

# Discriminant Analysis

<http://archive.ics.uci.edu/ml/>



The screenshot shows the homepage of the UCI Machine Learning Repository. At the top, there is a navigation bar with the UCI logo, a search bar, and links for 'About', 'Contribute', 'Data Set', and 'Contact'. Below the navigation bar, a welcome message states: 'Welcome to the UC Irvine Machine Learning Repository! We currently maintain 211 data sets as a service to the machine learning community. You may view all data sets through our searchable interface. Our REST API is still available for those who prefer the old format. For a general overview of the Repository, please visit our home page. For information about citing data sets in publications, please read our citation policy. If you wish to donate a data set, please consult our donation policy. For any other questions, feel free to contact the Repository Staffers. We have also set up a mailing list for the Repository.' Below the welcome message, there are logos for 'Supported By' (UCI) and 'In Collaboration With' (Data Science Central). The main content area is divided into three columns: 'Latest News' with a list of recent updates, 'Newest Data Sets' with a list of recently added datasets including 'Amazon Access Services', 'Vertical Columns', 'Viper Physical Action Data Set', 'EMG Physical Action Data Set', 'CoopBank Review Dataset', 'Relative location of CT slices on AMU sets', and 'Amazon Commerce reviews'; and 'Most Popular Data Sets (since 2007)' with a list of popular datasets including 'Iris', 'Adult', 'Wave', 'Breast Cancer Wisconsin (Diagnostic)', 'Car Evaluation', 'Mushrooms', and 'Poker Hand'. A 'Featured Data Set' section is also visible, featuring 'Spectra' with a 'Task' label and a small image of a spectrum.

# Discriminant analysis

## Tools & neural approaches

1. <http://134.208.26.59/AdvancedNA/Lecture6/Lecture62.files/frame.htm>
2. <http://134.208.26.59/NM/Lecture6/Lecture622.files/frame.htm>

