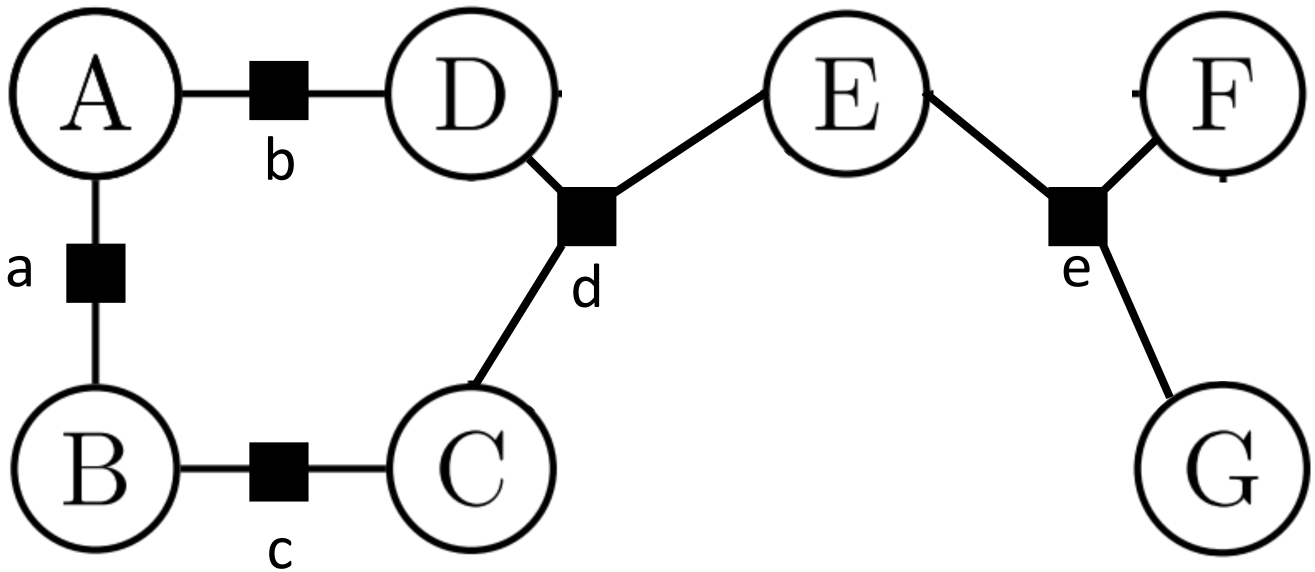
MRF

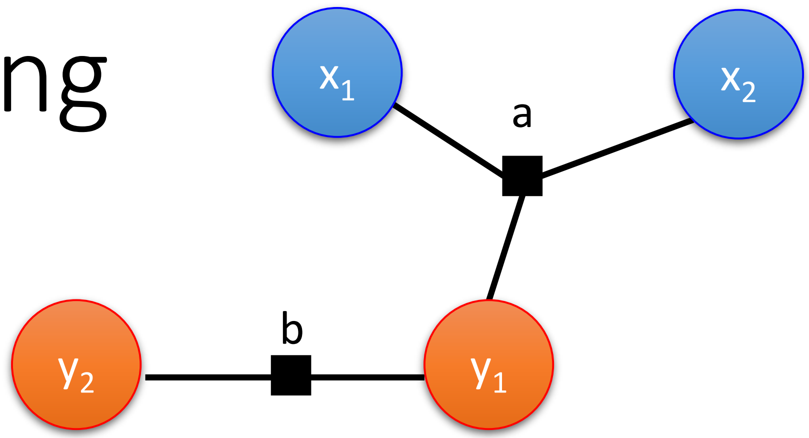
MRF → ***Factor Graph***



Evaluation Function

$$f_a(A,B) + f_b\,(A,D) + f_c(B,C) + f_d(C,D,E) + f_e(E,F,G)$$

Training



$$F(x,y) = f_a(x_1,x_2,y_1) + f_b(y_1,y_2)$$

$$= w_a \cdot \phi_a(x_1,x_2,y_1) + w_b \cdot \phi_b(y_1,y_2)$$

$$= \begin{bmatrix} w_a \\ w_b \end{bmatrix} \begin{bmatrix} \phi_a(x_1,x_2,y_1) \\ \phi_b(y_1,y_2) \end{bmatrix}$$

