


# Introduction of Structured Learning

Hung-yi Lee

# Structured Learning

- We need a more powerful function  $f$ 
  - Input and output are both objects with structures
  - *Object*: sequence, list, tree, bounding box ...

$$f : X \rightarrow Y$$


$X$  is the space of  
one kind of object

$Y$  is the space of  
another kind of object

In the previous lectures, the input and output are both vectors.

# Example Application

- **Speech recognition**
  - $X$ : Speech signal (sequence)  $\rightarrow$   $Y$ : text (sequence)
- **Translation**
  - $X$ : Mandarin sentence (sequence)  $\rightarrow$   $Y$ : English sentence (sequence)
- **Syntactic Paring**
  - $X$ : sentence  $\rightarrow$   $Y$ : parsing tree (tree structure)
- **Object Detection**
  - $X$ : Image  $\rightarrow$   $Y$ : bounding box
- **Summarization**
  - $X$ : long document  $\rightarrow$   $Y$ : summary (short paragraph)
- **Retrieval**
  - $X$ : keyword  $\rightarrow$   $Y$ : search result (a list of webpage)

# Unified Framework

## Training

- Find a function  $F$

$$F: X \times Y \rightarrow \mathbb{R}$$

- $F(x,y)$ : evaluate how compatible the objects  $x$  and  $y$  is

## Inference (Testing)

- Given an object  $x$

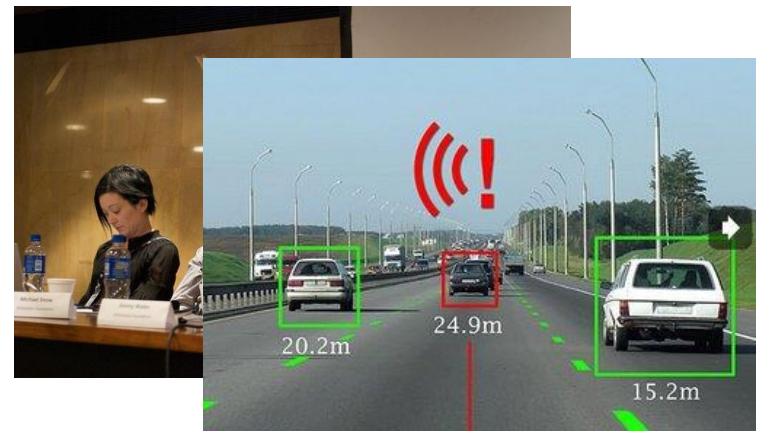
$$\tilde{y} = \arg \max_{y \in Y} F(x, y)$$

$$f: X \rightarrow Y \quad \Rightarrow \quad f(x) = \tilde{y} = \arg \max_{y \in Y} F(x, y)$$

# Unified Framework – Object Detection

- Task description

- Using a bounding box to highlight the position of a certain object in an image
- E.g. A detector of Haruhi



$X$  : Image  $\longrightarrow$   $Y$  : Bounding Box



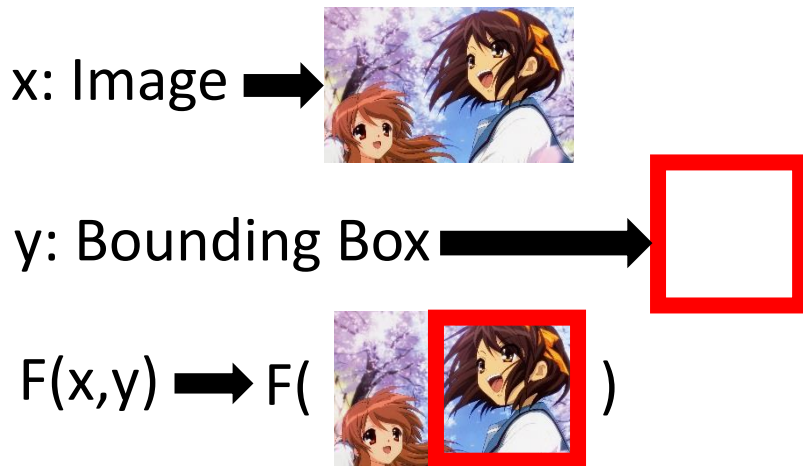
**Haruhi**

(the girl with  
yellow ribbon)

