot	Two-stage
Model:	Discriminant	Generative, Discriminative
Adv:	•Fewer assumptions •Direct towards the goal	•More flexible •Uncertainty & Domain-knowledge •Discovery and analysis power
Dis-Adv:	•No probabilistic signal	•More assumptions •Computational complexity
Usage:	Supervised	Supervised & Unsupervised
Techniques:	SVM, MLP, KNN, LDA, adaboost, CART, random forest	GMM, GDA, Graphical models, HMM, Naïve Bayes

2. The basic of machine learning

Machine learning revised:

- Theory: $$Eout(g) \leq Ein(g) \pm O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right)$$
- Goal: $$g = \underset{H}{argmin}\, Eout(h)$$
- Methods:

$$\min_{H} Eapp(h) \rightarrow \min_{H} Ein(h) \xrightarrow{\text{VC bound}} \text{low } Eout(g)$$

Outline

- 1. What is machine learning?
- 2. The basic of machine learning
- 3. Principles and effects of machine learning
- 4. Different machine learning techniques(hypotheses)
  - Linear classifier (numerical functions)
  - Non-parametric (Instance-based functions)
  - Non-metric (Symbolic functions)
  - Parametric (Probabilistic functions)
- 5. Practical usage and application: Pattern recognition
- 6. Conclusion and discussion

3. Principles and effects of machine learning

Effect 1: General performance

With probability δ: $$Eout(g) \leq Ein(g) \pm O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right)$$

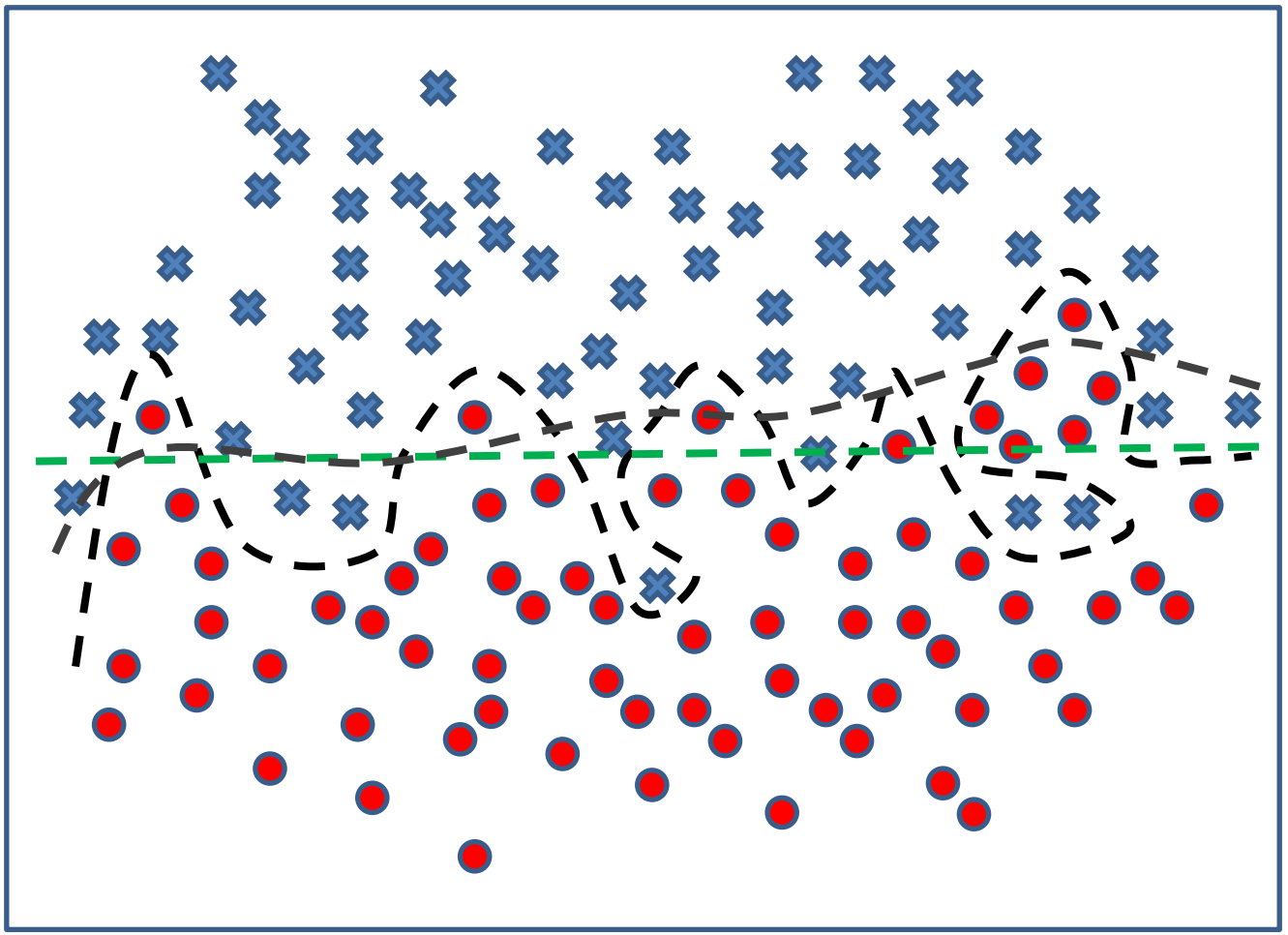
(VC bound, Generalization error)

Factors: (d_{VC} related to model complexity)

- Complex model: $d_{VC} \uparrow, Ein(g) \downarrow$
- Generative error: $d_{VC} \uparrow, O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right) \uparrow$

3. Principles and effects of machine learning

Model complexity:



d_{VC}:

:Low

:Middle

:high

3. Principles and effects of machine learning

Effect 2: *N*, VC dimension, and *d* on generalized error

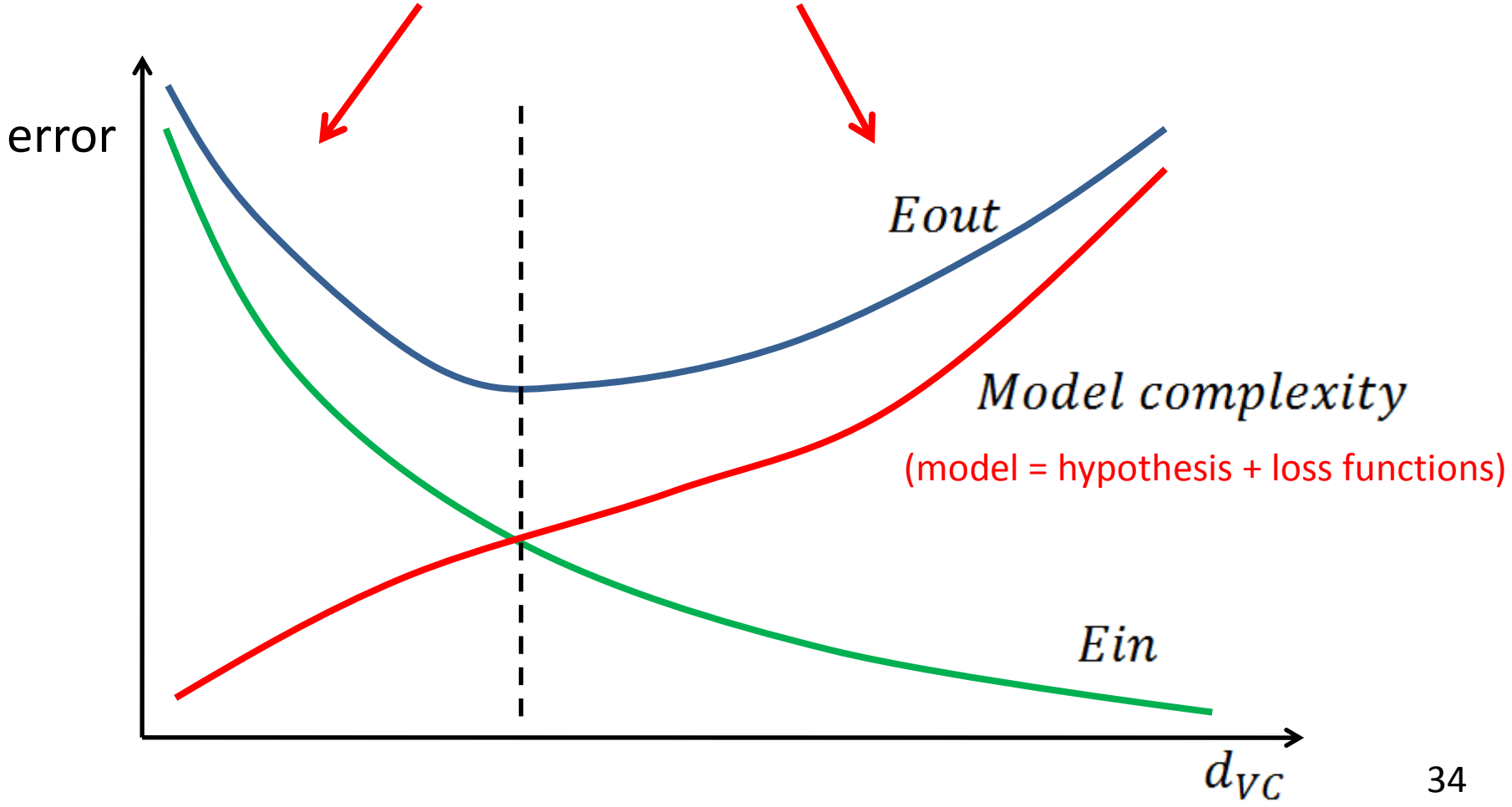
$$d_{VC} \uparrow, O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right) \uparrow$$

$$N \uparrow, O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right) \downarrow$$

$$d \uparrow, d_{VC} \uparrow, O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right) \uparrow$$

3. Principles and effects of machine learning

Effect 3: Under-fitting VS. Over-fitting (fixed *N*)