## Examples

1. Hill and Valley

<http://134.208.26.59/INA/Discriminant%20analysis.files/frame.htm>

2. Hand written character recognition

<http://134.208.26.59/INA/PenWriting.files/frame.htm>

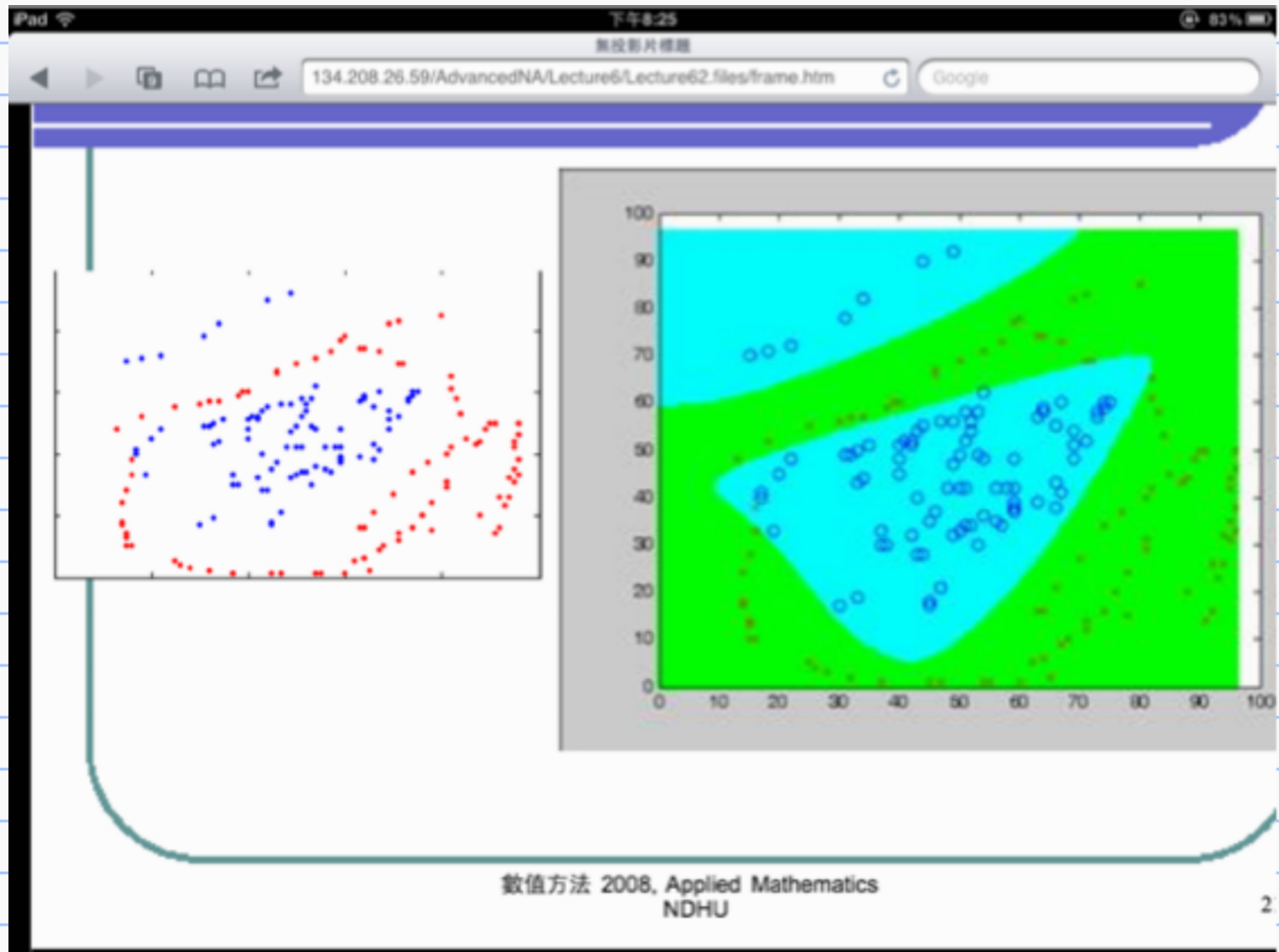
3. Breast cancer diagnosis

[http://134.208.26.59/INA/Cancer\\_Diagnosis.files/frame.htm](http://134.208.26.59/INA/Cancer_Diagnosis.files/frame.htm)

The screenshot shows the homepage of the UCI Machine Learning Repository. At the top, there is a navigation bar with the UCI logo, the text "Machine Learning Repository", and a search bar. Below the navigation bar, there is a welcome message and a brief description of the repository. The main content area is divided into three columns: "Latest News", "Newest Data Sets", and "Most Popular Data Sets (sets since 2007)".

Latest News	Newest Data Sets	Most Popular Data Sets (sets since 2007)
2010-03-01: <a href="#">Notice from donor regarding Netflix data</a>	2011-09-13: <a href="#">UCI Amazon Access Samples</a>	201058: <a href="#">Tic</a>
2009-10-06: <a href="#">Two new data sets have been added</a>	2011-08-09: <a href="#">UCI YachtRival Columns</a>	204423: <a href="#">Adult</a>
2009-09-04: <a href="#">Several data sets have been added</a>	2011-07-27: <a href="#">UCI Year.Physical.Action.Data.Set</a>	178344: <a href="#">Wine</a>
2008-07-23: <a href="#">Repository .html has been set up</a>	2011-07-27: <a href="#">UCI EMG.Physical.Action.Data.Set</a>	142871: <a href="#">Breast.Cancer.Wisconsin.Characteristics</a>
2008-03-24: <a href="#">New data sets have been added</a>	2011-07-26: <a href="#">UCI CreditCard.Review.Dataset</a>	128879: <a href="#">Car.Evaluation</a>
2007-06-26: <a href="#">Two new data sets have been added: GUI.Pan.Character, WISC.Gemini.Telescope</a>	2011-07-27: <a href="#">UCI Relative.location.of.ICI.written.on.MSIC.pdf</a>	113141: <a href="#">Abalone</a>
2007-04-03: <a href="#">Research papers that cite the repository have been associated to specific data sets</a>	2011-07-27: <a href="#">UCI Amazon.Commerce.reviews</a>	104186: <a href="#">Vowel.Vowel</a>
Featured Data Set: <a href="#">Internet Advertisements</a>		

Discriminant analysis & Classification  
Nonlinear separability vs. Linear separability  
Data driven neural approaches  
Conditional Probability Estimation  
Dimensionality reduction  
Similarity measure  
Temporal and spatial Feature extraction  
Training and testing



# Training data

- predictors
- $x[t]$
- multivariate
- desired targets :
  - $y[t]$
  - a set of labels or categories
  - two categories  $\{+1, -1\}$  or  $\{0, 1\}$
  - multiple categories  $\{e_1, e_2, \dots, e_K\}$
- paired data :  $(x[t], y[t])$
- a target function underlying paired data