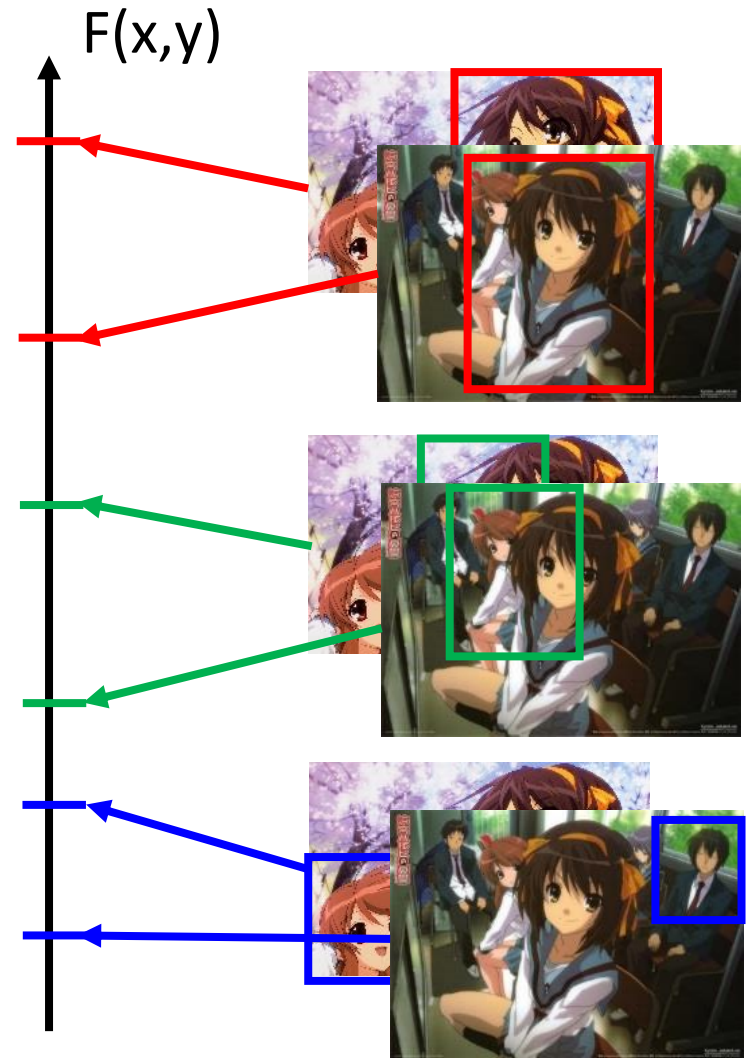
# Unified Framework – Object Detection

## Training

- Find a function  $F$   
$$F: X \times Y \rightarrow \mathbb{R}$$
- $F(x,y)$ : evaluate how compatible the objects  $x$  and  $y$  is



the correctness of taking range of  $y$  in  $x$  as “Haruhi”



# Unified Framework – Object Detection

## Training

- Find a function  $F$   
$$F: X \times Y \rightarrow \mathbb{R}$$
- $F(x,y)$ : evaluate how compatible the objects  $x$  and  $y$  is