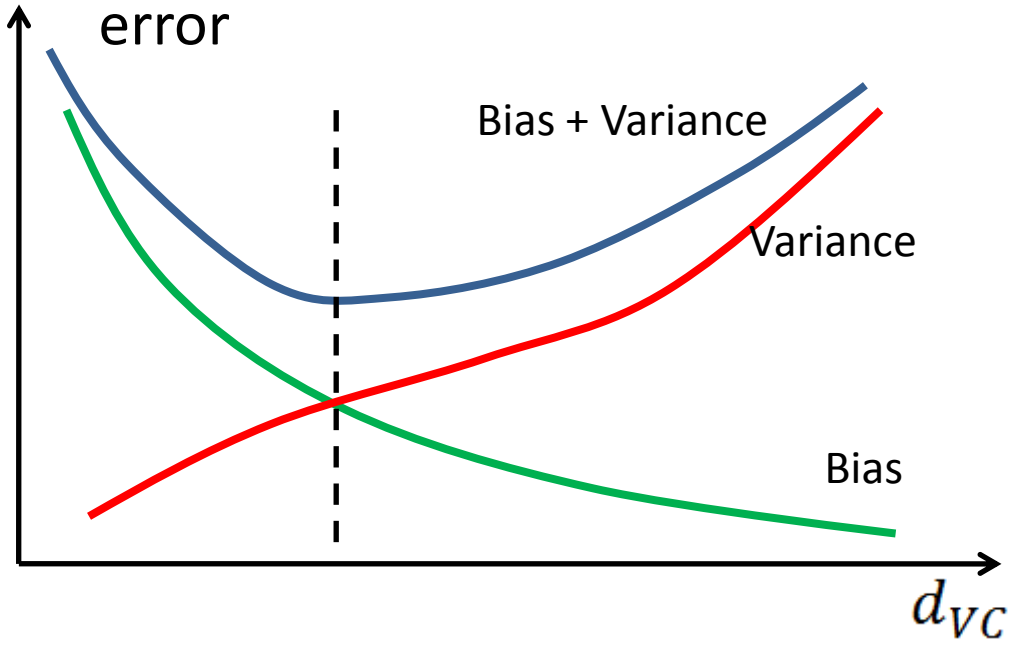


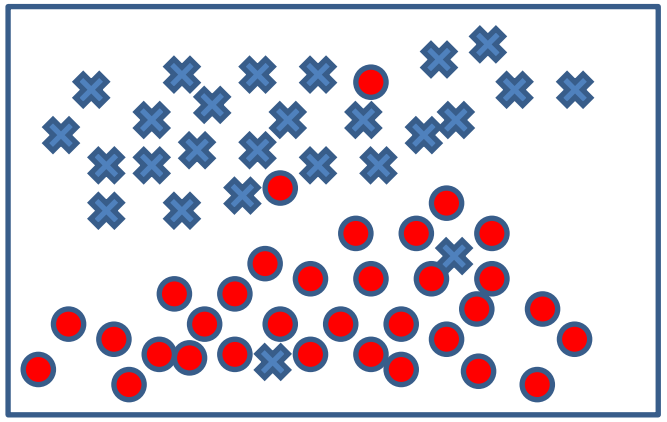
3. Principles and effects of machine learning

Effect 4: Bias VS. Variance (fixed *N*)

(Frequently used in statistics regression)

- Bias: Ability to fit the training data (lower bias the better)
- Variance: Final hypothesis variance with the training set

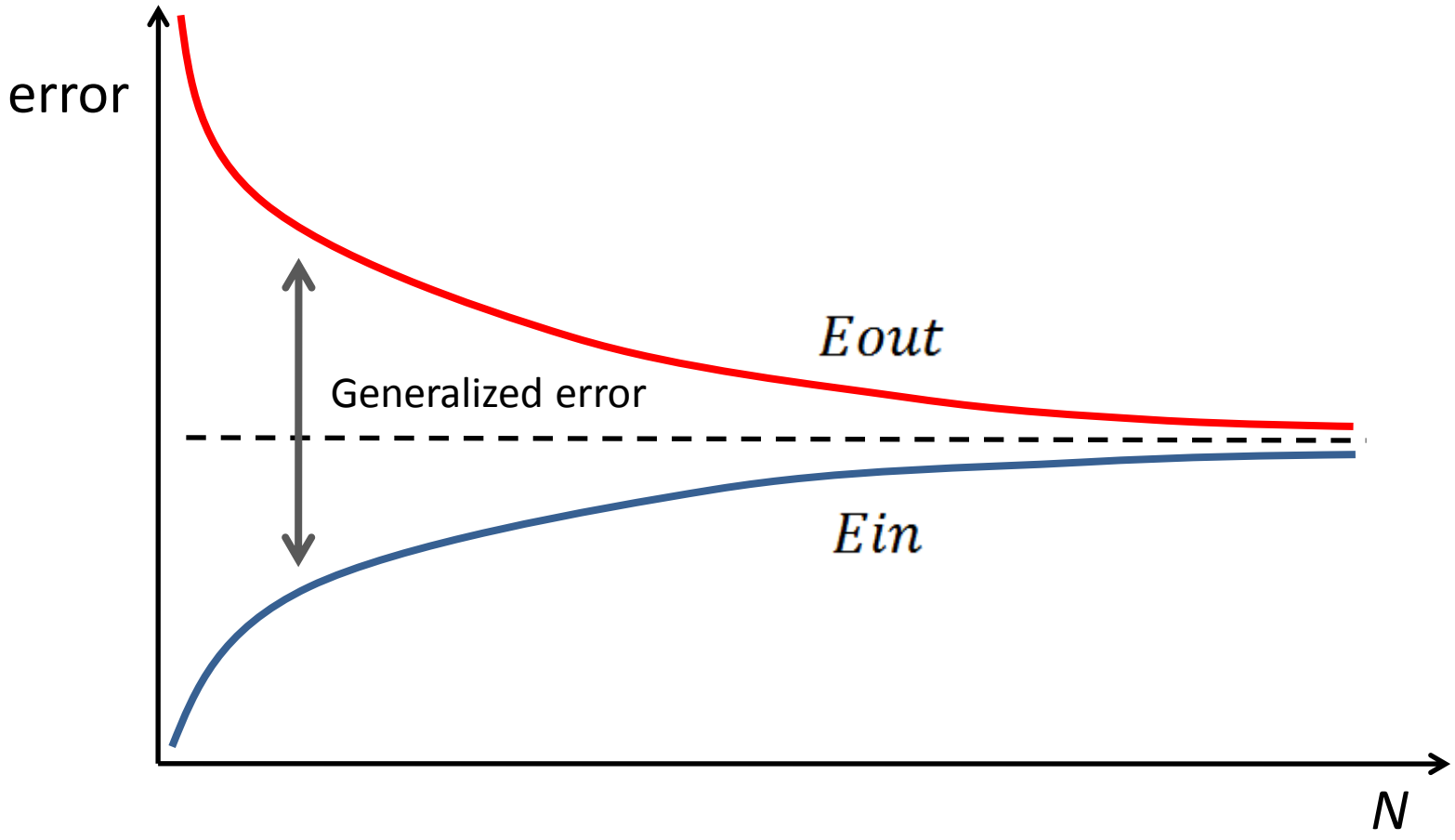




Universal set
(unobserved)

3. Principles and effects of machine learning

Effect 5: Learning curve (fixed d_{VC})



3. Principles and effects of machine learning

Principles:

- Occam's Razor: The simplest model that fits the data is also the most plausible!
- Sampling bias: If the data is sampled in a biased way, learning will produce a similarity biased outcome!
- Data snooping: If a data set has affected any step in the learning process, it cannot be fully trusted in assessing the outcome!