$\mathbb{R}^d$

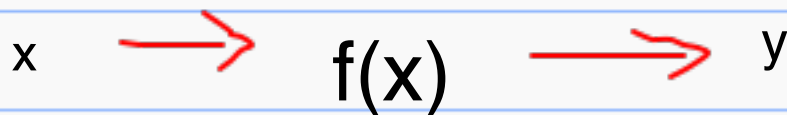
$d = 1$

2 - 3

1, 8

> 10

> 100



$$K = 3$$

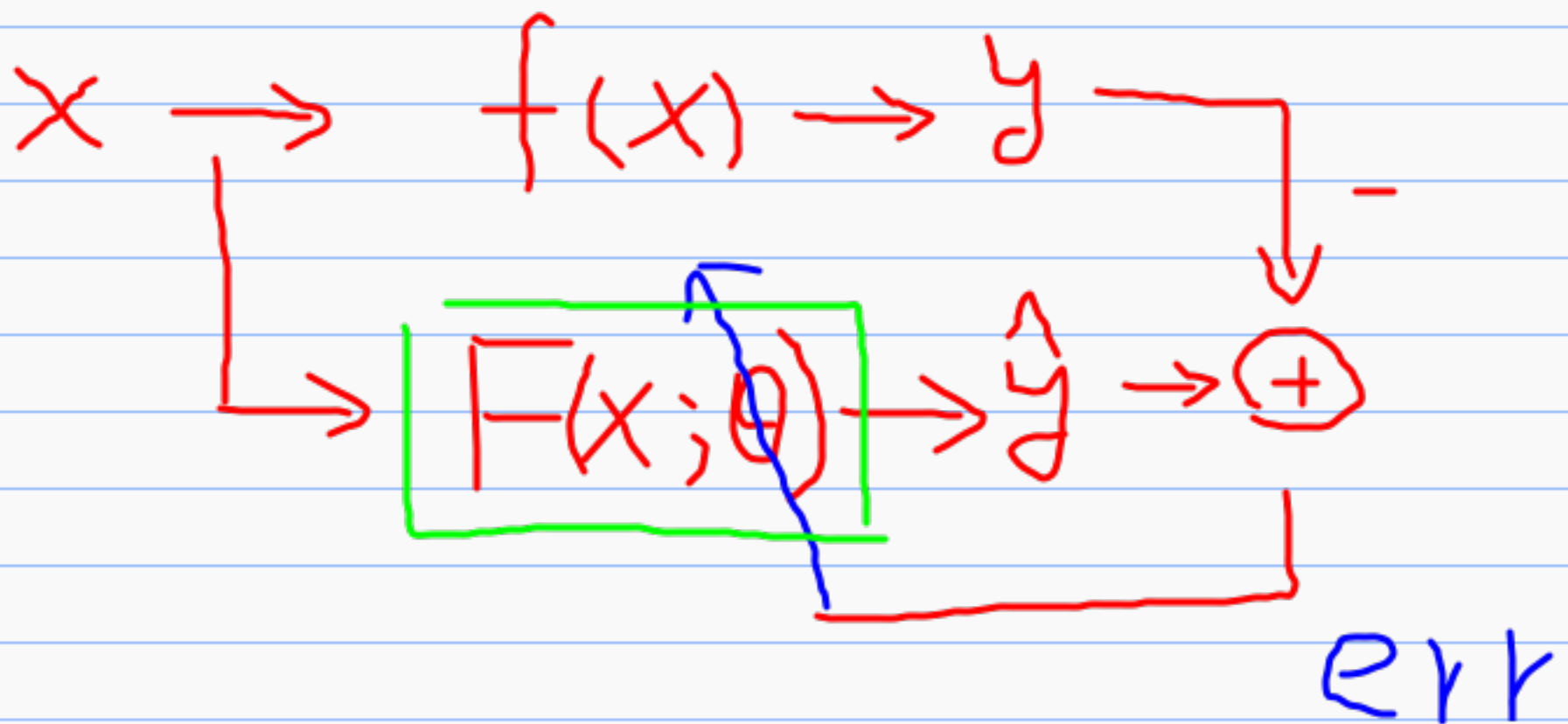
$$\phi_1 = (1 \ 0 \ 0)^T$$

$$\phi_2 = (0 \ 1 \ 0)^T$$

$$\phi_3 = (0 \ 0 \ 1)^T$$

# Training and Testing

- Training phase
- $F(x; \theta)$
- $\theta$  denotes adaptable semi-parametric representations