組合起來

$$= w \cdot \phi(x,y)$$

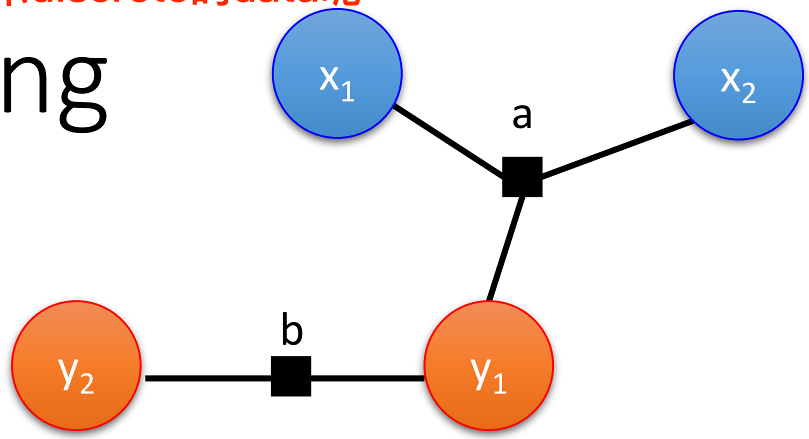
Simply training by
structured perceptron
or structured SVM

Max-Margin Markov Networks (M3N)

用structure perceptron來train的方法即為上

如何用linear來表示discrete的data呢

Training



$$F(x, y) = f_a(x_1, x_2, y_1) + f_b(y_1, y_2)$$

$$= w_a \cdot \phi_a(x_1, x_2, y_1) + w_b \cdot \phi_b(y_1, y_2)$$

$$y_1, y_2 \epsilon \{+1, -1\}$$

y_1	y_2	$f_b(y_1, y_2)$
+1	+1	w_1
+1	-1	w_2
-1	+1	w_3
-1	-1	w_4

窮舉

$$w_b = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}$$

$$\phi_b(+1, +1) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

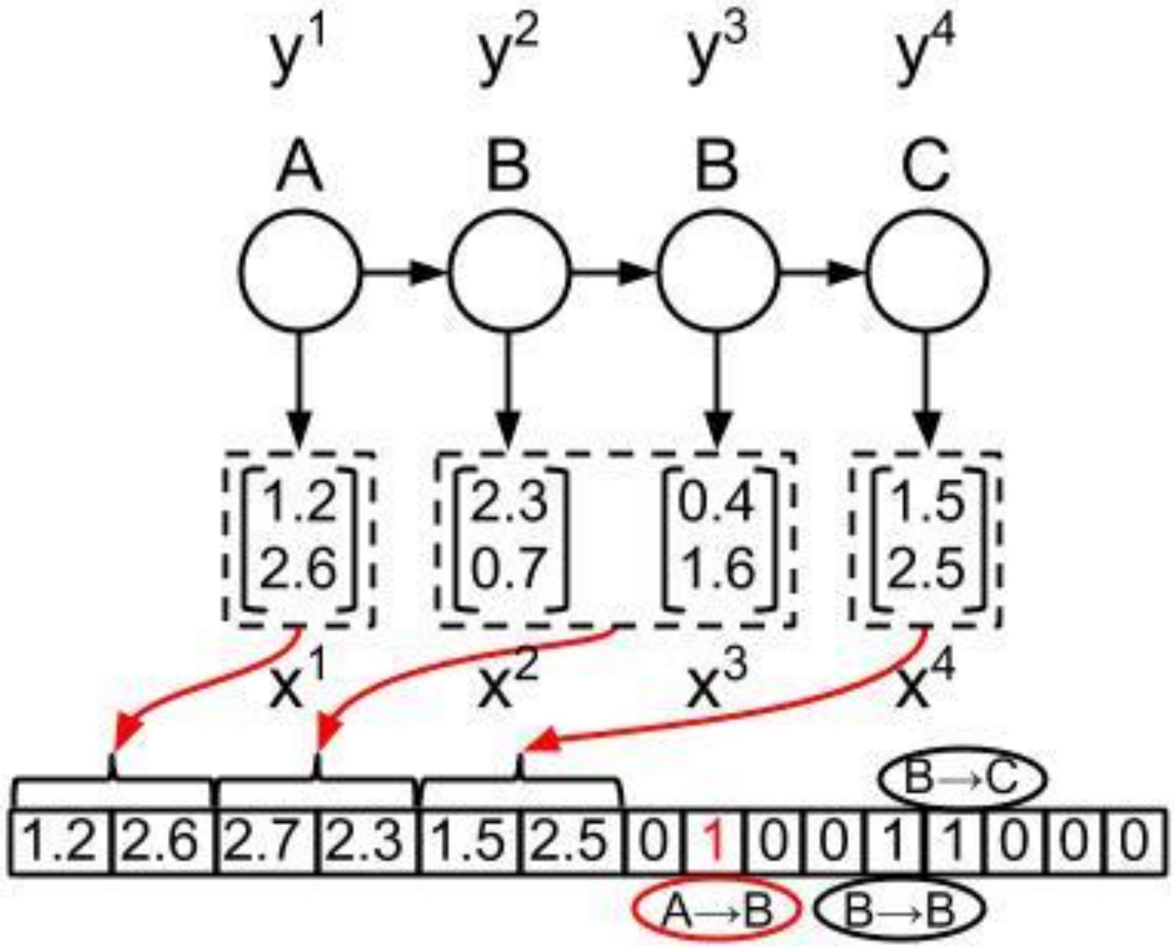
$$\phi_b(+1, -1) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\phi_b(-1, +1) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\phi_b(-1, -1) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

