# An Overview of Recognition System based on the Support Vector Machines

Author : Balwant A.Sonkamble,
D.D.Doye

Professor: 陳嘉平
Repotor: 吳國豪

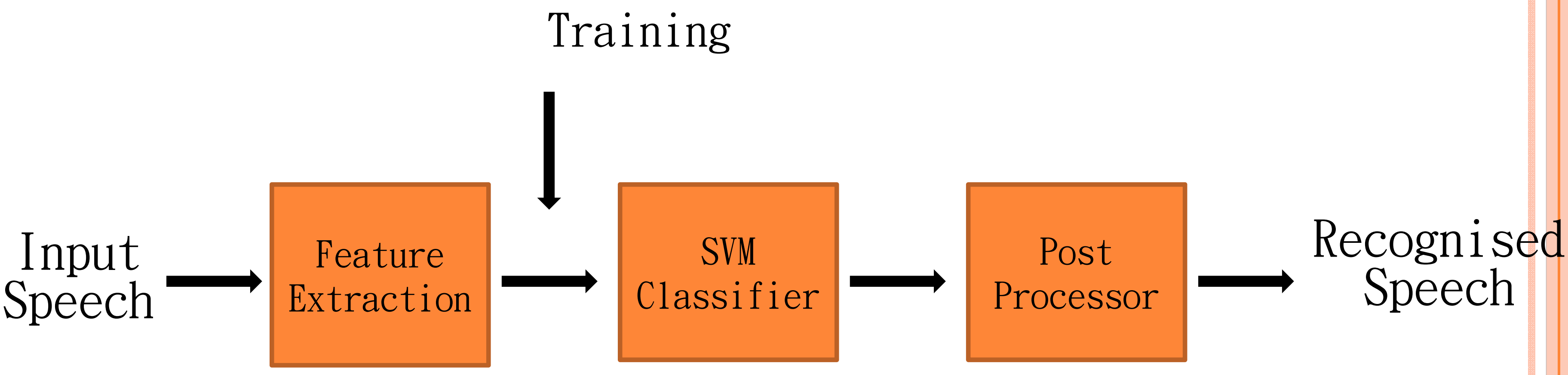
# OUTLINE

- Introduction
- Speech Recognition System
- Support Vector Machine (SVM) Classifier

# INTRODUCTION

- The classification modules are playing a very important role in most of the modern speech recognition systems.
- The speech recognition uses neural networks, hidden Markov models, Bayesian networks, DTW and other tools for recognizing the particular speech.
- The SVM is designed for solving the binary classification problems

# Speech Recognition System



# Feature Extraction

- To have a good discrimination for distinguishing between the similar speeches, this is easy to model statistically by considering fewer amounts of training data

- to have statistical properties which are invariant across speakers and over a wide range of speaking environments.