# Numerical simulations

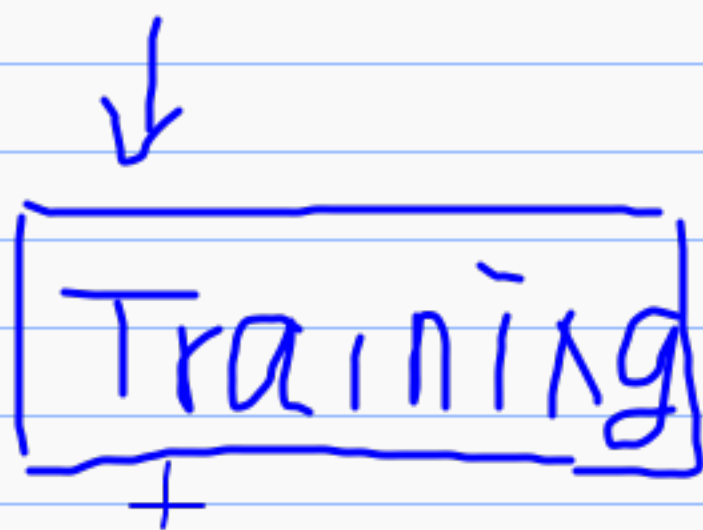
## Data flow

$$S = \{(x_t, y_t)\}$$

random split

$S_{\text{train}}$

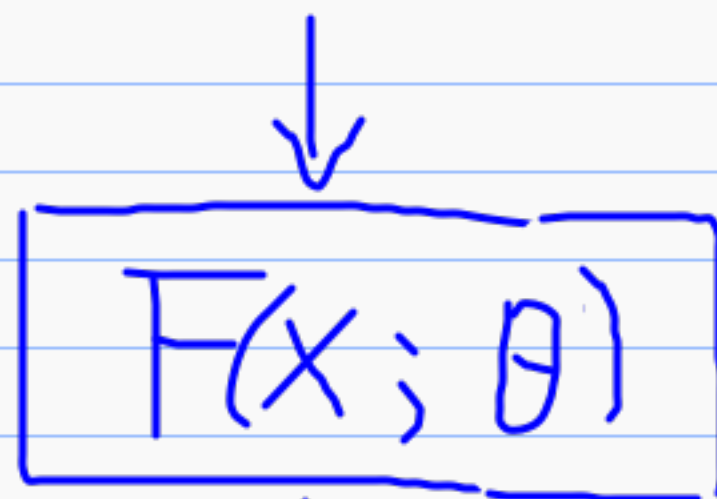
$S_{\text{test}}$



mse

Training

$\theta$



mean square error

# Discriminant analysis

1. The XOR problem: the DA is expected to extract an XOR rule.

- true table

- a. 1 1 0 0

- b. 1 0 1 0

- c. 0 1 1 0



- training data

- $x = (a \ b)$

- $y = c$

- linear separability

- no probability

- no noise

- boolean algebra versus numerical functions

$(0, 1)$   
0

0  $(1, 1)$

0

0


$(0, 0)$

$(1, 0)$



http://archive.ics.uci.edu/ml/

## 2. Tic-Tac-Toe Endgame data set

**UCI**   
**Machine Learning Repository**  
Center for Machine Learning and Intelligent Systems

About Citation Policy Donate a Data Set Contact


Repository Web

[View ALL Data Sets](#)

### Tic-Tac-Toe Endgame Data Set

Download: [Data Folder](#), [Data Set Description](#)

**Abstract:** Binary classification task on possible configurations of tic-tac-toe game



<b>Data Set Characteristics:</b>	Multivariate	<b>Number of Instances:</b>	958	<b>Area:</b>	Game
<b>Attribute Characteristics:</b>	Categorical	<b>Number of Attributes:</b>	9	<b>Date Donated</b>	1991-08-19
<b>Associated Tasks:</b>	Classification	<b>Missing Values?</b>	No	<b>Number of Web Hits:</b>	36805

**Source:**

**Creator:**  
David W. Aha ([aha@wpi.cs.bzu.edu](mailto:aha@wpi.cs.bzu.edu))

**Donor:**  
David W. Aha ([aha@wpi.cs.bzu.edu](mailto:aha@wpi.cs.bzu.edu))

**Data Set Information:**

This database encodes the complete set of possible board configurations at the end of tic-tac-toe games, where "x" is assumed to have played first. The target concept is "win for x" (i.e., true when "x" has one of 8 possible ways to create a "three-in-a-row").

Interestingly, this raw database gives a stripped-down decision tree algorithm (e.g., ID3) fits. However, the rule-based CN2 algorithm, the simple IB1 instance-based learning algorithm, and the CITRE feature-constructing decision tree algorithm perform well on it.

**Attribute Information:**

### 3. Sampling from two normal distributions

- $p(x; A, b)$  parametric representation
- two sources:  $p(x; A, b_1)$  and  $p(x; A, b_2)$
- linear separability, Quadratic separability

$$P(x; A, b_1) \propto A = \Sigma^{-1}$$

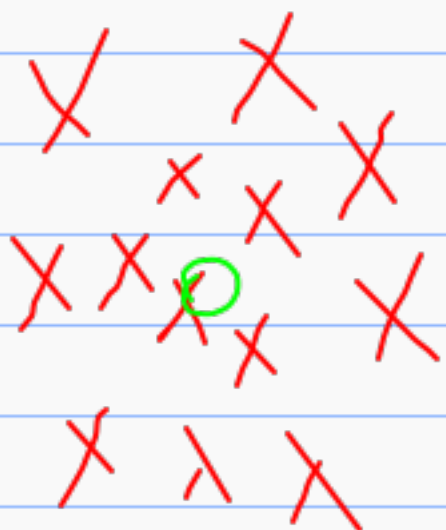
$$\exp\left(-\frac{1}{2}(x - b_1)^T A (x - b_1)\right)$$

