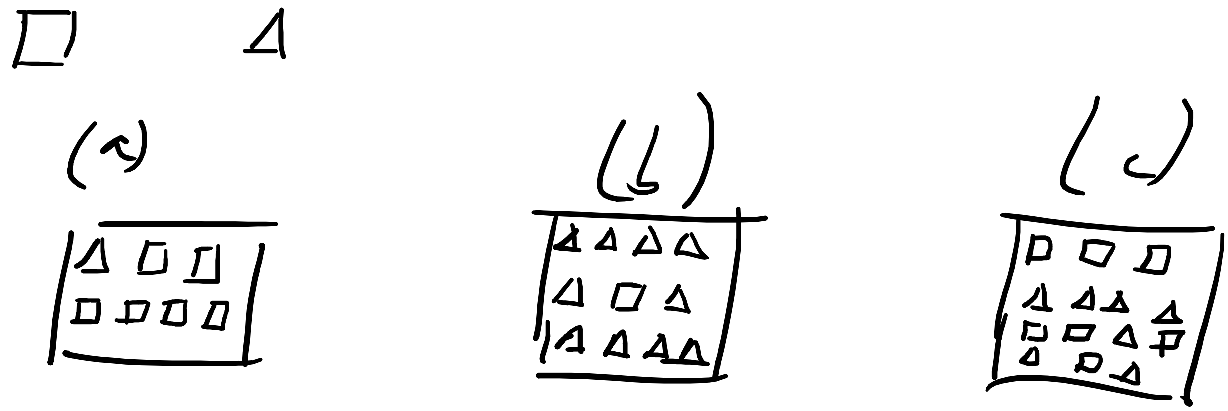
→ **Entropy**: Degree of disorder or uncertainity in a system.

↳ Basis of something called mutual information.

↳ Quantifies the relationship b/w 2 things.



→ **Surprise**: Surprise is high when rare items probability is low.

C —1→ (H) $\log\left(\frac{1}{p(H)}\right) = \log\left(\frac{1}{1}\right) = \underline{0}$

C —0→ (T) $\log\left(\frac{1}{p(T)}\right) = \log\left(\frac{1}{0}\right) = \log(1) - \log(0)$

$= \text{undefined}.$

→ Surprise $= \log\left(\frac{1}{p}\right)$

0.9 — (H) $\log\left(\frac{1}{p(H)}\right) = \log\left(\frac{1}{0.9}\right) = 0.152$

$$C \xrightarrow{0.9} H \qquad \log\left(\frac{1}{p(H)}\right) = \log\left(\frac{1}{0.9}\right) = 0.152$$

$$C \xrightarrow{0.1} T \qquad \log\left(\frac{1}{p(T)}\right) = \log\left(\frac{1}{0.1}\right) = 3.32$$

$\rightarrow$ H H T

$$0.9 \times 0.9 \times 0.1$$

$$\text{Surprise} = \log\left(\frac{1}{0.9 \times 0.9 \times 0.1}\right)$$

$$= \log(1) - \log(0.9 \times 0.9 \times 0.1)$$

$$= \log(1) - [\log(0.9) + \log(0.9) + \log(0.1)]$$

$$= 0 - \log(0.9) - \log(0.9) - \log(0.1)$$

$$= 0.15 + 0.15 + 3.32 = 3.62$$

H H T

$$0.152 + 0.152 + 3.32 = \underline{\underline{3.62}}$$

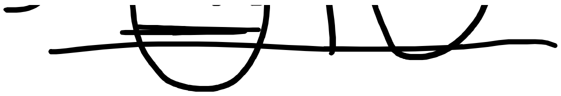
T T H T

$$3.32 + 3.32 + 0.152 + 3.32$$

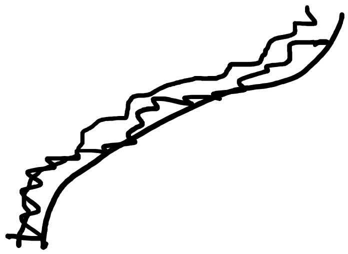
$\rightarrow$

	H	T
P	0.9	0.1
S	0.152	3.32

$$E(\text{Surprise}) = (0.9 \times 0.152) + (0.1 \times 3.32)$$



$$(0.1 \times 3.32)$$



$$= \underline{\underline{0.45}}$$

$$E(\text{Surprise}) = \sum_x P(X = x)$$

Specific value for surprise

The probability of observing that specific value for surprise

$$E = \sum \underbrace{\log\left(\frac{1}{p(x)}\right)}_{\text{Surprise}} \times \underbrace{p(x)}_{\text{Probability of surprise}}$$