窮舉

Now can you interpret this?



$\phi(x, y)$

Probability Point of View

- $F(x, y)$ can be any real number
- If you like probability 轉成一個機率

Between 0 and 1

$$P(x, y) = \frac{e^{F(x,y)}}{\sum_{x',y'} e^{F(x',y')}}$$

To be positive

normalization

$$P(y|x) = \frac{P(x, y)}{P(x)}$$

$$= \frac{P(x, y)}{\sum_{y''} P(x, y'')} = \frac{\frac{e^{F(x,y)}}{\cancel{\sum_{x',y'} e^{F(x',y')}}}}{\boxed{\sum_{y''}} \frac{e^{F(x,y'')}}{\cancel{\sum_{x',y'} e^{F(x',y')}}}} = \frac{e^{F(x,y)}}{\sum_{y''} e^{F(x,y'')}}$$

summation所有的y

posterior probability

Gibbs Sampling

Inference for the dumb

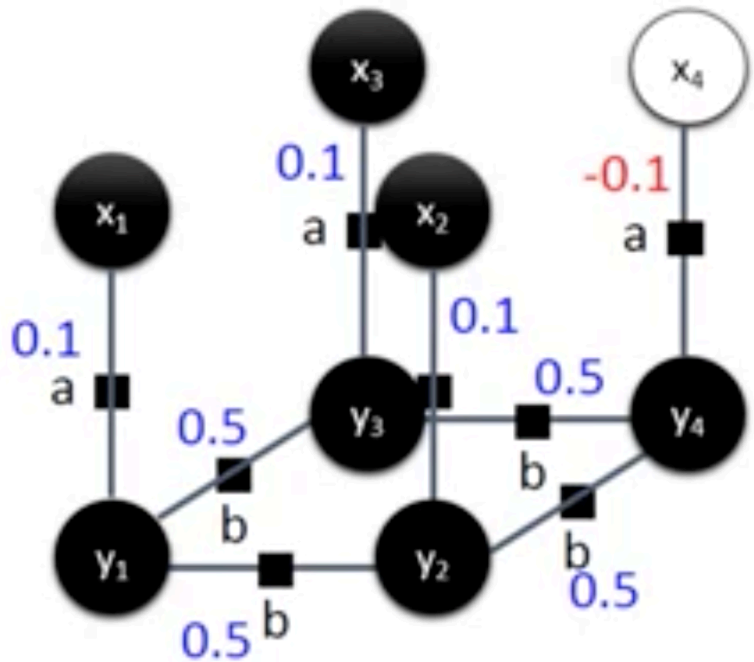
窮舉所有的y找出最大的evaluation function output，最後及輸出這組structure

$$f_a(x_i, y_i) = \begin{cases} 0.1 & x_i = y_i \\ -0.1 & x_i \neq y_i \end{cases}$$

$$f_b(y_i, y_j) = \begin{cases} 0.5 & y_i = y_j \\ -0.5 & y_i \neq y_j \end{cases}$$