$$P(x; A, b_1) \geq P(x; A, b_2)$$

$$\ln P(x; A, b_1) \geq \ln P(x; A, b_2)$$

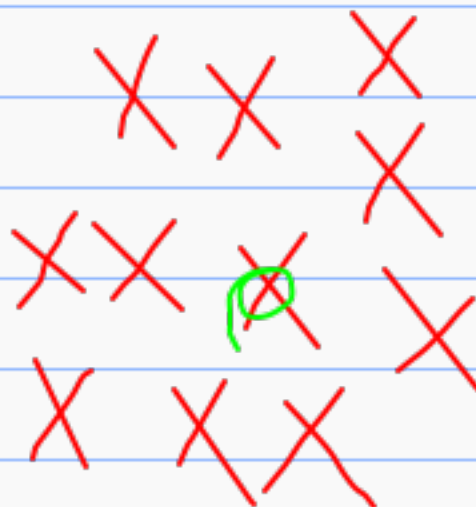
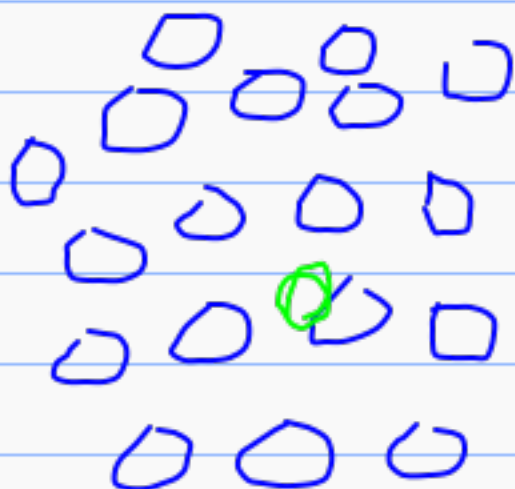
$$\|x - b_1\|_A \leq \|x - b_2\|_A$$