$$E = \sum p(x) \cdot \log\left(\frac{1}{p(x)}\right)$$

$$= \sum p(x)\left[\log(1) - \log(p(x))\right]$$

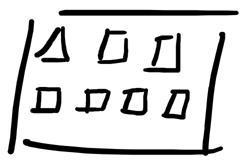
$$= \sum p(x)\left[0 - \log(p(x))\right]$$

$$= \sum -p(x) \cdot \log(p(x))$$

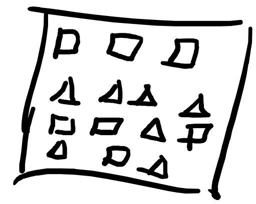
$$= \sum -p(x) \cdot \log p(x)$$

$$\boxed{\therefore\ E = -\sum p(x) \cdot \log p(x)}$$

(a)



(c)

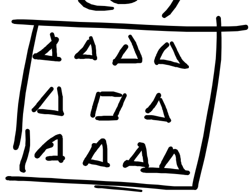


A) $E = \sum p(x) \cdot \log\left(\frac{1}{p(x)}\right)$

$$= \frac{6}{7} \cdot \log\left(\frac{1}{\frac{6}{7}}\right) + \frac{1}{7} \cdot \log\left(\frac{1}{\frac{1}{7}}\right)$$

$$= \underline{\underline{0.58}}$$

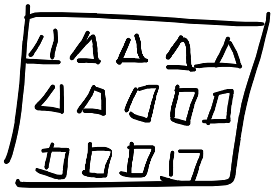
(b)



B) $E = \sum p(x) \cdot \log\left(\frac{1}{p(x)}\right)$

$$= \frac{1}{11} \times \log\left(\frac{1}{1/11}\right) + \frac{10}{11} \times \log\left(\frac{1}{10/11}\right)$$

$$= \underline{\underline{0.439}}$$

c) 

c) $E = \sum p_x \cdot \log\left(\frac{1}{p(x)}\right)$

$= 7/14 \times \log\left(\frac{1}{7/14}\right) + 7/14 \times \log\left(\frac{1}{7/14}\right)$

$= \underline{\underline{1}}$

; 0.58, 0.432, 1

Maximum Entropy Model.

$= \arg\max\left(\sum p(x) \cdot \log\left(\frac{1}{p(x)}\right)\right)$

$\arg\max(E)$

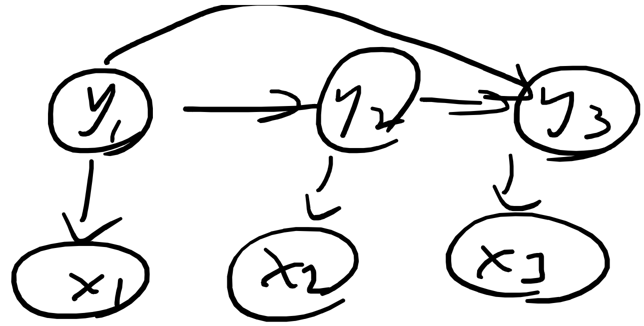
→ MEM : Discriminative Model.

→ HMM ; Undetectable Flaws;

Generative : Hidden States.