# Support Vector Machine (SVM) Classifier

- The SVM is designed for solving the binary classification problems

- SVM: (1) Linear SVM
  (2) Nonlinear SVM

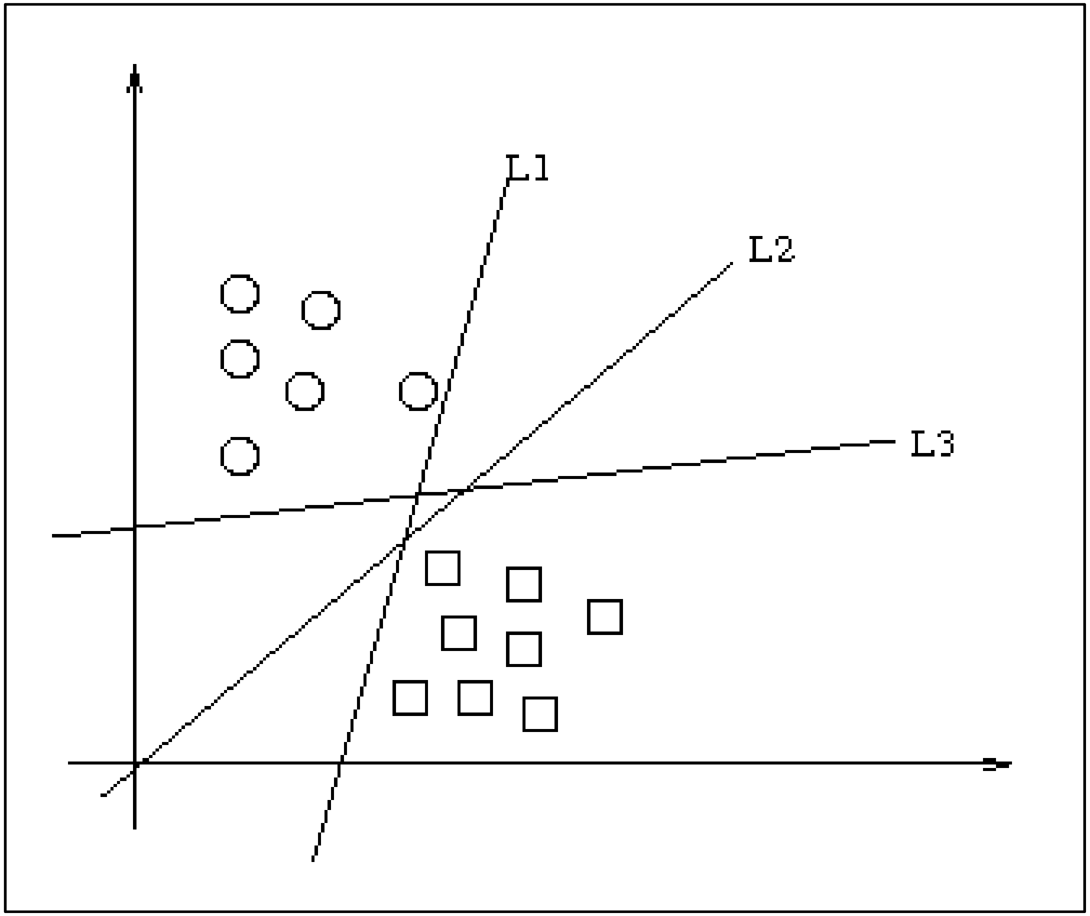



*Fig.*1

# Linear SVM

- SVM is a binary classifier that makes its decisions by constructing a hyperplane that optimally separates the two classes.

- We are given some training data, a set of points of the form:

$$\{(x_i, y_i)\}, x_i \in R^d, y_i \in \{-1,1\}, i = 1, \ldots, N$$

$x_i$ : condition attribute

$y_i$ : class attribute

# Linear SVM

- The hyperplane is defined :