$b_1$



$b_2$

$b_3$

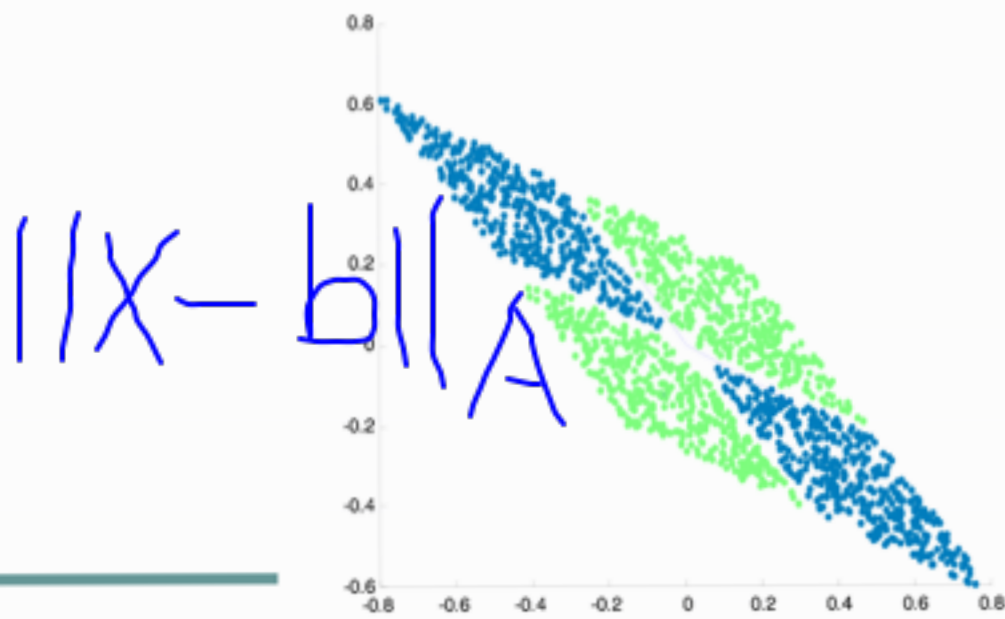


$b_4$

$$\|X - b\|_I$$

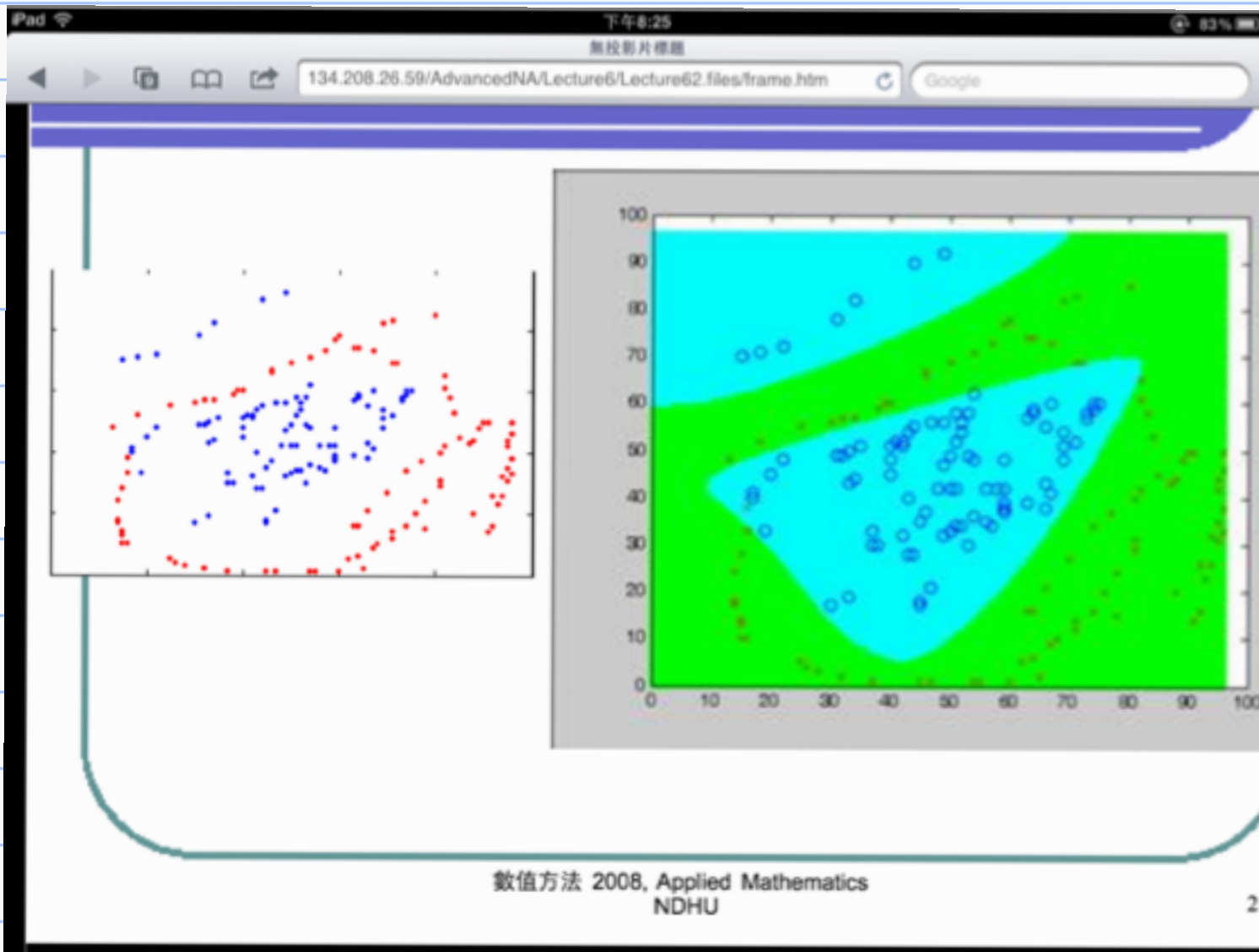
Euclidean distance

- Four local means
- Non-overlapping distributions
- A common weight matrix for rotation



Mahalanobis distance

# 4 Pen drawing



## Methodologies

- hyperplane
- quadratic hyperplane
- a linear combination of radial basis functions
- a linear combination of simple Perceptrons

$$Ax + b$$

$$A_1x + b_1 + A_2x + b_2$$

$$x^T A_1 x + x^T A_2 x$$

$$\rightarrow x^T (A_1 + A_2) x$$

$$\exp\left(-\left(x^T A_1 x + b_1 x\right)\right)$$

$$+ \exp\left(-\left(x^T A_2 x + b_2 x\right)\right)$$

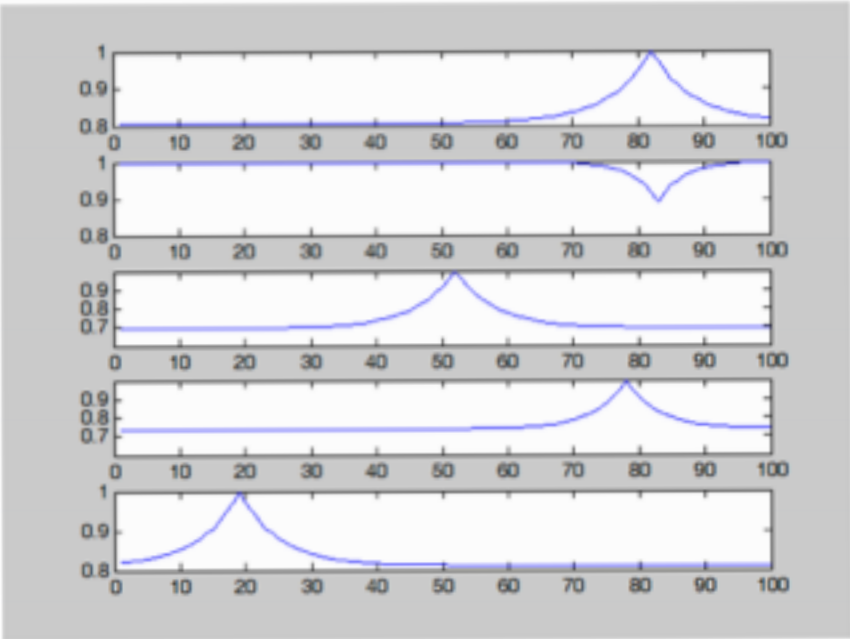
---

radial basis  
function

# 5 Hill Valley

Microsoft PowerPoint - Lecture6.htm  
134.208.26.59/AdvancedNA/Lecture6/Lecture6.pdf