3. Principles and effects of machine learning

Model selection:

- For a machine learning problems, we have many hypothesis, loss functions, and algorithms to try!
- 2 parameters for a hypothesis set:

  (1) Set parameter (manual), (2) Hypothesis parameter (learned)
- We need a method or strategy to find the best model (hypothesis set + loss function) for training and testing!

** Directly using the achieved training error (E-in) of each model for model selection is risky!!

3. Principles and effects of machine learning

Model selection:

$$error = \frac{1}{N}\sum_{n=1}^{N}[y_n \neq g(x_n)]$$

$$\min_H Eapp(h)$$

Regularization:

$$\min_H [Eapp(h) + Complexity(h)]$$

Validation:

Training set→ Base set + Validation set
(1)Train parameters on Base set
(2)Test performance on Validation set

3. Principles and effects of machine learning

Machine learning revised:

- Theory: $$Eout(g) \leq Ein(g) \pm O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right)$$
- Goal: $$g = \underset{H}{argmin}\, Eout(h)$$
- Methods:

$$\min_{H} Eapp(h) \rightarrow \min_{H} Ein(h)$$ VC bound ⟹ low $Eout(g)$

Outline

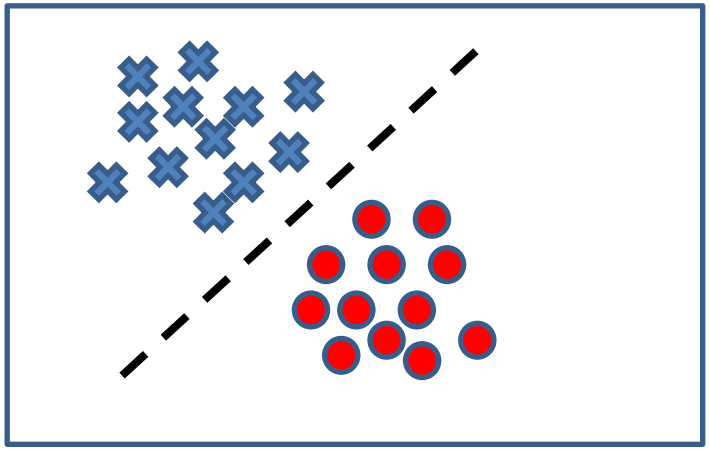
- 1. What is machine learning?
- 2. The basic of machine learning
- 3. Principles and effects of machine learning
- 4. Different machine learning techniques(hypotheses)
  - Linear classifier (numerical functions)
  - Non-parametric (Instance-based functions)
  - Non-metric (Symbolic functions)
  - Parametric (Probabilistic functions)
- 5. Practical usage and application: Pattern recognition
- 6. Conclusion and discussion

4. Different machine learning techniques

- Linear classifier (numerical functions)
- Non-parametric (Instance-based functions)
- Non-metric (Symbolic functions)
- Parametric (Probabilistic functions)
- Aggregation

4. Different machine learning techniques

Linear classifier (numerical functions):



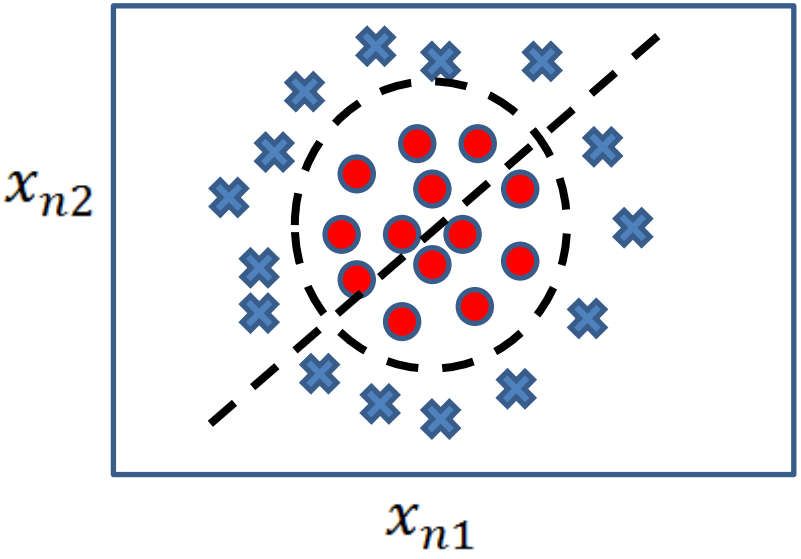
$$g(x_n) = sign(w^T x_n)$$

, where *w* is an *d*-dim vector (learned)

- Techniques: Logistic regression, perceptron, Ada-line, Support vector machine (SVM), Multi-layer perceptron (MLP)
- Linear to nonlinear: Feature transform and Kernel

4. Different machine learning techniques

Feature transform:



$$x_n = [x_{n1}, x_{n2}]$$

$$x_n = [x_{n1}, x_{n2}, x_{n1} * x_{n2}, {x_{n1}}^2, {x_{n2}}^2]$$

$$g(x_n) = sign(w^T x_n)$$

Support vector machine (SVM):