Hidden :

Observed :



$$P(y, x) = \prod P(y_i \mid y_{i-1}) P(x_i \mid y_i)$$

-> Limitations :

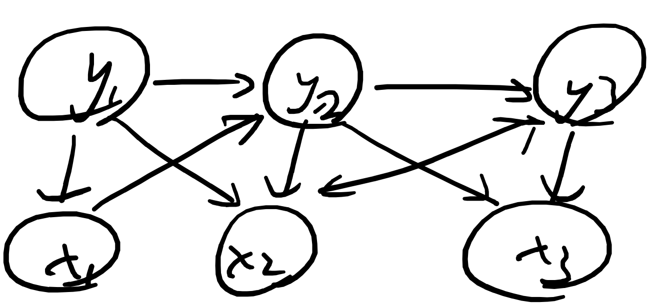
↳ Static Transmission & Emission

↳ Limited Dependencies

-> Conditional Random Field (CRF) :

Goal : Model $P(y|x)$
Conditional Probability (Discriminative)

->



-> Linear Chain CRF :

$X \; f(x, y_{i-1}, y_i, i) \; y_i$

$$X \; \underbrace{f(x, y_{i-1}, y_i, i)}_{\text{Feature Function}} \; y_i$$

X - Observed Data $\quad$ y_{i-1} - Previous Hidden Data

y_i - Current Hidden State $\quad$ i - Index (Timestamp of current state)

-> Bruce Wayne lives in Gotham city

Linear Chain CRF:

Let $F(X, y_{i-1}, y_i, i) = \sum w_j \underbrace{f_j(x, y_{i-1}, y_i, i)}_{j^{th} \text{ feature function}}$

-> $p(y|x) = \frac{1}{Z} e^{\gamma}$ $\quad$ Feature function Sum

Normalizing Parameter

$$\gamma = \sum F(x, y_{i-1}, y_i, i)$$

-> **Formal Definition:**

→ Formal Definition:

$$P(y|x) = \frac{1}{Z} \prod \exp(\phi_k(x,y))$$

Normalizer

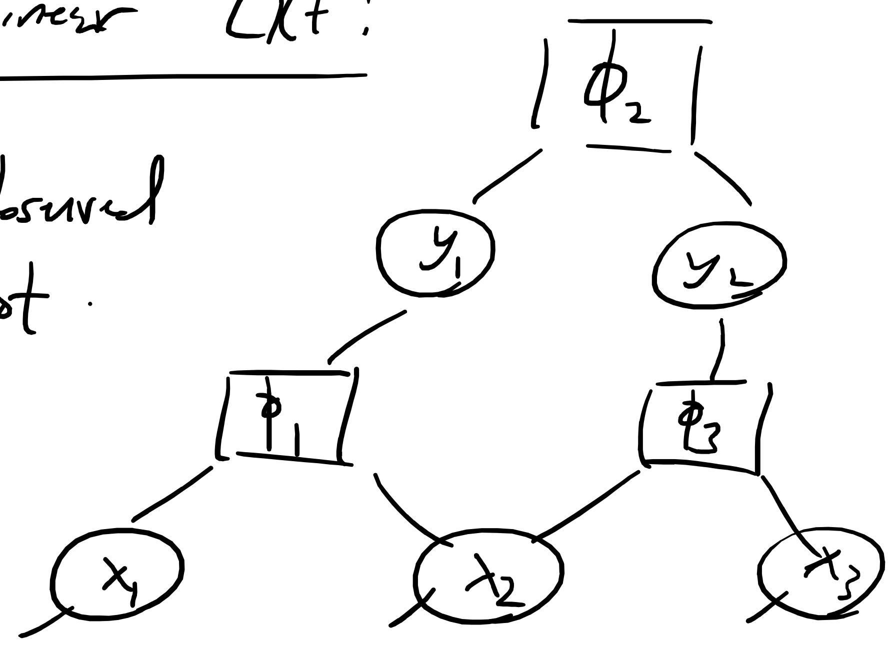
Any real valued scoring function of an argument.

→ $\phi_k(x,y) = w^T f_k(x,y)$ ~ Logistic Regression

→ $$P(y|x) = \frac{1}{Z} \exp\left(\sum_{k=1}^{n} w^T f_k(x,y)\right)$$

→ Log-Linear CRF:

→ x is observed
y is not.