## Noise-free Hill Valley

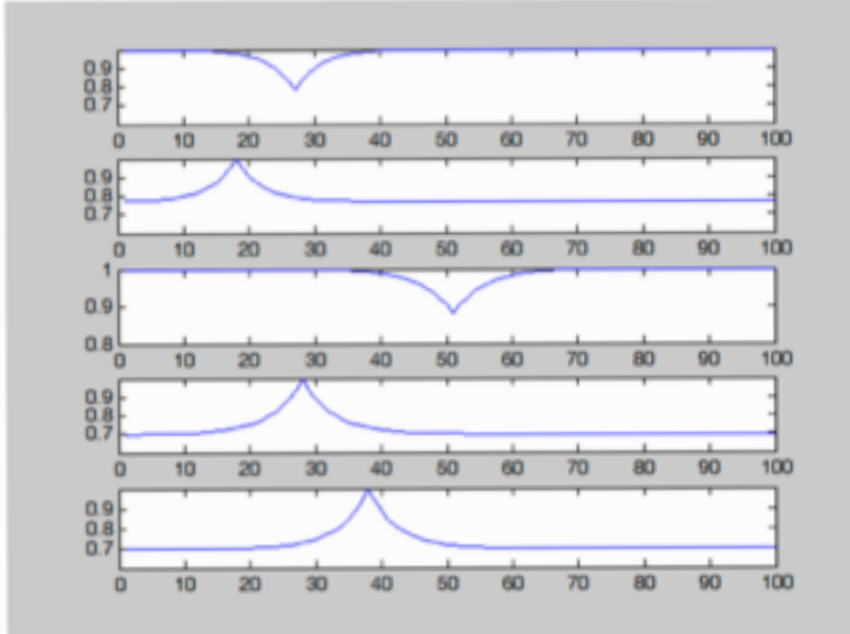


數值方法2008, Applied Mathematics  
NDHU

8

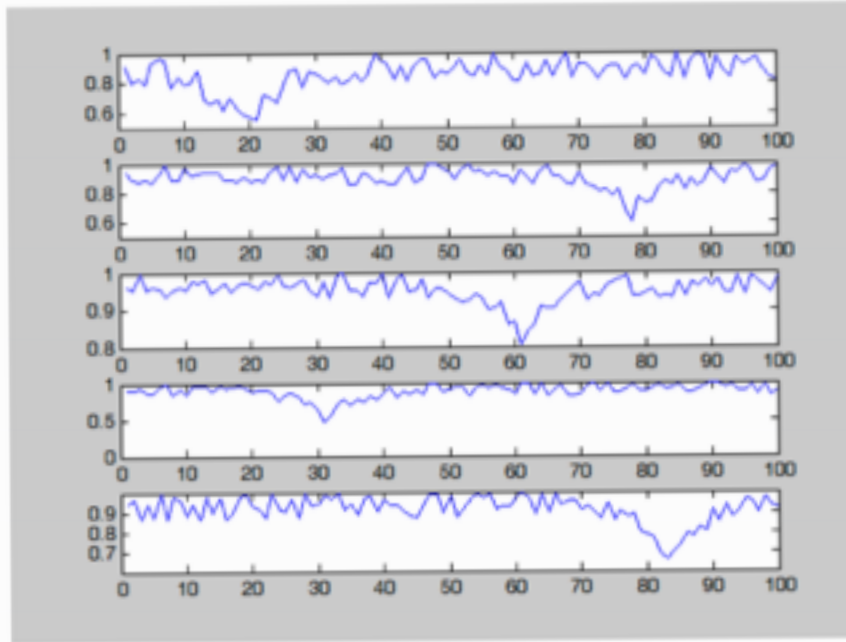


# Noise-free Hill Valley



10/43

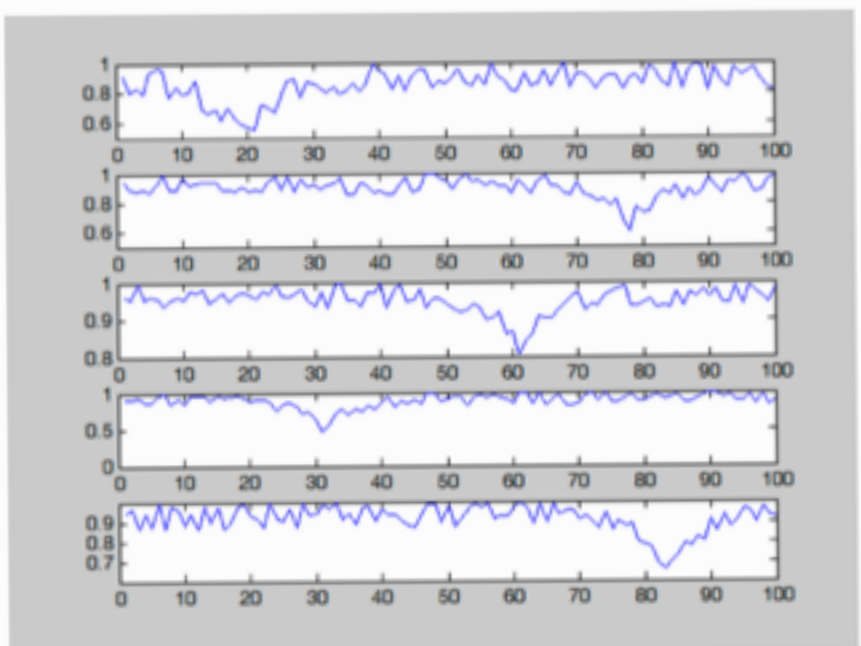
