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

Unified Framework

Training

- Find a function F

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

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

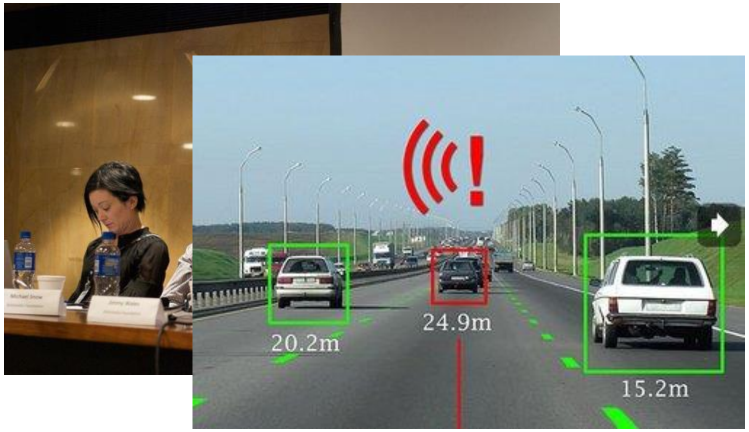
Inference (Testing)

- Given an object x

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

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

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 → Y : Bounding Box

Haruhi

(the girl with yellow ribbon)

Unified Framework – Object Detection

Training

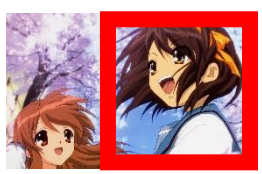
- Find a function F

$$F: X \times Y \rightarrow R$$

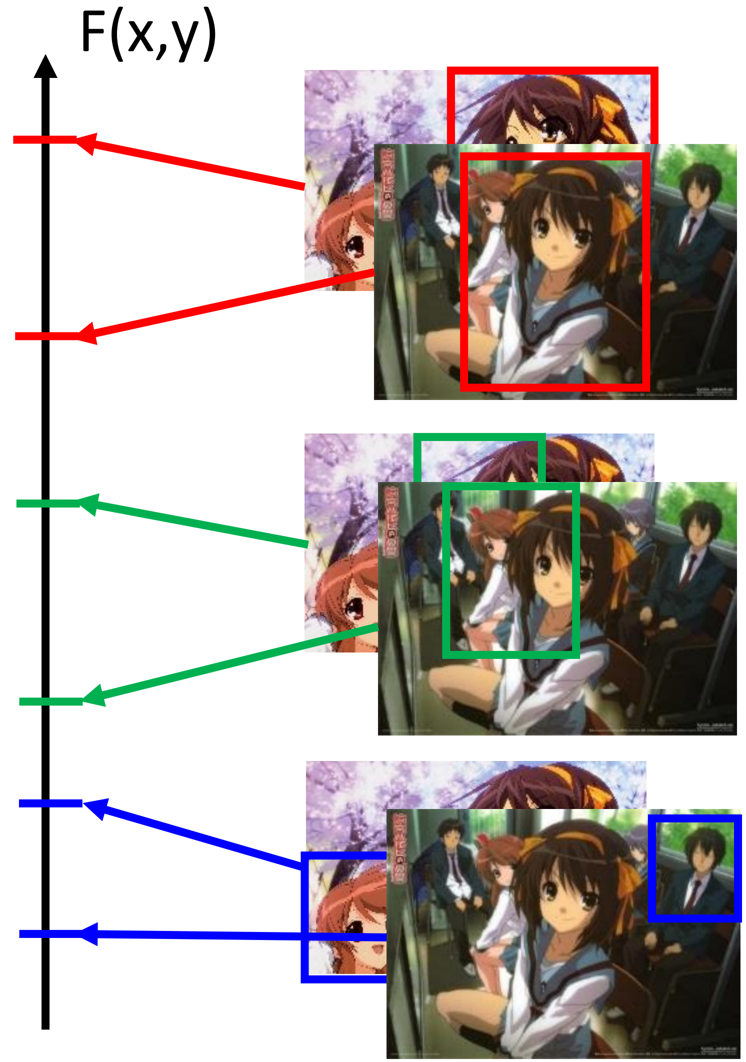
- F(x,y): evaluate how compatible the objects x and y is

x: Image ➡ 

y: Bounding Box ➡

F(x,y) ➡ F()