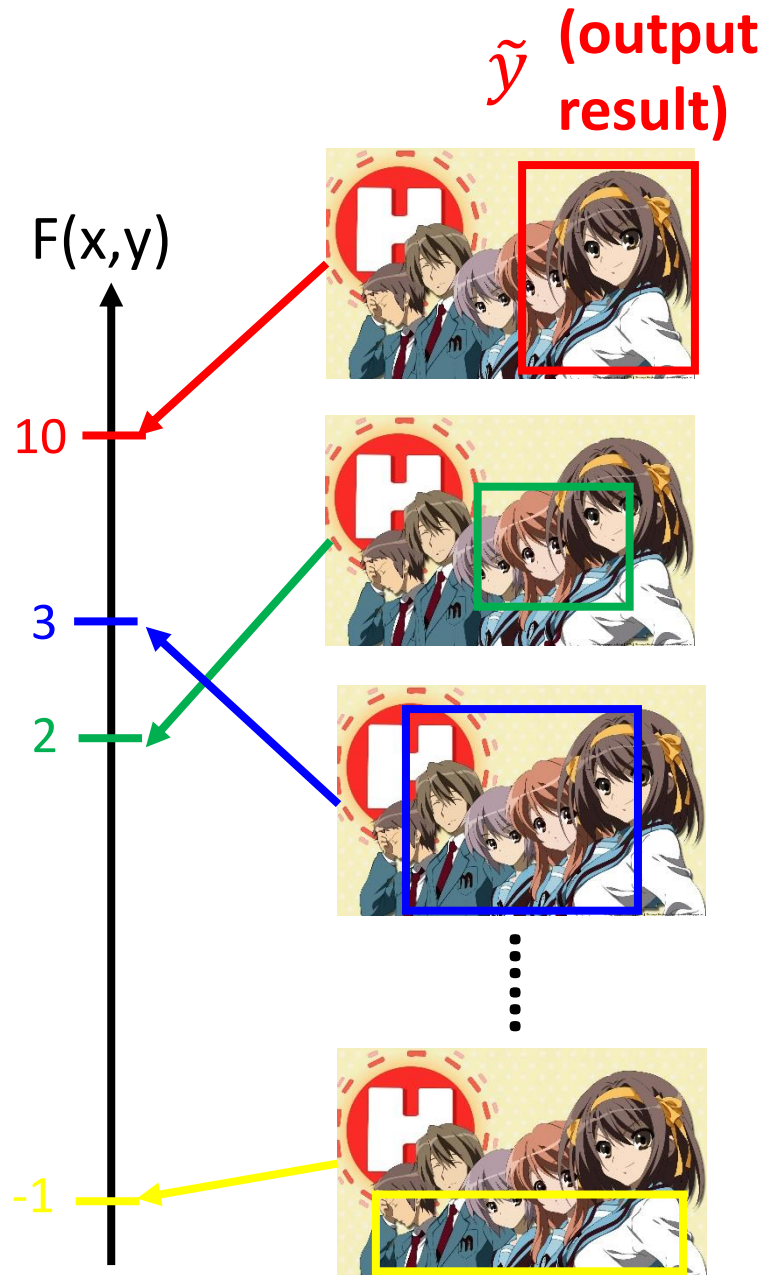
## Inference (Testing)

- Given an object  $x$   
$$\tilde{y} = \arg \max_{y \in Y} F(x, y)$$

input  $x =$



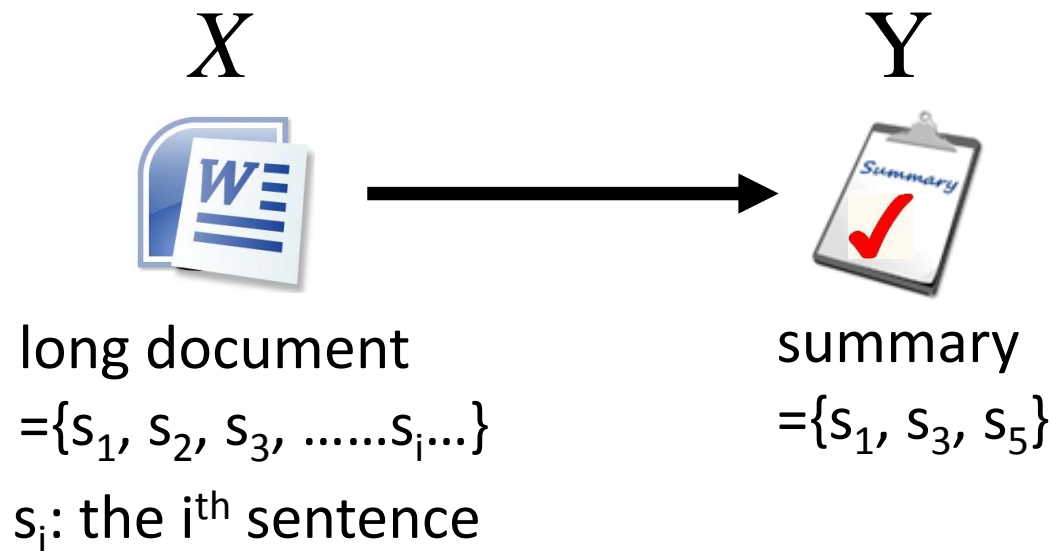
Enumerate all possible  
bounding box  $y$



