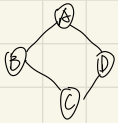
· Markov Network

teach made a misconception that may propagate in class

A and C never speak to each other, B and D never speak to each other.

$A = \begin{cases} a^1 & \text{A has misconception} \\ a^0 & \text{B has misconception} \end{cases}$



instead of CPD, the prob of one variable given another.

we need a more symmetric parameterization

eg. affinity, how likely 2 people are to agree with each other

factor is a function $\phi: Val(D) \mapsto \mathbb{R}$ where D is the domain of ϕ

$\phi_1(A,B)$		
a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_2(B,C)$		
b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

$\phi_3(C,D)$		
c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_4(D,A)$		
d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

Assignment				Unnormalized	Normalized
a^0	b^0	c^0	d^0	300,000	0.04
a^0	b^0	c^0	d^1	300,000	0.04
a^0	b^0	c^1	d^0	300,000	0.04
a^0	b^0	c^1	d^1	30	$4.1 \cdot 10^{-6}$
a^0	b^1	c^0	d^0	500	$6.9 \cdot 10^{-5}$
a^0	b^1	c^0	d^1	500	$6.9 \cdot 10^{-5}$
a^0	b^1	c^1	d^0	5,000,000	0.69
a^0	b^1	c^1	d^1	500	$6.9 \cdot 10^{-5}$
a^1	b^0	c^0	d^0	100	$1.4 \cdot 10^{-5}$
a^1	b^0	c^0	d^1	1,000,000	0.14
a^1	b^0	c^1	d^0	100	$1.4 \cdot 10^{-5}$
a^1	b^0	c^1	d^1	100	$1.4 \cdot 10^{-5}$
a^1	b^1	c^0	d^0	10	$1.4 \cdot 10^{-6}$
a^1	b^1	c^0	d^1	100,000	0.014
a^1	b^1	c^1	d^0	100,000	0.014
a^1	b^1	c^1	d^1	100,000	0.014

local interactions between directly related variables

$$p(a,b,c,d) = \frac{1}{Z}\, \phi_1(a,b) \cdot \phi_2(b,c) \cdot \phi_3(c,d) \cdot \phi_4(d,a)$$

$$Z = \sum_{a,b,c,d} \phi_1(a,b) \cdot \phi_2(b,c) \cdot \phi_3(c,d) \cdot \phi_4(d,a)$$ is the partition function

$(X \perp Y \mid Z)$ iff $P(\cdots) = \phi_1(X,Z)\, \phi_2(Y,Z)$

$$P(A,B,C,D) = \left[\frac{1}{Z}\, \phi_1(A,B)\, \phi_2(B,C)\right] \phi_3(C,D) \cdot \phi_4(A,D)$$

$$= f_1(B,A,C) \cdot f_2(D,A,C)$$

$\therefore (B \perp D \mid A, C)$

similarly $(A \perp C \mid B, D)$

$(X \perp Y \mid Z) \Longleftrightarrow P(X,Y,Z) = \phi_1(X,Z) \cdot \phi_2(Y,Z)$

right to left: $P(X,Z) = \sum_y P(X,y,Z) = \phi_1(X,Z) \cdot \sum_y \phi_2(y,Z)$

$P(Y,Z) = \sum_x P(x,Y,Z) = \phi_2(Y,Z) \sum_x \phi_1(x,Z)$

$P(Z) = \sum_x \sum_y P(x,y,Z) = \sum_x \sum_y \phi_1(x,Z)\, \phi_2(y,Z)$

$$P(X|Z) \cdot P(Y|Z) = \frac{P(X,Z) \cdot P(Y,Z)}{P(Z)^2} = \frac{\phi_1(X,Z) \cdot \phi_2(Y,Z) \cdot \left(\sum_x \phi_1(x,Z)\right)\left(\sum_y \phi_2(y,Z)\right)}{P(Z)^2}$$

$$= \frac{\phi_1(X,Z)\, \phi_2(Y,Z)}{P(Z)} = \frac{P(X,Y,Z)}{P(Z)} = P(X,Y|Z)$$

$\therefore P(X|Z)\, P(Y|Z) = P(X,Y|Z)$

$\therefore (X \perp Y \mid Z)$

left to right: $P(X,Y,Z) = P(Z) \cdot P(X|Z) \cdot P(Y|X,Z)$

$$= \underbrace{P(Z) \cdot P(X|Z)}_{\phi_1(X,Z)} \cdot \underbrace{P(Y|Z)}_{\phi_2(Y,Z)}$$

· Gibbs Distribution and Markov Network

A distribution P_Φ is a Gibbs distribution parameterized by a set of factors

$$\Phi = \{ \phi_1(D_1), \dots, \phi_k(D_k) \}$$

$$P_\Phi(X_1, \dots, X_n) = \frac{1}{Z} \tilde{P}_\Phi(X_1, \dots, X_n)$$

$$= \frac{1}{\sum_{X_1 \dots X_n} \phi_1(D_1) \times \dots \times \phi_n(D_n)} \phi_1(D_1) \times \dots \times \phi_n(D_n)$$

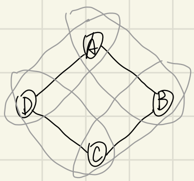
A factor is not a joint distribution, it's only one contribution to the overall joint distribution
The distribution as a whole has to take into consideration to contributions from all of the factors involved