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



Unified Framework – Object Detection

Training

- Find a function F

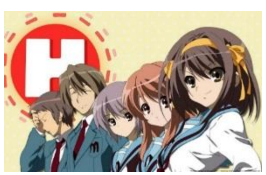
$$F : X \times Y \rightarrow R$$

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

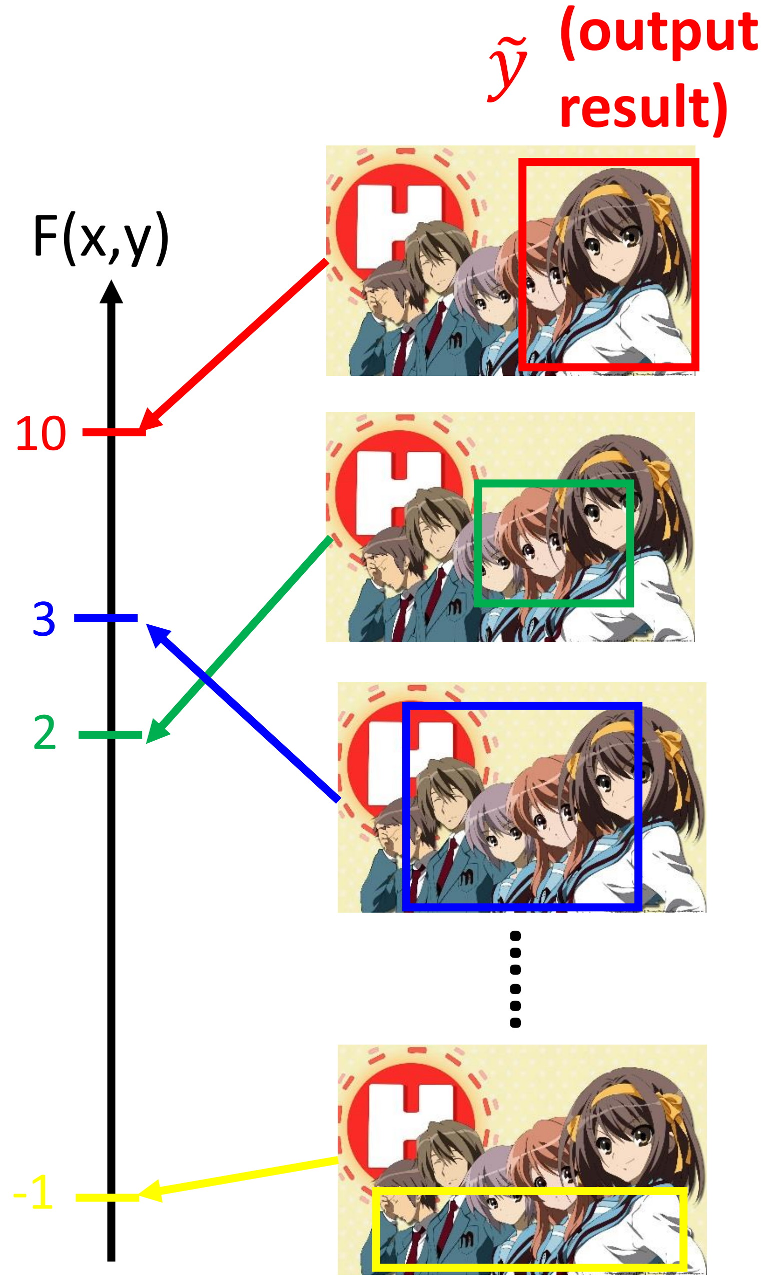
Inference (Testing)