# Unified Framework

## - Summarization

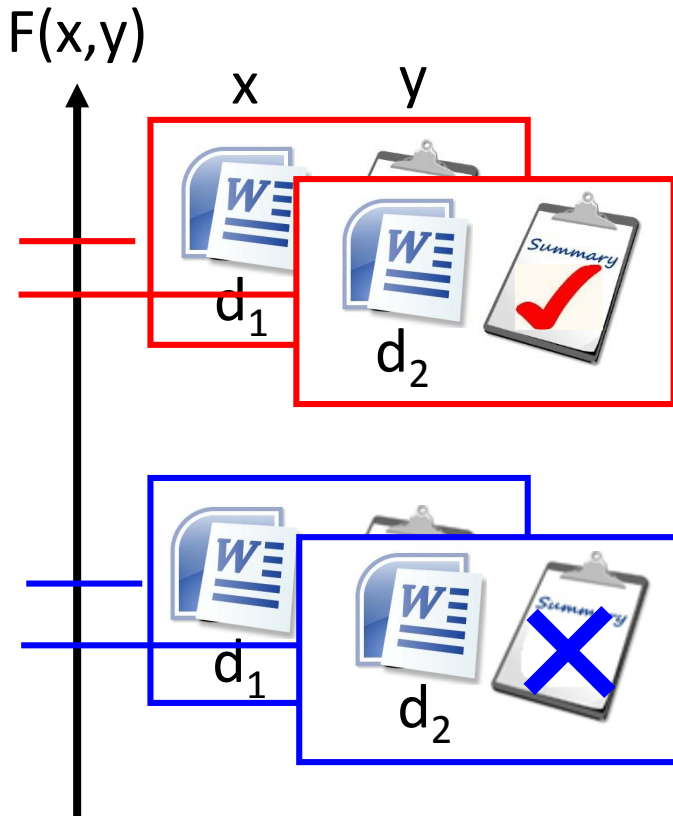
- Task description
  - Given a long document
  - Select a set of sentences from the document, and cascade the sentences to form a short paragraph