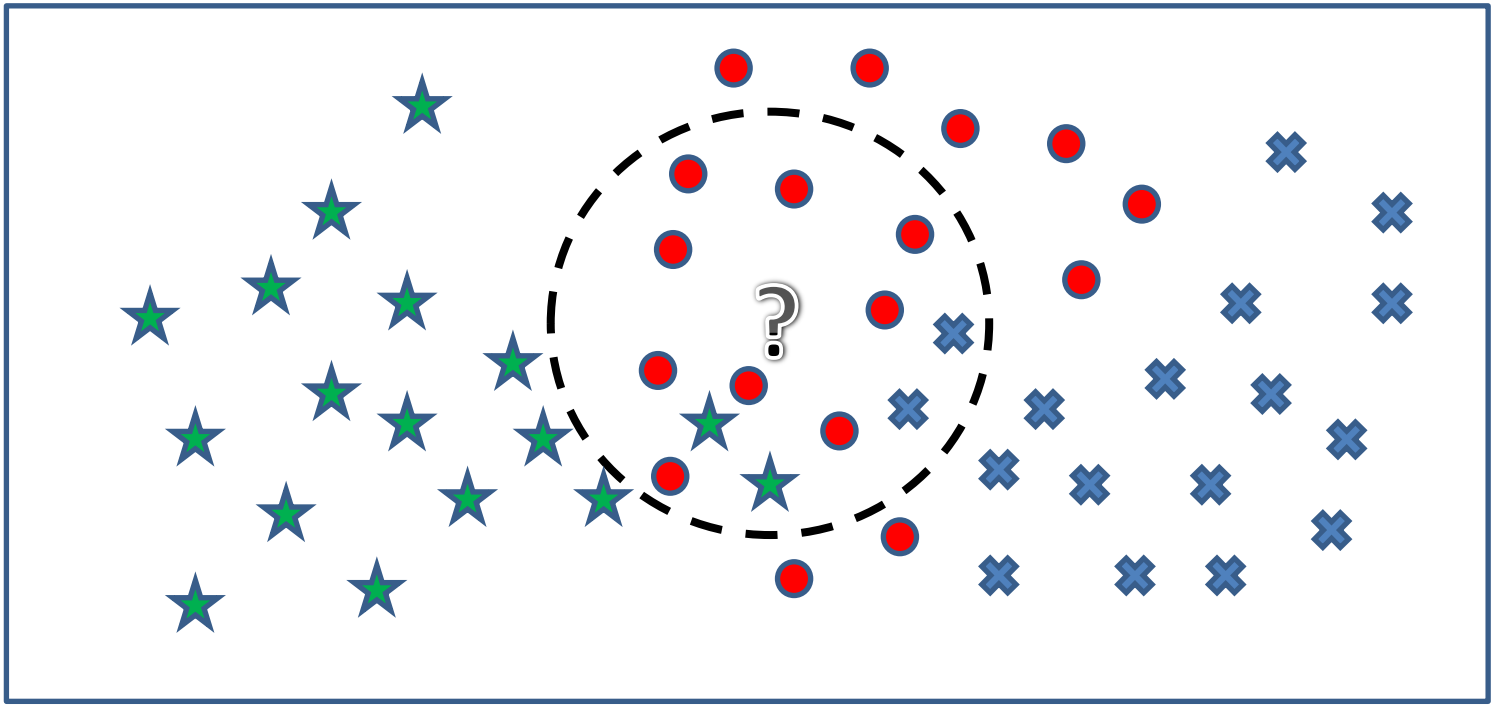
- Useful because it combines regularization with feature transform

4. Different machine learning techniques

- Linear classifier (numerical functions)
- Non-parametric (Instance-based functions)
- Non-metric (Symbolic functions)
- Parametric (Probabilistic functions)
- Aggregation

4. Different machine learning techniques

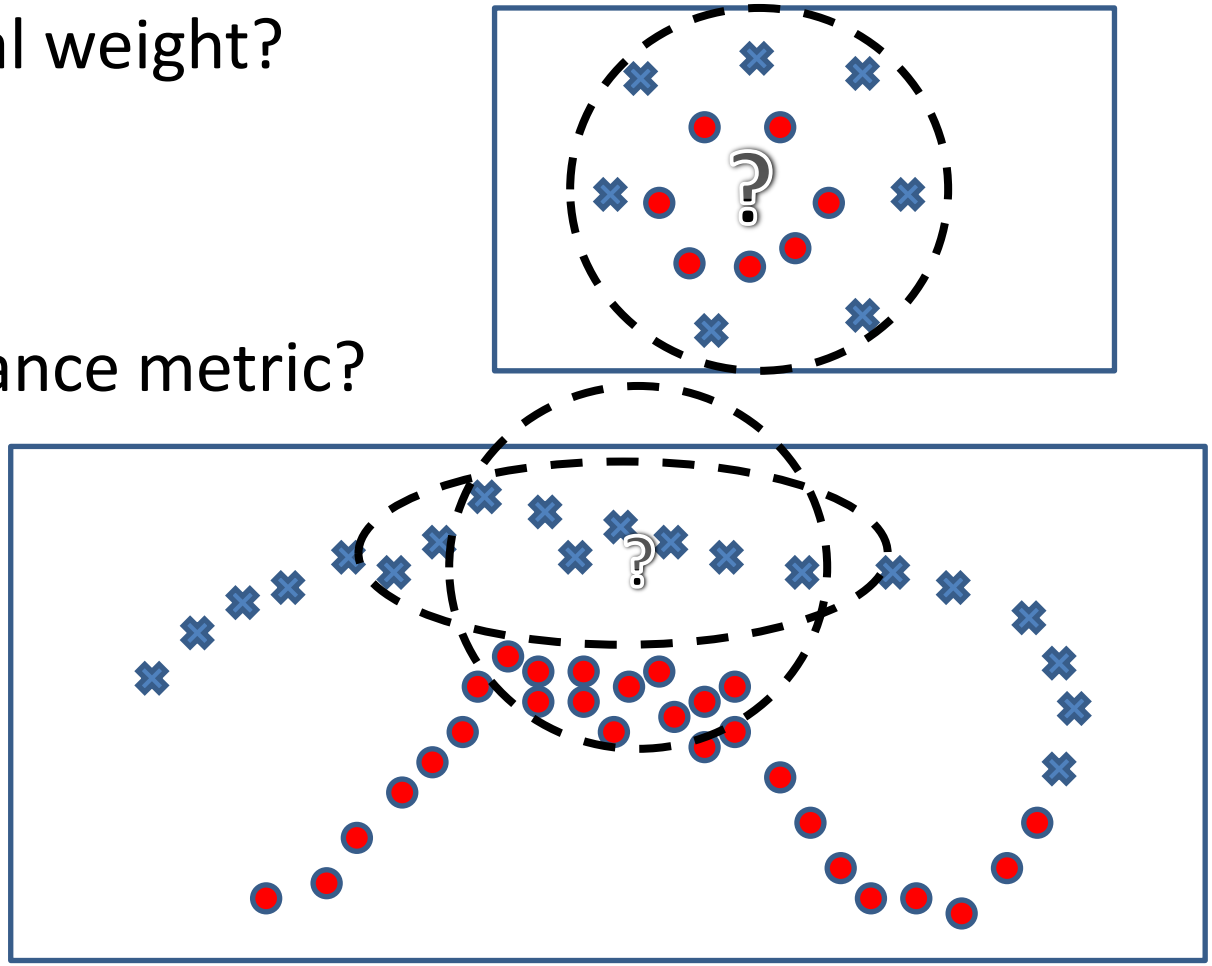
Non-parametric (Instance-based functions):