- 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

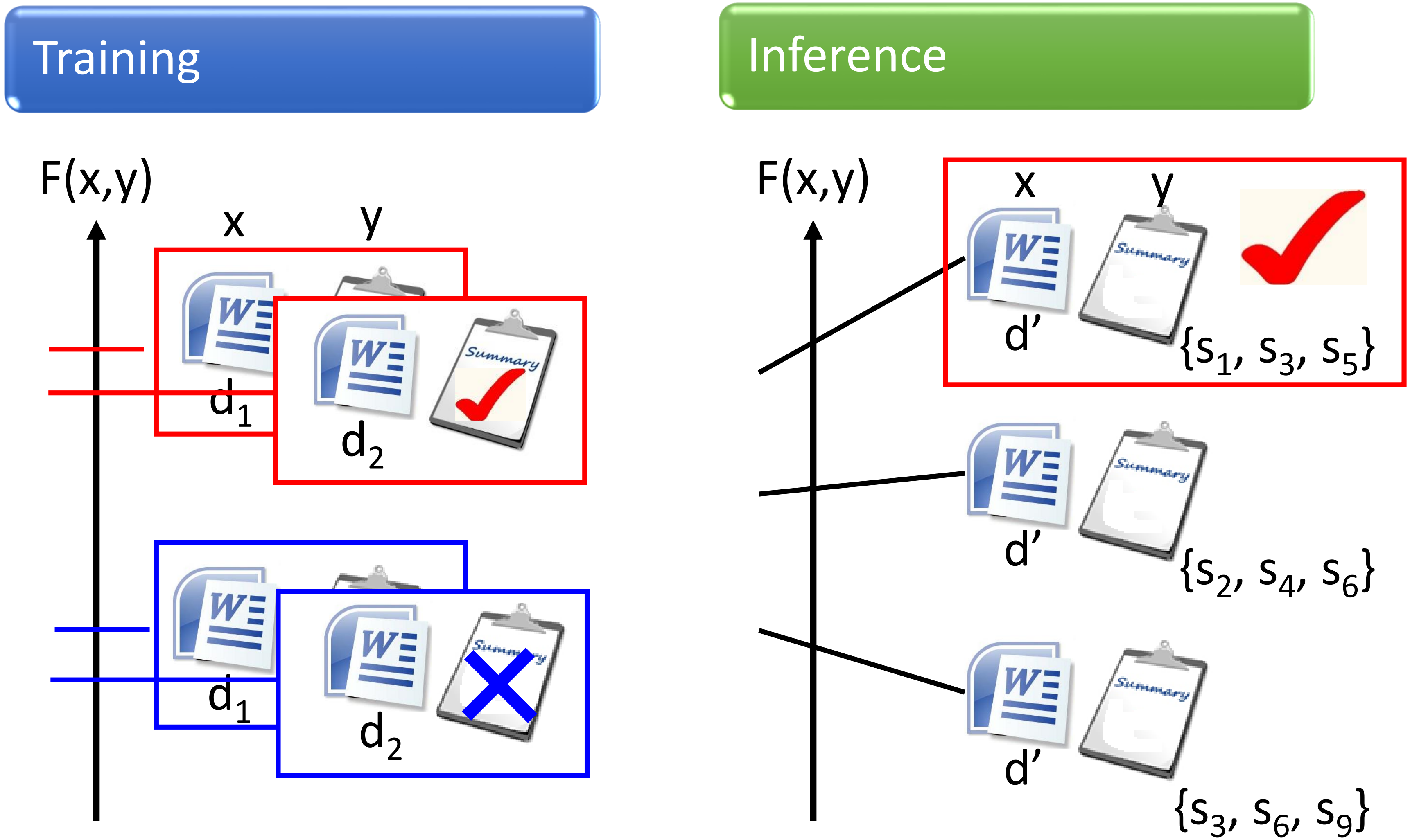
- Task description
  - Given a long document
  - Select a set of sentences from the document, and cascade the sentences to form a short paragraph

X → Y

long document
$=\{s_1, s_2, s_3, \ldots\ldots s_i \ldots\}$
s_i: the i^{th} sentence

summary
$=\{s_1, s_3, s_5\}$

Unified Framework
- Summarization



Unified Framework
- Retrieval

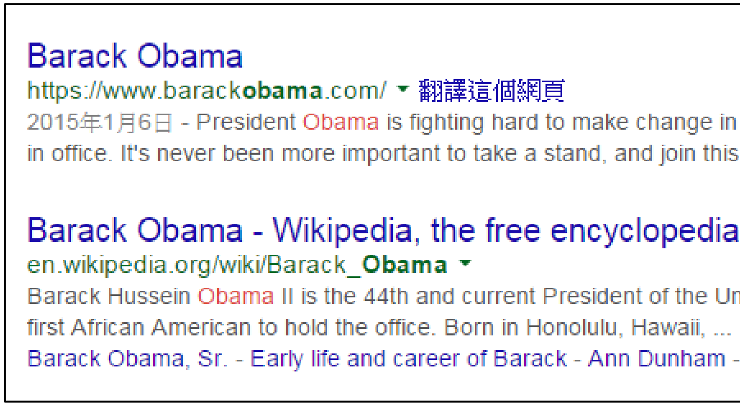
- Task description
  - User input a keyword Q
  - System returns a ***list*** of web pages