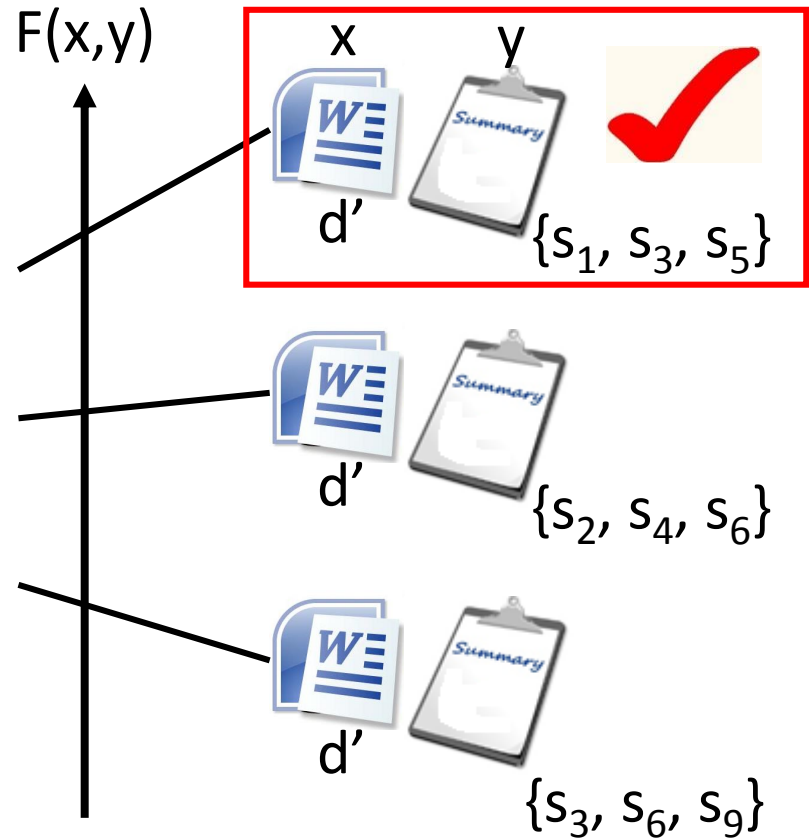


# Unified Framework - Summarization

Training



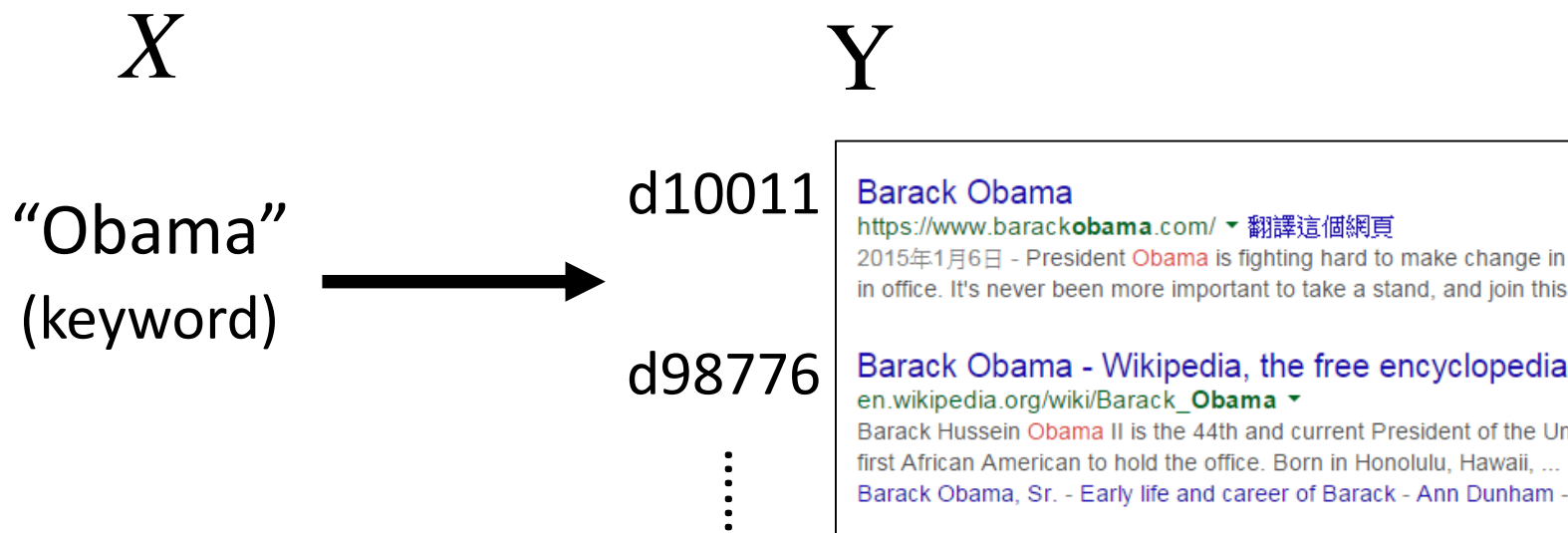
Inference



# Unified Framework

## - Retrieval

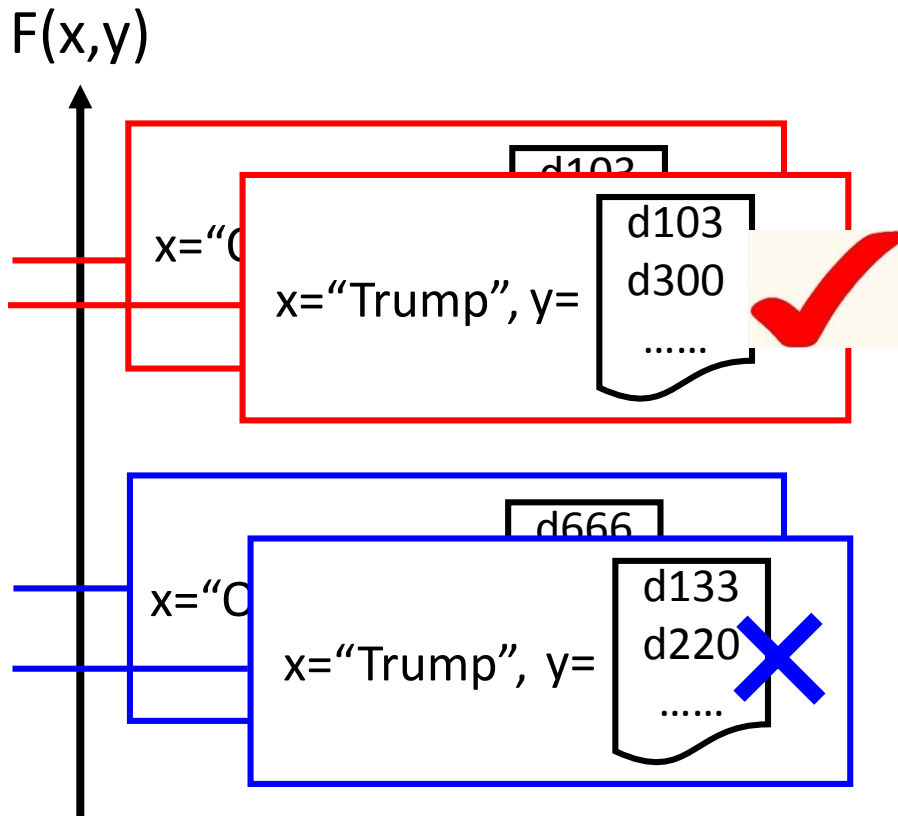
- Task description
  - User input a keyword  $Q$
  - System returns a *list* of web pages