# Hill-Valley with noise



數值方法2008, Applied Mathematics  
NDHU

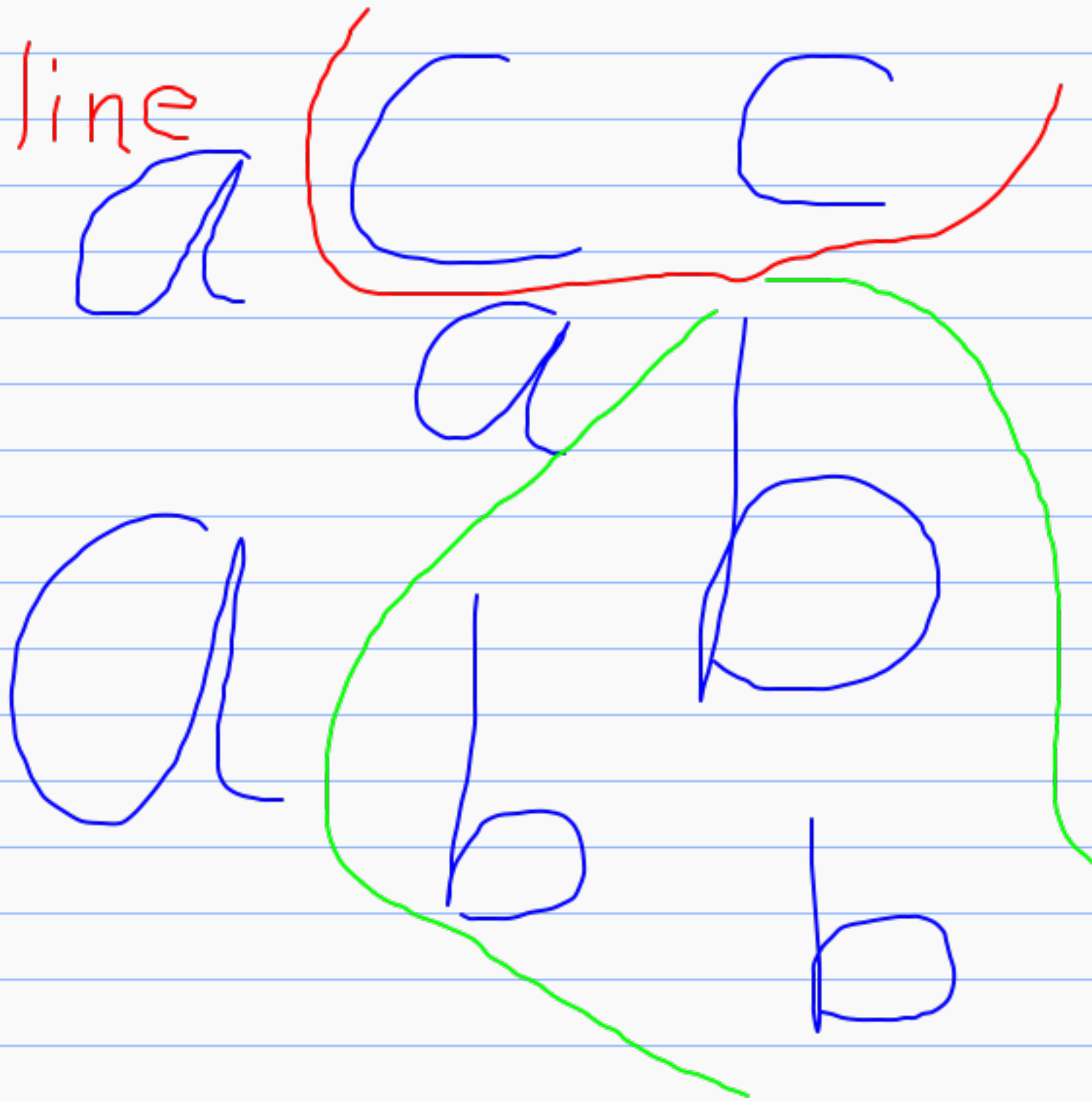
10

# Hill-Valley with noise



# 6 Hand written character recognition

on-line



## 7. Segmentation and recognition