X $\quad\quad$ Y

“Obama”
(keyword)

→

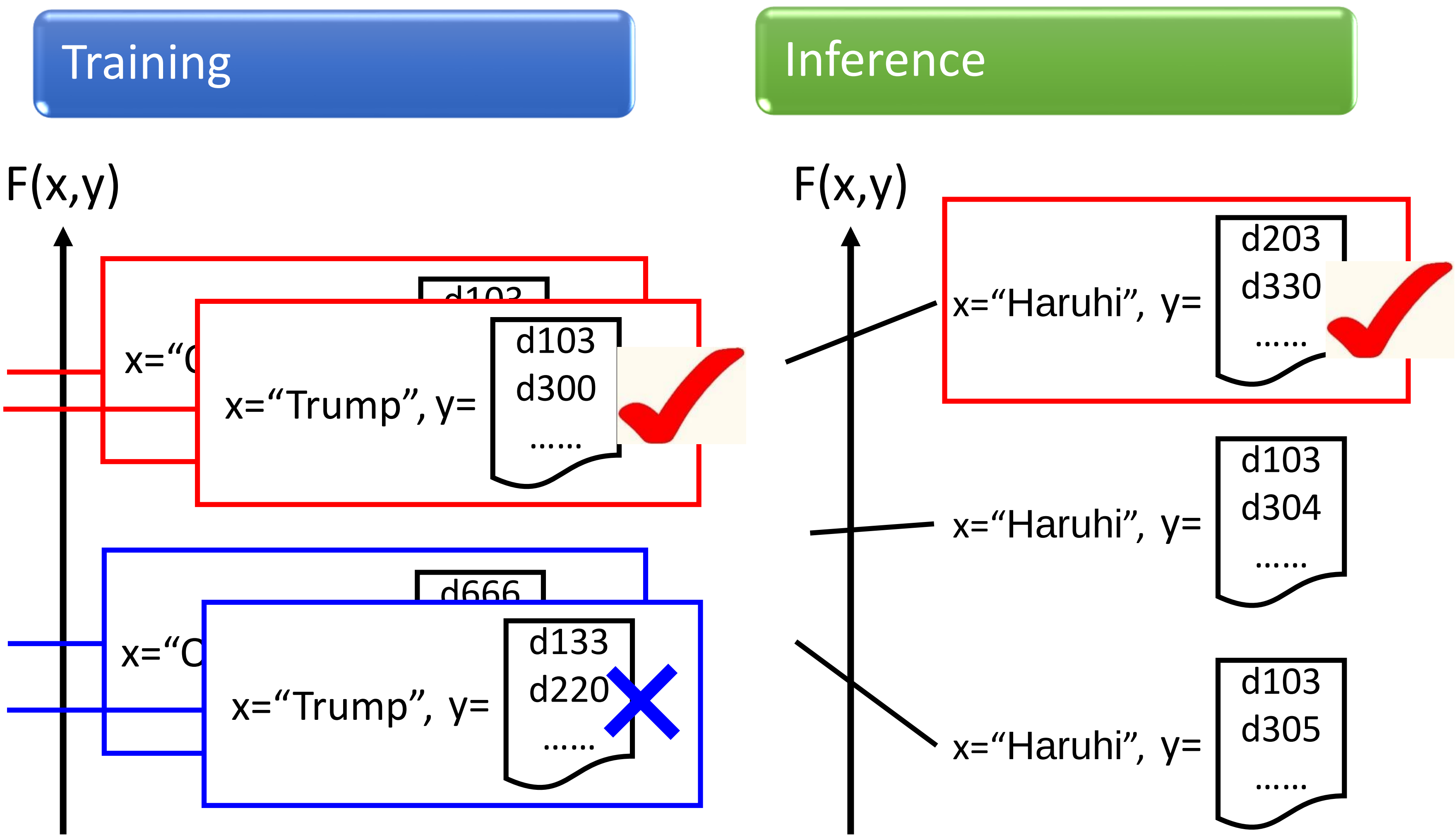
d10011
d98776
⋮


Barack Obama
https://www.barackobama.com/ ▾ 翻譯這個網頁
2015年1月6日 - President Obama is fighting hard to make change in
in office. It's never been more important to take a stand, and join this