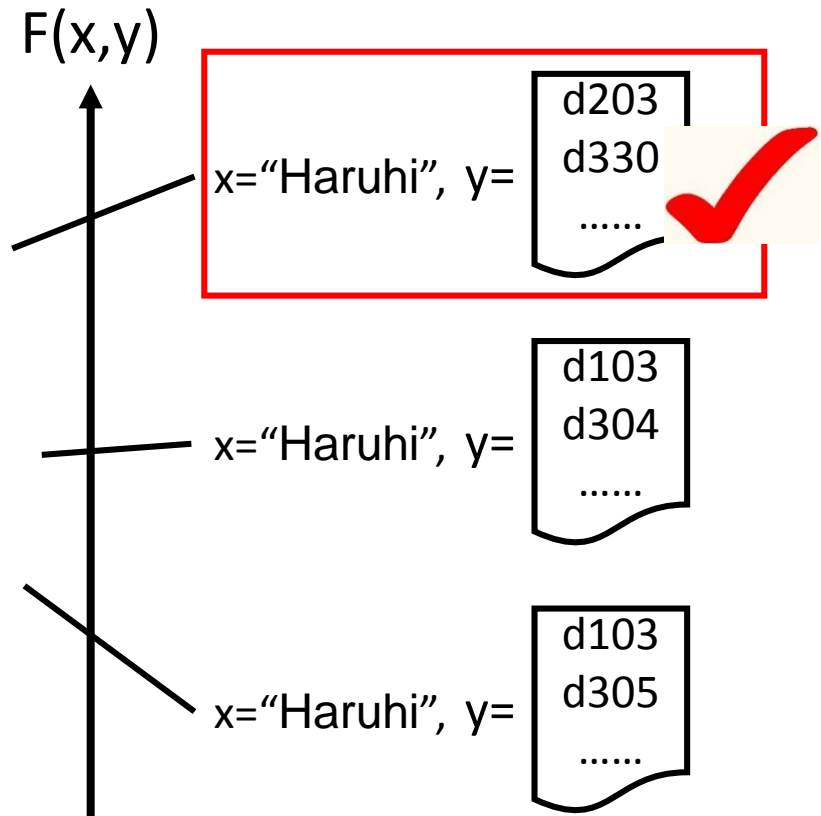
A list of web pages (Search Result)

# Unified Framework - Retrieval

Training



Inference



# Statistics

## Unified Framework

### Training

- Find a function  $F$

$$F: X \times Y \rightarrow \mathbb{R}$$

- $F(x,y)$ : evaluate how compatible the objects  $x$  and  $y$  is

### Inference

- Given an object  $x$

$$\tilde{y} = \arg \max_{y \in Y} F(x, y)$$

$$F(x, y) = P(x, y)?$$

### Training

- Estimate the probability  $P(x,y)$

$$P: X \times Y \rightarrow [0,1]$$

### Inference

- Given an object  $x$

$$\tilde{y} = \arg \max_{y \in Y} P(y | x)$$

$$= \arg \max_{y \in Y} \frac{P(x, y)}{P(x)}$$

$$= \arg \max_{y \in Y} P(x, y)$$

# Statistics

## Unified Framework

$$F(x, y) = P(x, y)?$$

### Drawback for probability

- Probability cannot explain everything
- 0-1 constraint is not necessary

### Strength for probability

- Meaningful

Energy-based Model:  
<http://www.cs.nyu.edu/~yann/research/ebm/>

## Training

- Estimate the probability  $P(x, y)$

$$P: X \times Y \rightarrow [0, 1]$$

## Inference

- Given an object  $x$

$$\tilde{y} = \arg \max_{y \in Y} P(y | x)$$

$$= \arg \max_{y \in Y} \frac{P(x, y)}{P(x)}$$

$$= \arg \max_{y \in Y} P(x, y)$$

# Unified Framework

That's it!?

## Training

- Find a function  $F$

$$F: X \times Y \rightarrow \mathbb{R}$$

- $F(x,y)$ : evaluate how compatible the objects  $x$  and  $y$  is

## Inference (Testing)


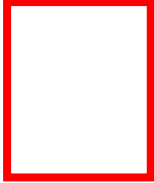
- Given an object  $x$

$$\tilde{y} = \arg \max_{y \in Y} F(x, y)$$


There are three problems in this framework.