$$w \cdot x + b = 0$$

$w$: the normal to the plane

- The optimal hyperplane is chosen according to the maximum margin criterion

$$x_i \cdot w + b \geq +1 \quad \text{for } y_i = +1 \qquad (1)$$

$$x_i \cdot w + b \leq -1 \quad \text{for } y_i = -1 \qquad (2)$$

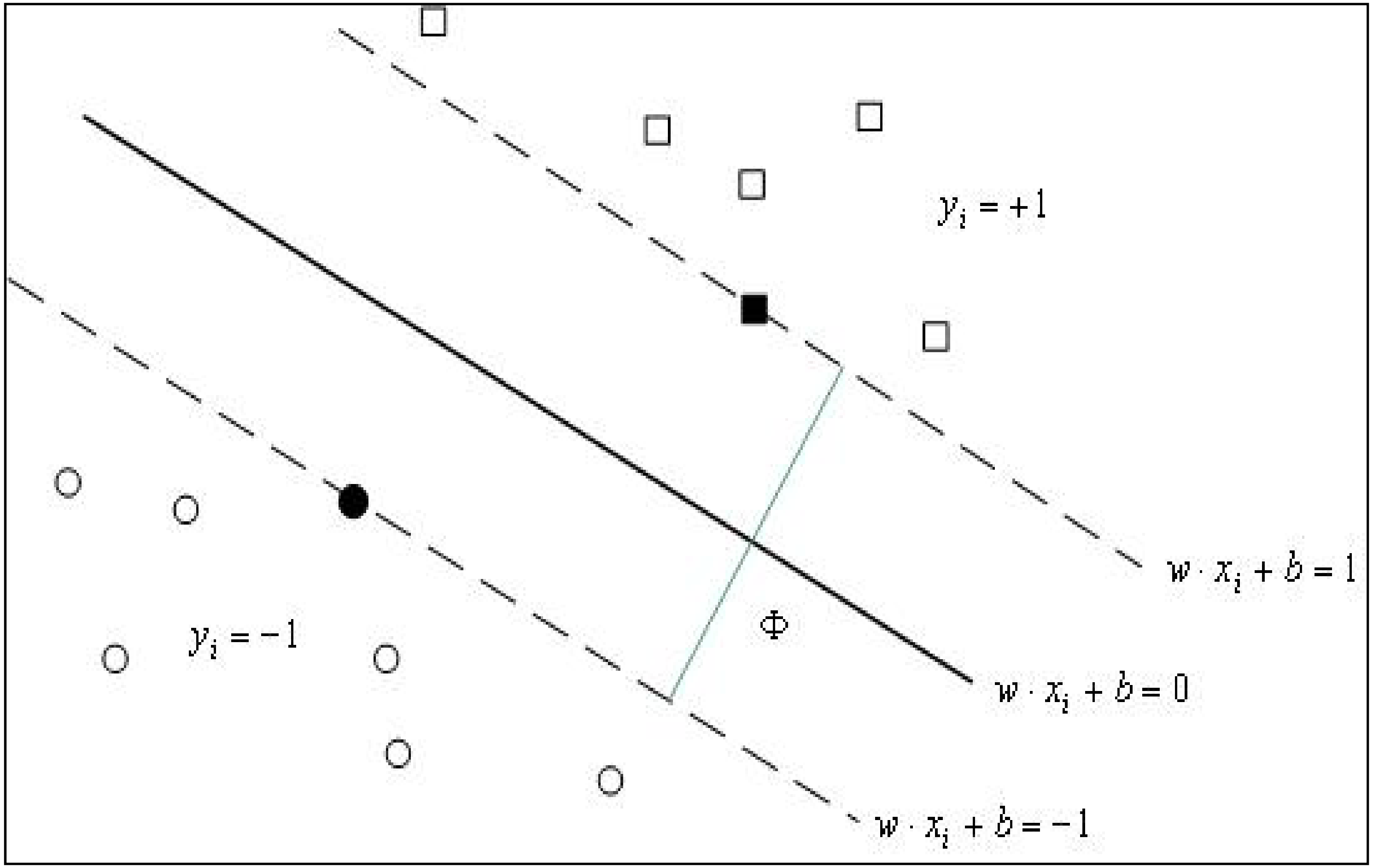



*Fig*.2

令 $x_i \cdot w + b = +1$ 爲 $H_1$, $d_1 = \dfrac{1}{\|w\|}$ ; 令 $x_i \cdot w + b = -1$ 爲 $H_2$, $d_2 = \dfrac{1}{\|w\|}$

$H_1, H_2$ is support hyperplane

# Linear SVM

- Margin: $d_1 + d_2 = \frac{2}{\|w\|}$ minimize: $\|w\|^2$

$$\min \phi(w) = \frac{1}{2}\|w\|^2$$

- Combine(1)(2): $y_i(x_i \cdot w + b) \geq 1, \forall i$

- We switch to Lagrange formulation:

$$L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} a_i \left[ y_i(w \cdot x_i + b) - 1 \right]$$

# Linear SVM

- 對 $w$ 偏微分: $\dfrac{\partial}{\partial w} L = 0 \Rightarrow w = \sum_{i=1}^{N} a_i y_i x_i$
- 對 $b$ 偏微分: $\dfrac{\partial}{\partial b} L = 0 \Rightarrow \sum_{i=1}^{N} a_i y_i = 0$

$$L_d = \sum a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j (x_i \cdot x_j) \qquad (3)$$

$$Subjectto : a_i \geq 0, i = 1, \ldots, N \quad \text{and} \quad \sum_{i=1}^{N} a_i y_i = 0$$