- Techniques: (adaptive) *K* nearest neighbor, ……

4. Different machine learning techniques

- How large *K* is? About the model complexity
- Equal weight?
- Distance metric?



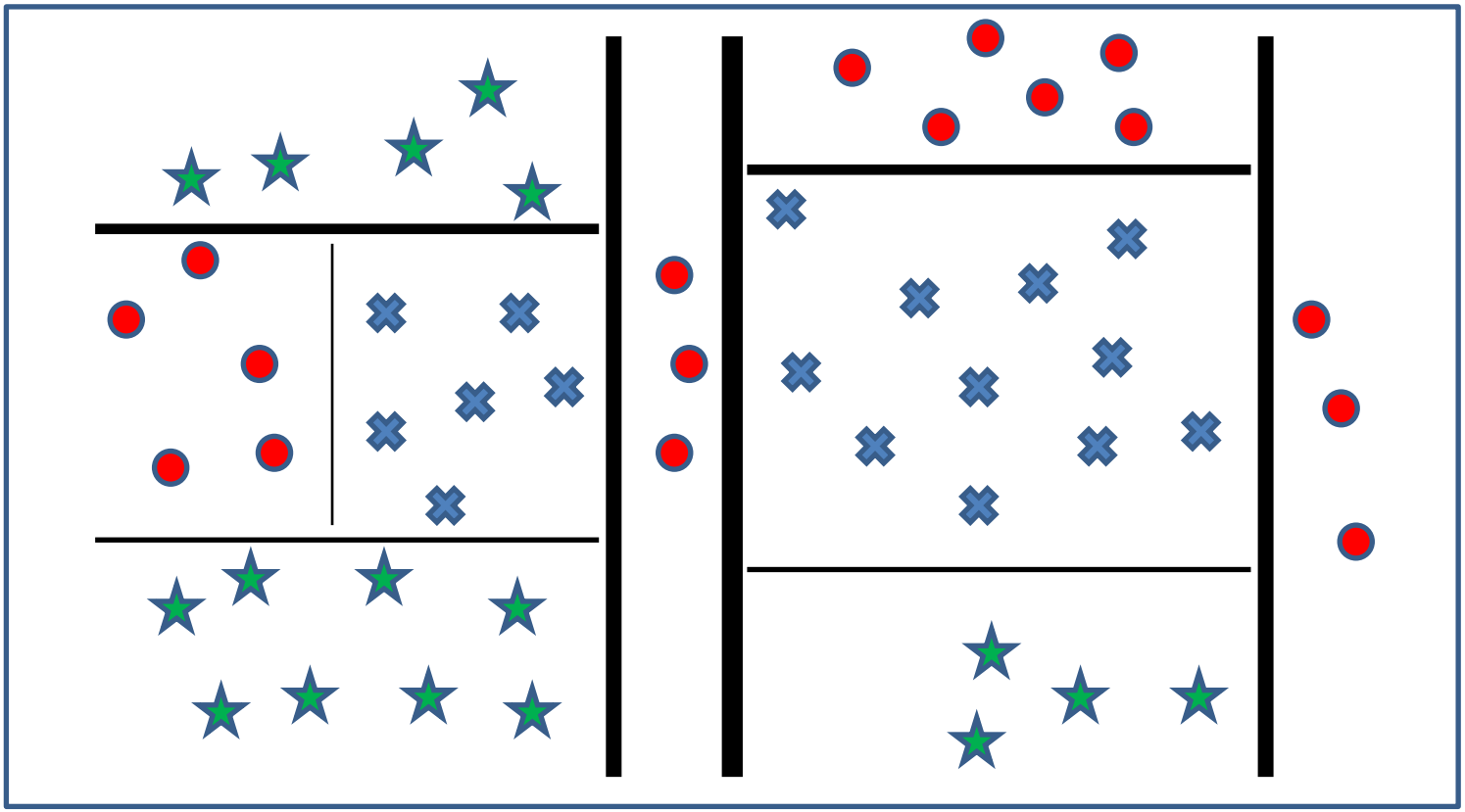
4. Different machine learning techniques

- Linear classifier (numerical functions)
- Non-parametric (Instance-based functions)
- Non-metric (Symbolic functions)
- Parametric (Probabilistic functions)
- Aggregation

4. Different machine learning techniques

Non-metric (Symbolic functions):