# Problem 1

- **Evaluation:** What does  $F(x,y)$  look like?
  - How  $F(x,y)$  compute the “compatibility” of objects  $x$  and  $y$

Object Detection:  $F(x=$   ,  $y=$   )

Summarization:  $F(x=$   ,  $y=$   )  
(a long document) (a short paragraph)

Retrieval:  $F(x=$  “Obama” (keyword) ,  $y=$   )  
(Search Result)

# Problem 2

- **Inference:** How to solve the “arg max” problem

$$y = \arg \max_{y \in Y} F(x, y)$$

The space  $Y$  can be extremely large!

**Object Detection:**  $Y$ =All possible bounding box (maybe tractable)

**Summarization:**  $Y$ =All combination of sentence set in a document ...

**Retrieval:**  $Y$ =All possible webpage ranking ....



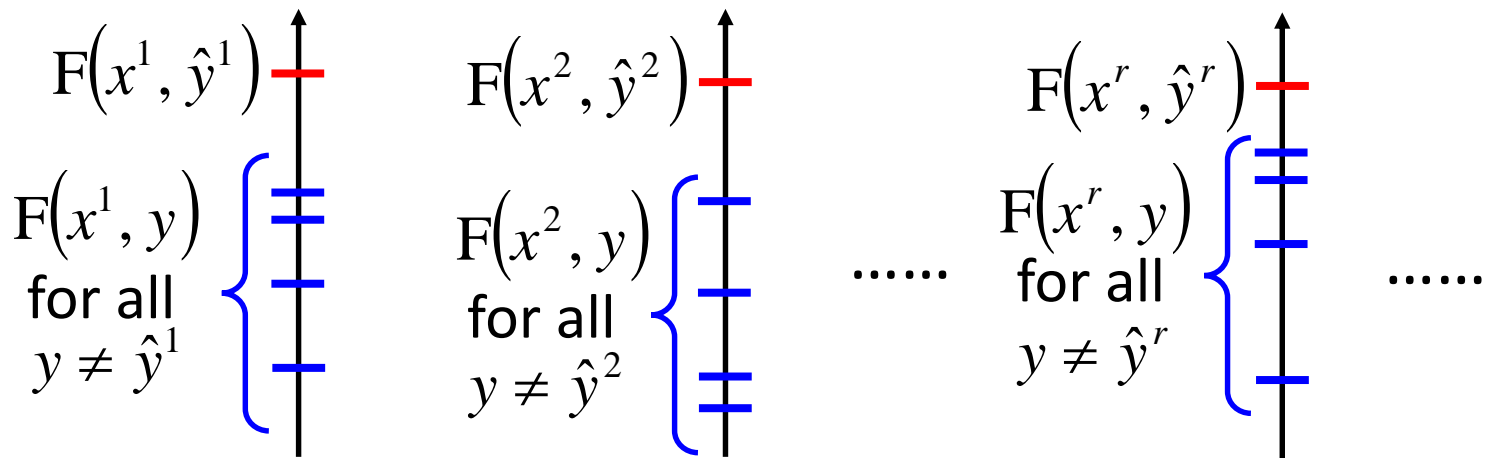
# Problem 3

- **Training**: Given training data, how to find  $F(x,y)$

## Principle

Training data:  $\{(x^1, \hat{y}^1), (x^2, \hat{y}^2), \dots, (x^r, \hat{y}^r), \dots\}$

We should find  $F(x,y)$  such that .....



# Three Problems

## Problem 1: Evaluation

- What does  $F(x,y)$  look like?



## Problem 2: Inference

- How to solve the “arg max” problem

$$y = \arg \max_{y \in Y} F(x, y)$$



## Problem 3: Training

- Given training data, how to find  $F(x,y)$



## Three Problems

### Problem 1: Evaluation

- What does  $F(x,y)$  look like?

### Problem 2: Inference

- How to solve the “arg max” problem?

$$y = \arg \max_x F(x,y)$$

### Problem 3: Training

- Given training data, how to find the best model?

Have you heard the three problems elsewhere?

## Hidden Markov Model

### • Three Basic Problems for HMMs

Given an observation sequence  $\bar{O}=(o_1,o_2,\dots,o_T)$ , and an HMM

$\lambda=(A,B,\pi)$

– Problem 1 :

How to *efficiently* compute  $P(\bar{O}|\lambda)$  ?

⇒ *Evaluation problem*

– Problem 2 :

How to choose an optimal state sequence  $\mathbf{q}=(q_1,q_2,\dots,q_T)$  ?