Barack Obama - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Barack_Obama ▾
Barack Hussein Obama II is the 44th and current President of the Un
first African American to hold the office. Born in Honolulu, Hawaii, ...
Barack Obama, Sr. - Early life and career of Barack - Ann Dunham -

A list of web pages (Search Result)

Unified Framework - Retrieval



Statistics

Unified Framework

Training

- Find a function F

$$F : X \times Y \to R$$

- 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)

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

Inference

- Given an object x

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

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

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

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)

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

Inference

- Given an object x

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

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

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

Unified Framework

That's it!?

Training

- Find a function F

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

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

Inference (Testing)

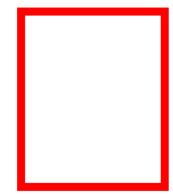
- Given an object x

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

There are three problems in this framework.

Problem 1

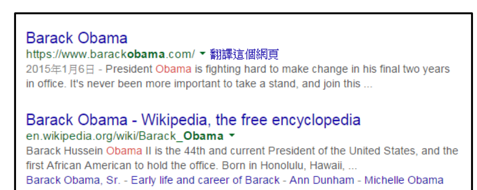
- ***Evaluation***: What does F(x,y) look like?
  - How F(x,y) compute the “compatibility” of objects x and y

Object Detection: F(x=  , y= )

Summarization: F(x=  , y= 

)

(a long document) (a short paragraph)

Retrieval: F(x= “Obama” , y= 
Barack Obama
https://www.barackobama.com/ 翻譯這個網頁
2015年1月6日 - President Obama is fighting hard to make change in his final two years in office. It's never been more important to take a stand, and join this ...

Barack Obama - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Barack_Obama
Barack Hussein Obama II is the 44th and current President of the United States, and the first African American to hold the office. Born in Honolulu, Hawaii, ...
Barack Obama, Sr. - Early life and career of Barack - Ann Dunham - Michelle Obama

)

(keyword) (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 ….

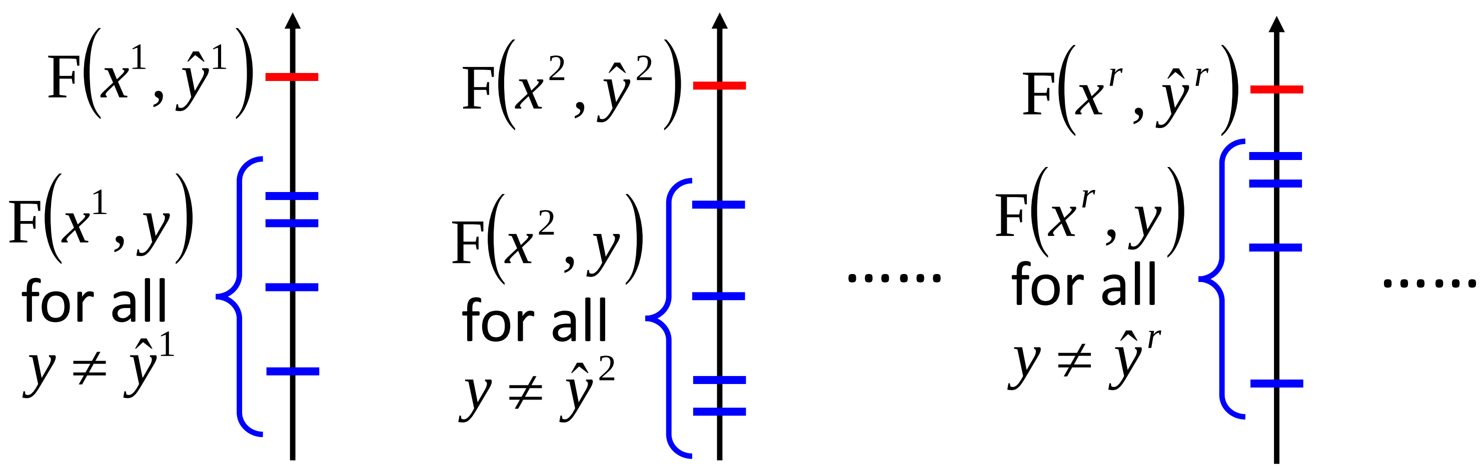
Problem 3

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

Principle

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

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

$$y = \arg m$$

Problem 3: Training

- Given training data, how t

Have you heard the three problems elsewhere?