- Techniques: Decision tree, ……



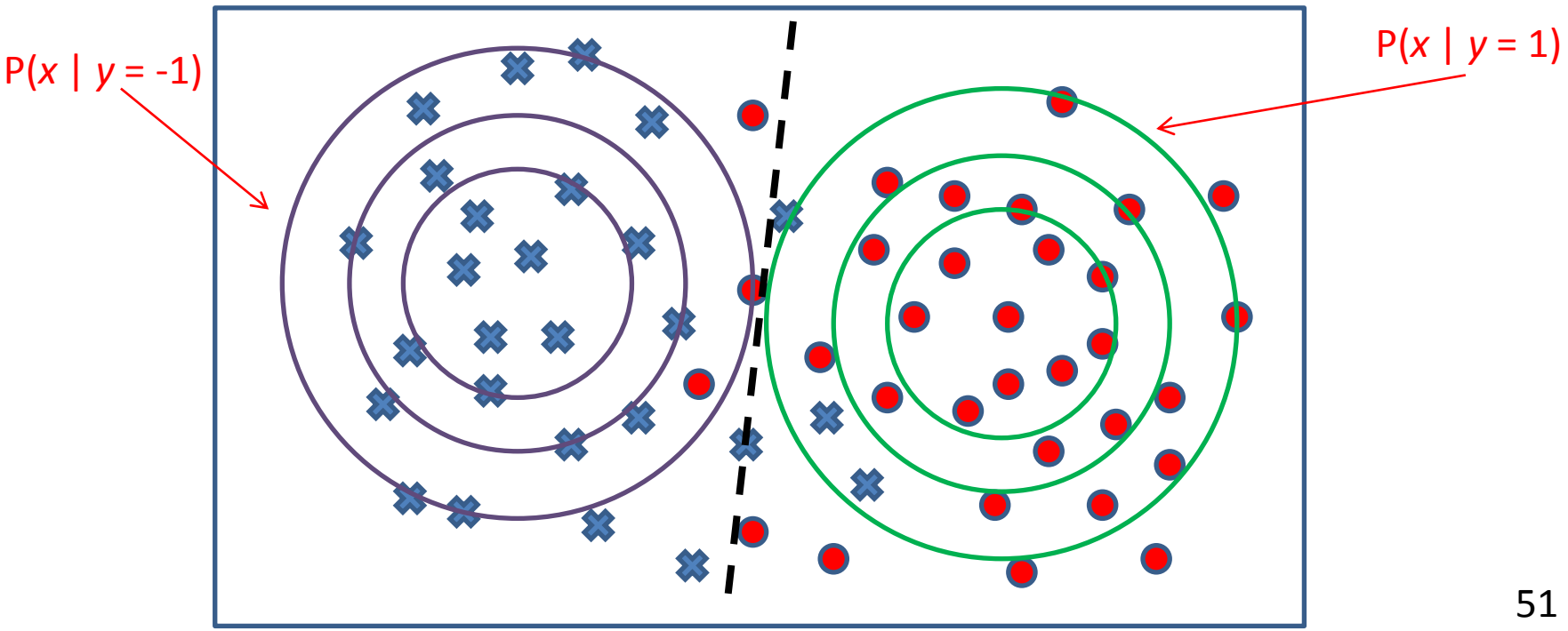
4. Different machine learning techniques

- Linear classifier (numerical functions)
- Non-parametric (Instance-based functions)
- Non-metric (Symbolic functions)
- Parametric (Probabilistic functions)
- Aggregation

4. Different machine learning techniques

Parametric (Probabilistic functions):

- Techniques: Gaussian discriminant analysis (GDA), Naïve Bayes, Graphical models, ……



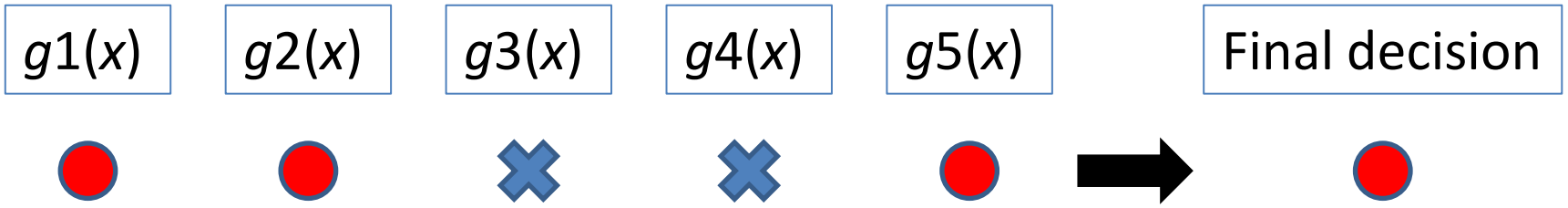
4. Different machine learning techniques

- Linear classifier (numerical functions)
- Non-parametric (Instance-based functions)
- Non-metric (Symbolic functions)
- Parametric (Probabilistic functions)
- Aggregation

4. Different machine learning techniques

Aggregation: (ensemble learning)

- Assume we have trained 5 models and get 5 final hypotheses
- For a test x



- Techniques: Adaboost, Bagging, Random forest, ……

4. Different machine learning techniques

Type	Adv	Dis-adv	Techniques
Linear	low d_{VC}	Too easy	Logistic regression, Ada-line, perceptron learning rule, ……
Linear + feature transform		Hard to design transform functions	2nd –order polynomial transform,……
SVM	Powerful Over-fitting	Computational complexity	C-SVM, v-SVM, SVR, …… SVM + RBF, SVM + poly, ……
MLP	Powerful	Convergence?	Back-propagation MLP
Non-parametric	Easy idea	Computational complexity	KNN, adaptive KNN, KNN with kernel smoother, ……
Decision tree	Easy idea	Over-fitting	CART, ……
Parametric	Flexible Discovery	Computational complexity	GDA, Naïve Bayes, Graphical models, PLSA, PCA, ……
Aggregation	Powerful Over-fitting	Computational complexity	Adaboost, Bagging, Random forest, Blending, ……

4. Different machine learning techniques

Machine learning revised:

- Theory: $$Eout(g) \leq Ein(g) \pm O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right)$$
- Goal: $$g = \underset{H}{argmin}\, Eout(h)$$
- Methods:

$$\min_{H} Eapp(h) \rightarrow \min_{H} Ein(h) \xrightarrow{\text{VC bound}} \text{low } Eout(g)$$