⇒ *Decoding Problem*

– Problem 3 :

Given some observations  $\bar{O}$  for the HMM  $\lambda$  , how to adjust the model parameter  $\lambda=(A,B,\pi)$  to maximize  $P(\bar{O}|\lambda)$ ?

⇒ *Learning /Training Problem*

From 數位語音處理

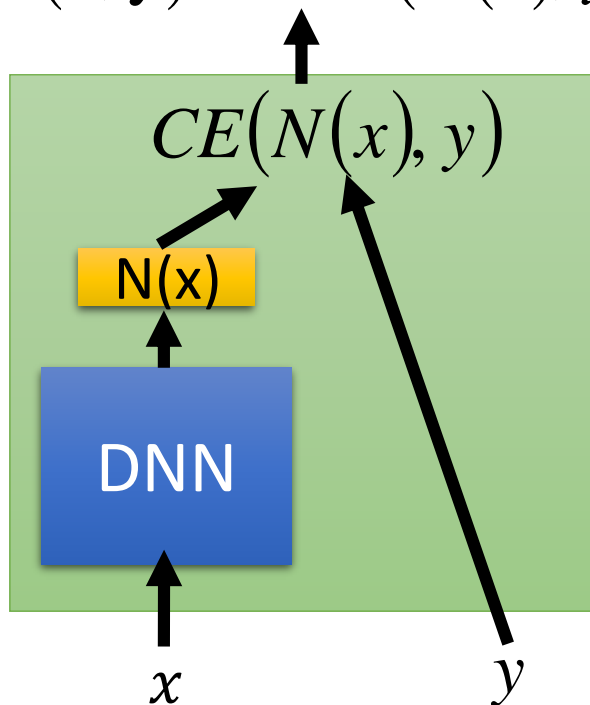
# Link to DNN?

The same as what we have learned.

Training

$$F: X \times Y \rightarrow \mathbb{R}$$

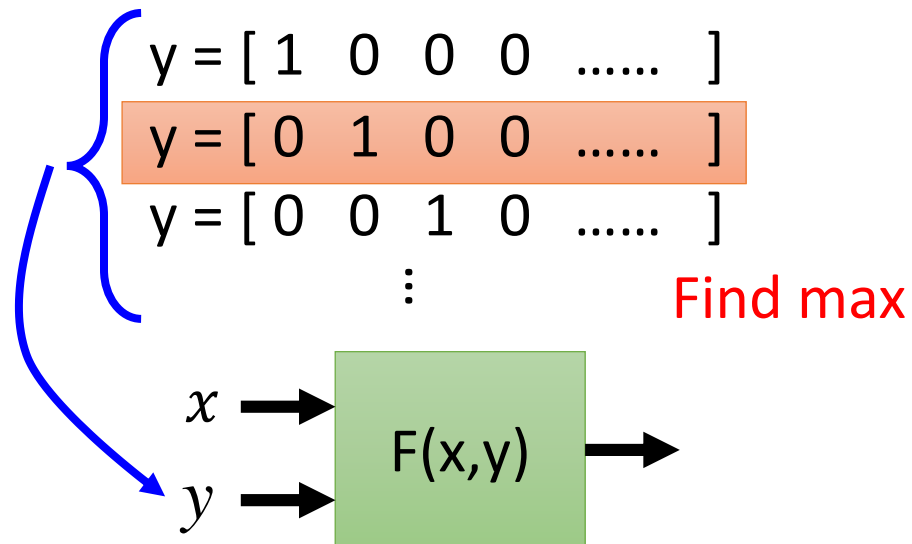
$$F(x, y) = -CE(N(x), y)$$



Inference

$$\tilde{y} = \arg \max_{y \in Y} F(x, y)$$

In handwriting digit classification, there are only 10 possible  $y$ .



# You have to know .....

- **Viterbi Algorithm**

- 數位語音處理:

- [http://speech.ee.ntu.edu.tw/DSP2015Autumn/Videos/20150930\\_4.0.fsp.wmv/index.html](http://speech.ee.ntu.edu.tw/DSP2015Autumn/Videos/20150930_4.0.fsp.wmv/index.html) (請用 IE 開啟)

- [http://speech.ee.ntu.edu.tw/DSP2015Autumn/Videos/20151007\\_4.0.fsp.wmv/index.html](http://speech.ee.ntu.edu.tw/DSP2015Autumn/Videos/20151007_4.0.fsp.wmv/index.html) (請用 IE 開啟)

- 演算法

- 數位通信相關課程