Hidden Markov Model

- **Three Basic Problems for HMMs**

Given an observation sequence $\overline{O}=(o_1,o_2,\ldots,o_T)$, and an HMM λ =(A,B,π)

- Problem *1* :
  How to *efficiently* compute P($\overline{\mathbf{O}}$| λ) ?
  ⇨ *Evaluation problem*
- Problem *2* :
  How to choose an optimal state sequence **q**=($q_1,q_2,\ldots\ldots, q_T$) ?
  ⇨ *Decoding Problem*
- Problem *3* :
  Given some observations $\overline{O}$ for the HMM λ , how to adjust the model parameter λ =(A,B,π) to maximize P($\overline{\mathbf{O}}$| λ)?
  ⇨ *Learning /Training Problem*

From 數位語音處理

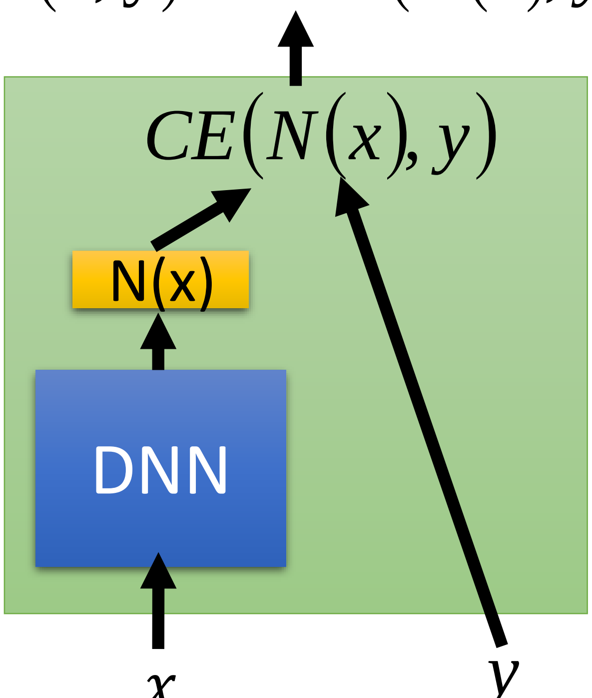
Link to DNN?

The same as what we have learned.

Training

$$F : X \times Y \rightarrow R$$

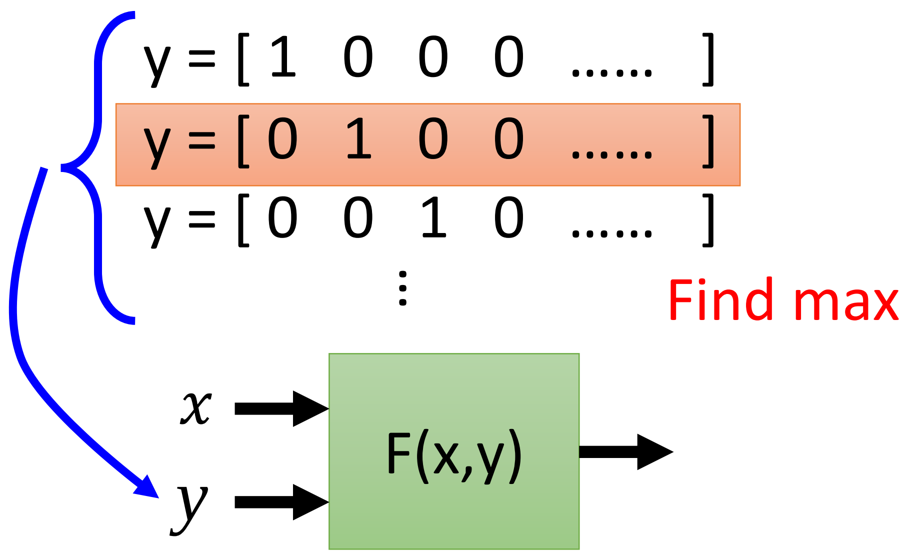
$$F(x, y) = -CE(N(x), y)$$



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 (請用 IE 開啟)
    - 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