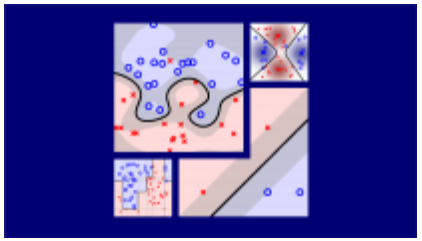
Machine Learning Techniques

(_hx?€?)



Lecture 1: Linear Support Vector Machine

Hsuan-Tien Lin (??0)

htlin@csie.ntu.edu.tw

Department of Computer Science
& Information Engineering

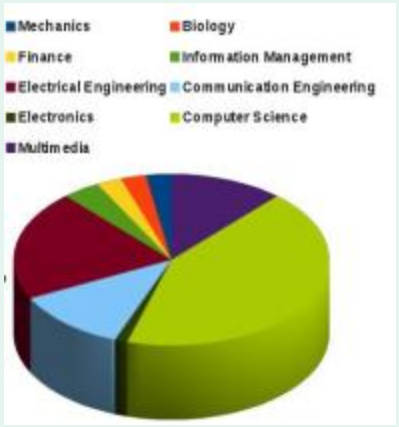
National Taiwan University

(??c"x? ? ?)

Course History

NTU Version

- 15-17 weeks (2+ hours)
- highly-praised with **English and blackboard teaching**



Coursera Version

- 8 weeks of 'foundations' (previous course) + 8 weeks of 'techniques' (**this course**)
- **Mandarin teaching** to reach more audience in need
- **slides teaching** improved with Coursera's quiz and homework mechanisms

goal: **try** making Coursera version even better than NTU version

Course Design

from Foundations to Techniques

- mixture of philosophical illustrations, key theory, core algorithms, usage in practice, and hopefully jokes **:-)**
- three major techniques surrounding **feature transforms**:
  - Embedding Numerous Features: how to **exploit** and **regularize** numerous features?
    —inspires **Support Vector Machine** (SVM) model
  - Combining Predictive Features: how to **construct** and **blend** predictive features?
    —inspires **Adaptive Boosting** (AdaBoost) model
  - Distilling Implicit Features: how to **identify** and **learn** implicit features?
    —inspires **Deep Learning** model

allows students to **use ML professionally**

Fun Time

Which of the following description of this course is true?

1. the course will be taught in Taiwanese
2. the course will tell me the techniques that create the android Lieutenant Commander Data in Star Trek
3. the course will be 16 weeks long
4. the course will focus on three major techniques

Fun Time

Which of the following description of this course is true?

1. the course will be taught in Taiwanese
2. the course will tell me the techniques that create the android Lieutenant Commander Data in Star Trek
3. the course will be 16 weeks long
4. the course will focus on three major techniques

Reference Answer: 4

1. no, my Taiwanese is unfortunately not good enough for teaching (yet)
2. no, although what we teach may serve as building blocks
3. no, unless you have also joined the previous course
4. yes, **let's get started**!

Roadmap

1. Embedding Numerous Features: Kernel Models

Lecture 1: Linear Support Vector Machine

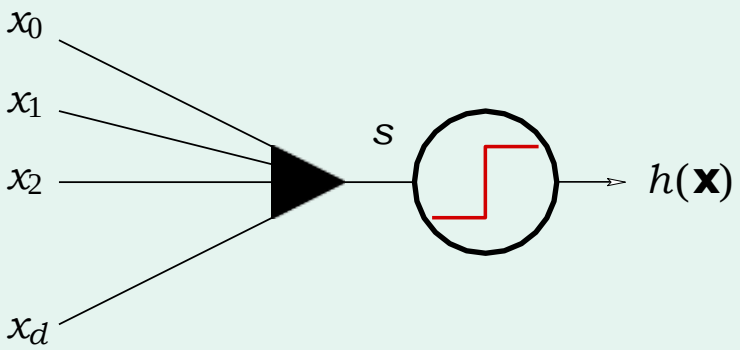
- Course Introduction
- Large-Margin Separating Hyperplane
- Standard Large-Margin Problem
- Support Vector Machine
- Reasons behind Large-Margin Hyperplane

2. Combining Predictive Features: Aggregation Models
3. Distilling Implicit Features: Extraction Models

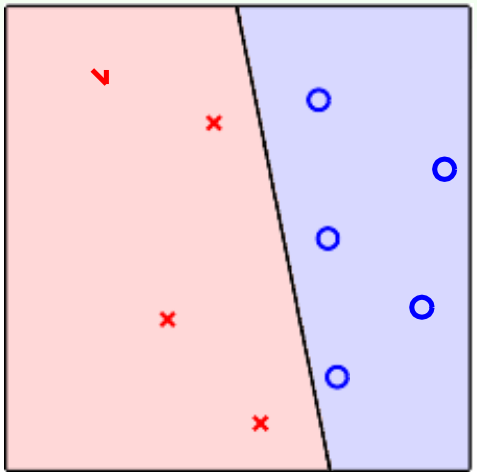
Linear Classification Revisited

PLA/pocket

$h(\mathbf{x}) = \text{sign}(s)$



plausible err = 0/1
(small flipping noise)
minimize **specially**



(linear separable)

linear (hyperplane) classifiers:
$h(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x})$