# LINEAR SVM

- 由(3)得到optimal solution:

$$f(x) = sgn\left(\sum_{i=1}^{N} y_i a_i (x_i x_j) + b\right)$$

$$\text{if } f(x_j) > 0 \Rightarrow y_j = +1$$

$$\text{if } f(x_j) < 0 \Rightarrow y_j = -1$$

# Nonliear SVM

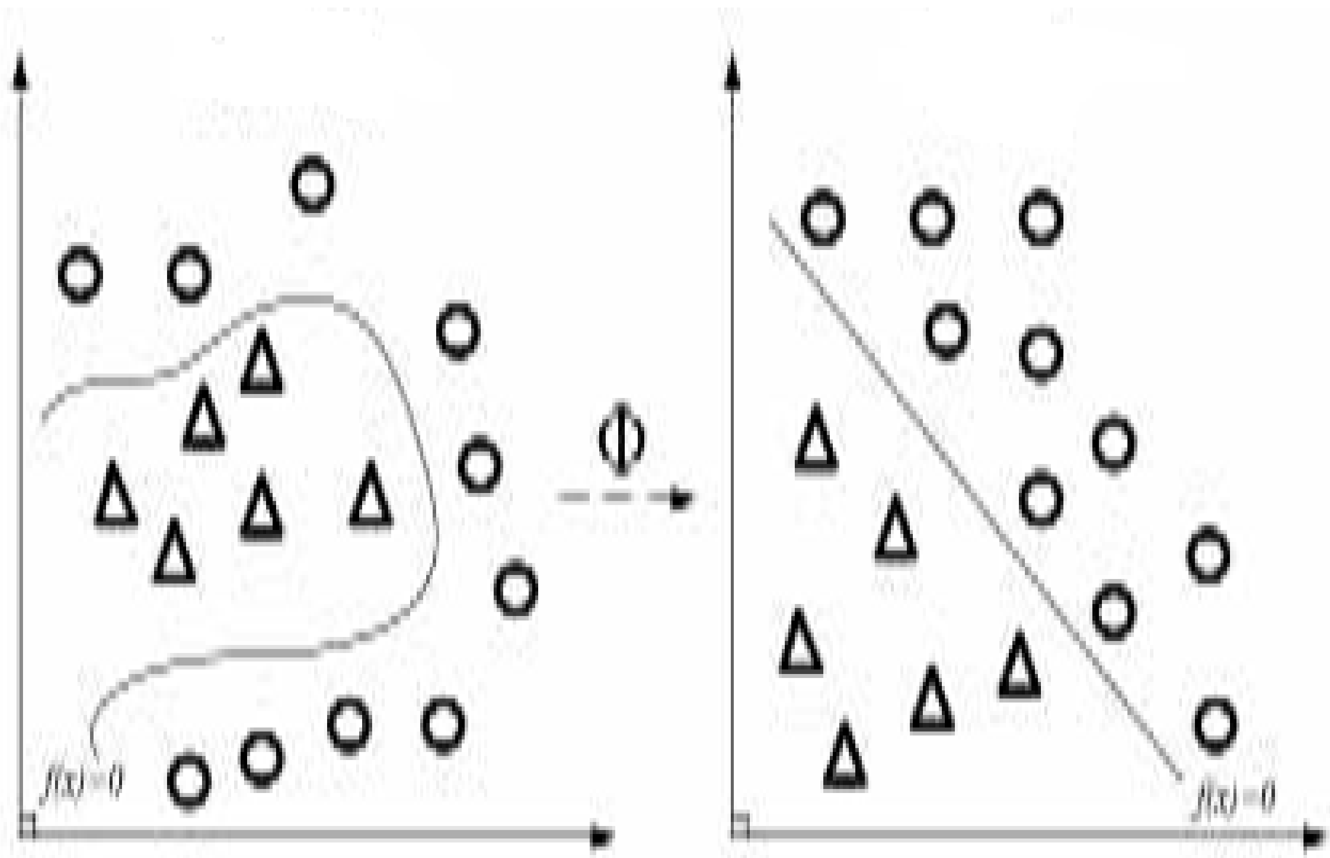
- For the nonlinear case, the problem is transformed as that of linear classification in the space *H by* projecting the original data into a high dimensional space *H* using a nonlinear map:

$$\Phi(x_i):\ \mathrm{R}^d \rightarrow \mathrm{H}$$

- Kernel function : $K(x_i, x_j) = (\ \Phi(x_i)\cdot\Phi(x_j)\ )$

$$f(x) = sgn(\ \sum_{i=1}^{N} y_i a_i k(x_i x_j) + b\ )$$

# Nonliear SVM




$Fig.3 \qquad \Phi(x_i) : R^d \rightarrow H$

# Multi-Class Classification

- The construction of n-class classifer is done by using the construction of two-class classification

- One-against-rest(OAR)

- One-against-one(OAO)
  (Directed Acyclic Graph SVM)