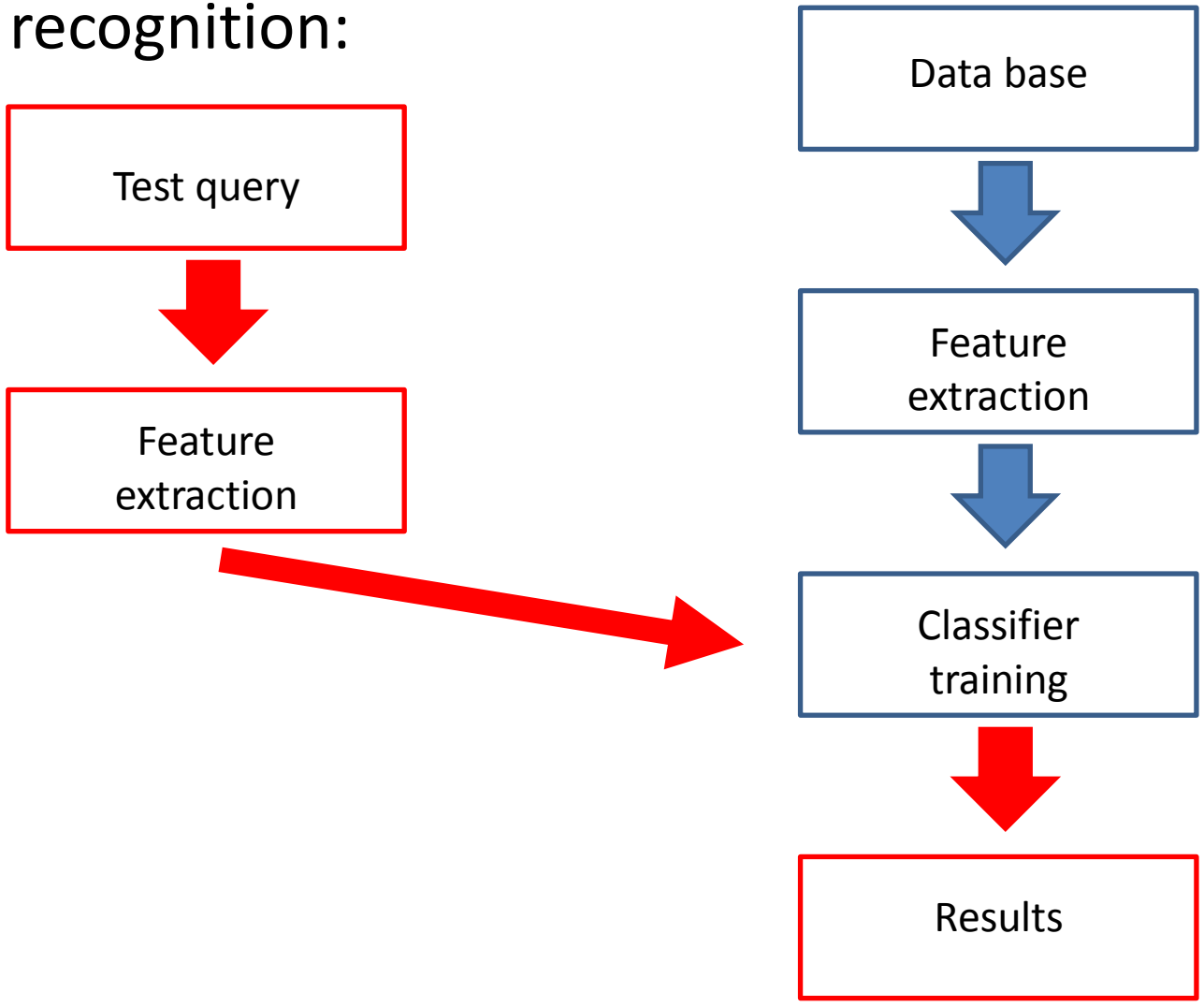
- Usage: Linear → SVM, Adaboost, Random forest

Outline

- 1. What is machine learning?
- 2. The basic of machine learning
- 3. Principles and effects of machine learning
- 4. Different machine learning techniques (hypotheses)
  - Linear classifier (numerical functions)
  - Non-parametric (Instance-based functions)
  - Non-metric (Symbolic functions)
  - Parametric (Probabilistic functions)
- 5. Practical usage and application: Pattern recognition
- 6. Conclusion and discussion

5. Practical usage and application

Pattern recognition:



5. Practical usage and application

Pattern recognition:

(1) Data pre-processing

- Raw data: music signal, image, ……
- Feature extraction: DCT, FT, Gabor feature, RGB, MFCC, ……
- Feature selection & feature transform: PCA, LDA, ICA, ……

(2) Classifier training

- Choose a model and set parameters
- Train the classifier
- Test the performance, either with validation or regularization
- Find another model?

Outline

- 1. What is machine learning?
- 2. The basic of machine learning
- 3. Principles and effects of machine learning
- 4. Different machine learning techniques (hypotheses)
  - Linear classifier (numerical functions)
  - Non-parametric (Instance-based functions)
  - Non-metric (Symbolic functions)
  - Parametric (Probabilistic functions)
- 5. Practical usage and application: Pattern recognition
- 6. Conclusion and discussion

6. Conclusion and discussion

Factors: Training set with *N* and *d*

Assumptions: Hypothesis, loss functions, and same distribution

Theory: (1)

$$Eout(g) \leq Ein(g) \pm O\left(\sqrt{\frac{d_{VC}}{N} ln\, N}\right)$$

(2) Probabilistic: Assume the basic distribution is known

Cautions:

- Over-fitting VS. Under-fitting
- Bias VS. Variance
- Learning curve

6. Conclusion and discussion

- Machine learning is powerful and widely-used in recent researches such as pattern recognition, data mining, and information retrieval
- Knowing how to use machine learning algorithms is very important
- Q&A?

Thank you for listening