Given input noisy image x

$$x_1, x_2, x_3, x_4 = -1, -1, -1, 1$$



Inference:

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

$$y_1, y_2, y_3, y_4 = -1, -1, -1, -1$$

$$\Rightarrow F(x, y) = 2.2 \quad \text{max}$$

$$y_1, y_2, y_3, y_4 = 1,1,1,1$$

$$\Rightarrow F(x, y) = 1.8$$

$$y_1, y_2, y_3, y_4 = -1,1,1,-1$$

$$\Rightarrow F(x, y) = -2.2$$

⋮

Enumerate all possible **y**

Design an efficient algorithm to do that is not always easy.

但是在現實的問題中y所有的可能組合往往是很大的，無法窮舉，因此我們可以使用Gibbs sampling

Sampling?

Probability point of view:

$$P(x,y) = \frac{e^{F(x,y)}}{\sum_{x',y'} e^{F(x',y')}}$$

$$\boxed{P(y|x) = \frac{e^{F(x,y)}}{\underline{\sum_{y''} e^{F(x,y'')}}} \quad \text{Independent of y}} \propto F(x,y)$$

$$\tilde{y} = \arg\max_y \mathrm{F}(x,y) \quad = \quad \tilde{y} = \arg\max_y P(y|x)$$

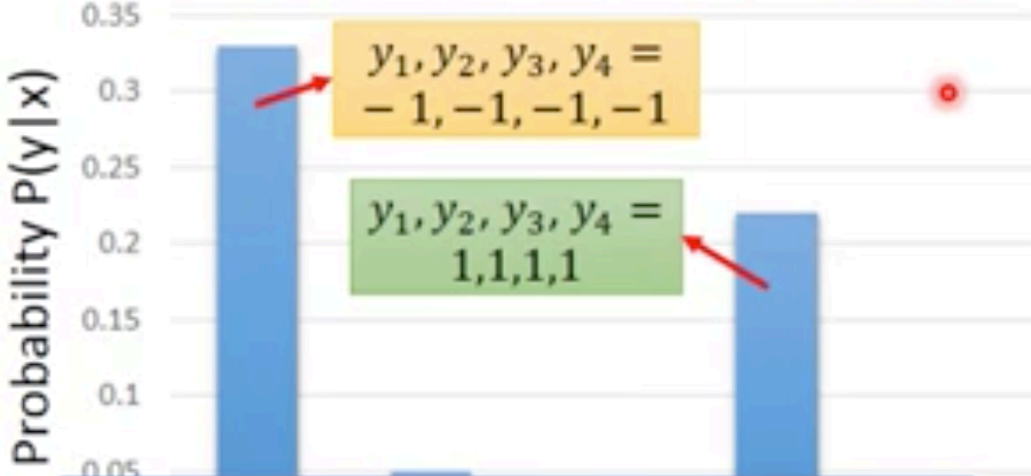
從distribution中sample一個值，找出sample中出現機率最高的作為max

$$P(y|x) = \frac{e^{F(x,y)}}{\sum_{y''} e^{F(x,y'')}}$$

Sampling?

- $P(y|x)$ is a distribution

Given $x_1, x_2, x_3, x_4 = -1, -1, -1, 1$



> It is hard to know the distribution.

> If we know the distribution, why bother with the sampling?

Sample from the distribution ……

$-1, -1, -1, -1$

$1,1,1,1$

$-1, -1, -1, 1$

$-1, -1, -1, -1$

$1,1,1,1$

$-1, -1, -1, -1$

……

$-1, -1, -1, -1$

Max probability

Inference result

然而既然知道distribution的話為何不直接找出max?

不合理！因此Gibbs Sampling就是可以在不知道distribution的情況下找出max值

Gibbs Sampling

- There is a probability distribution $P(\mathbf{y}|\mathbf{x})$
  - $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$

y太難窮舉

- We want to sample from $P(\mathbf{y}|\mathbf{x})$, but it is too complex to do that

假設這是可以成立的

- However, $P(y_i|y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_N,\mathbf{x})$ can be computed
- We can sample from $P(\mathbf{y}|\mathbf{x})$ by Gibbs sampling