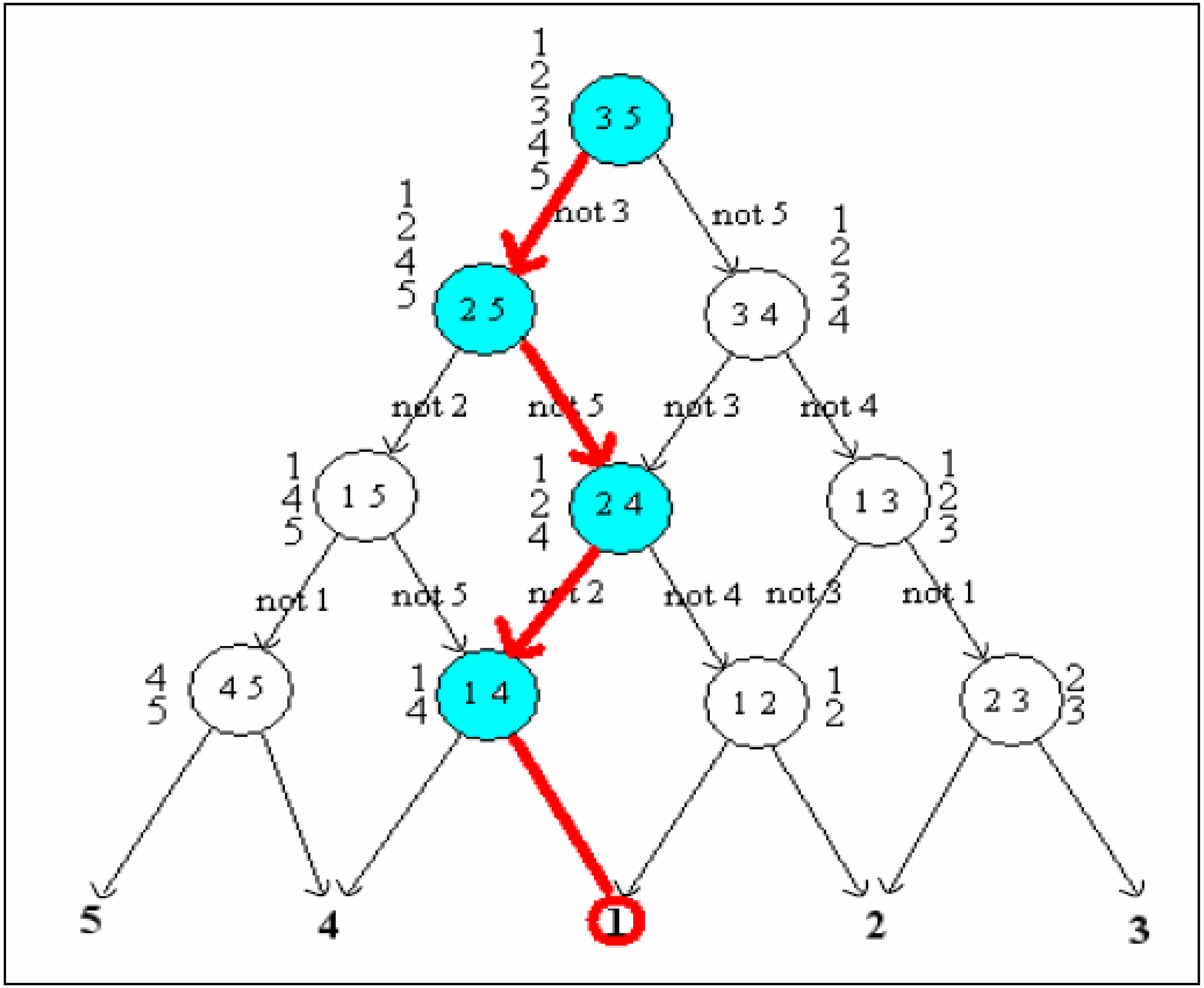



*Fig*.4 One-against-one(OAO)

# Post Processor

- In this process, the SVM outputs are converted into the probabilities and then finding the maximum probability at each instant and choosing the best probability for recognizing the speech inputs.

# Applications of SVM

- SVMs are applied on number of computer vision application such Natural language processing, speech/ speaker recognition, verification and Identification, Image Recognition.