→ Problem: Intractable Inference:

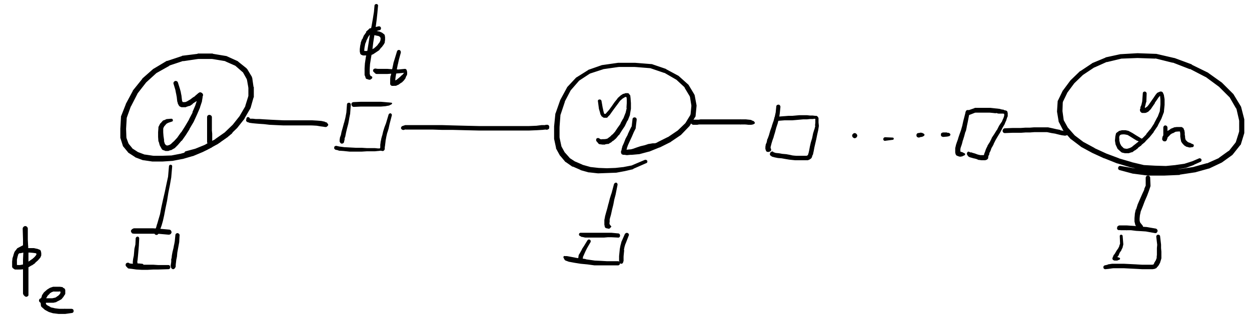
$$Z = \sum_{y'} \exp\left(\sum_{k=1}^{n} w^T f_k(x, y')\right)$$

-> Sequential CRF (Expressed w/ potentials ϕ)

$$\rightarrow p(y|x) = \frac{1}{Z} \prod_{i=2}^{n} \exp\left(\phi_t(y_{i-1}, y_i)\right) \prod_{i=1}^{n} \exp\left(\phi_e(y_{i-1}, x)\right)$$

-> Transitions ϕ_t

Emissions ϕ_e

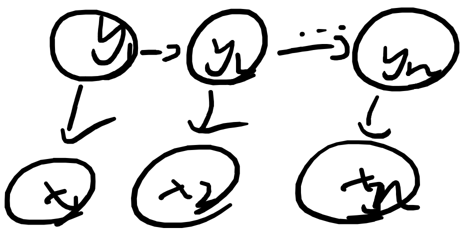


-> Special Case:

$$P(y|x) = \frac{1}{Z} \exp w^T \left[\sum_{i=2}^{n} f_t(y_{i-1}, y_i) + \sum_{i=1}^{n} f_e(y_i, i, x)\right]$$

-> Structure uses dynamic programming (Viterbi)

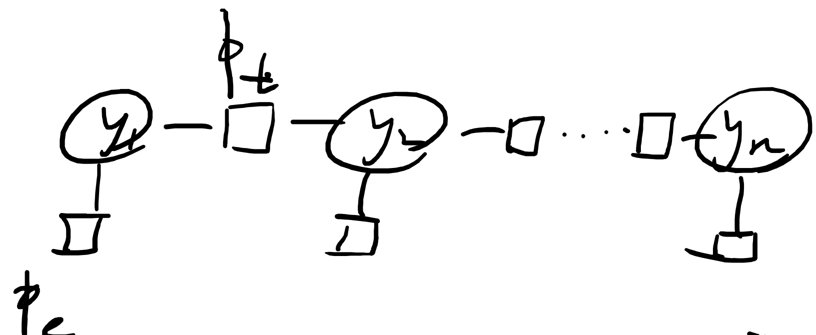
→ Structure uses dynamic programming (Viterbi) to sum or max over all sequences.



$$P(y,x) = P(y_1)P(x_1|y_1) \cdots\cdots$$

$$= P(y_1)\prod_{i=2}^{n} P(y_i|y_{i-1}) \prod_{i=1}^{n} P(x_i|y_i)$$

→



$$P(y|x) = \frac{1}{Z}\prod_{i=2}^{n} \exp\left(\phi_t(y_{i-1}, y_i)\right)\prod_{i=1}^{n} \exp\left(\phi_e(y_i, i, x)\right)$$

Naive Bayes	: Logistic Regression	:: HMM	: CRF
Local	Global	Local	Global
G	D	G	D

→ Locally normalized discriminative models do exist (MEMM).