A distribution P_Φ with $\Phi = \{\phi_1(D_1) \dots \phi_k(D_k)\}$ factorizes over a Markov network $\mathcal{H}$ if each D_k is a complete subgraph of $\mathcal{H}$

$\phi_1(A,B,D)$ $\phi_2(B,C,D)$
$\phi_1(A,B)$ $\phi_2(B,C)$ $\phi_3(C,D)$ $\phi_4(A,D)$ $\phi_5(B,D)$
$\phi_1(A,B,C)$ $\phi_2(B,C)$ $\phi_3(C,D)$

Cannot read the factorization from graph

The factors that parameterize a Markov network are often called clique potentials



4 cliques and the clique potentials for each of them

$\{A, B\}$

A	B	$\phi_1(A,B)$
a^0	b^0	⋮
a^0	b^1	⋮
⋮	⋮	⋮

$\{C, D\}$

C	D	$\phi_3(C,D)$
⋮	⋮	⋮

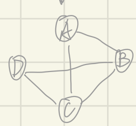
$\{B, C\}$

B	C	$\phi_2(B,C)$
⋮	⋮	⋮

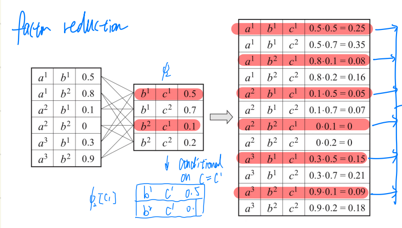
$\{A, D\}$

A	D	$\phi_4(A,D)$
⋮	⋮	⋮

$\phi_1(A,B,C)$ $\phi_2(B,C,D)$



· Reduced Markov Networks



a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

b^1	c^1	0.5
b^2	c^1	0.1

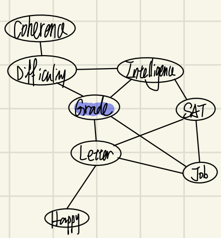
a^1	b^1	c^1	0.5·0.5 = 0.25
a^1	b^1	c^2	0.5·0.7 = 0.35
a^1	b^2	c^1	0.8·0.1 = 0.08
a^1	b^2	c^2	0.8·0.2 = 0.16
a^2	b^1	c^1	0.1·0.5 = 0.05
a^2	b^1	c^2	0.1·0.7 = 0.07
a^2	b^2	c^1	0·0.1 = 0
a^2	b^2	c^2	0·0.2 = 0
a^3	b^1	c^1	0.3·0.5 = 0.15
a^3	b^1	c^2	0.3·0.7 = 0.21
a^3	b^2	c^1	0.9·0.1 = 0.09
a^3	b^2	c^2	0.9·0.2 = 0.18

Figure 4.3 An example of factor product

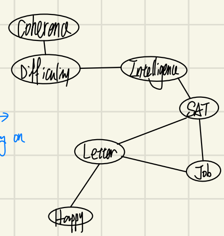
reduced Gibbs distribution defined by $\Phi[c_1] = \{ \phi_1[c_1], \phi_2[c_1] \dots \}$

$\mathcal{H}$ is a Markov network over X and U = u is a context.
the reduced Markov network $\mathcal{H}[u]$ is a Markov network over nodes X−U
(edges not associated with U stay the same

In a Markov network, conditioning on a variable eliminates the node and edges.
In a Bayesian network, conditioning on a variable can activate v-structures, creating new dependencies



conditioning on Grade



· Markov Independencies

In Markov networks, probabilistic influence flows along undirected paths in the graph, but it's blocked if we condition on the intervening nodes

A set of nodes $\mathbf{Z}$ separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{H}$, denoted $sep_{\mathcal{H}}(X;Y|Z)$
if there is no active path between and node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ given $\mathbf{Z}$

global independencies in $\mathcal{H}$ $\quad I(\mathcal{H}) = \{ (X \perp Y | Z) : sep_{\mathcal{H}}(X;Y|Z) \}$

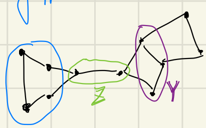
P is a Gibbs distribution that factorizes over $\mathcal{H}$ $\Longrightarrow$ $\mathcal{H}$ is an I-map for P

Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be any 3 disjoint subsets of all variables s.t. $sep_{\mathcal{H}}(\mathbf{X};\mathbf{Y}|\mathbf{Z})$

Z_X is the set of cliques that are contained in $\mathbf{X} \cup \mathbf{Z}$

Z_Y is the set of cliques that are contained in remaining nodes

$$P(X_1, \dots X_n) = \frac{1}{Z} \prod_{i \in Z_X} \phi_i(D_i) \cdot \prod_{i \in Z_Y} \phi_i(D_i)$$
$$= \frac{1}{Z} f(X,Z) \cdot g(Y,Z)$$

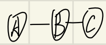


$\therefore (X \perp Y | Z)$

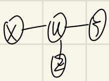
Any conditional independence revealed in $\mathcal{H}$ holds true in P

If $\mathcal{H}$ is an I-map for positive P, then P is a Gibbs distribution that factorizes over $\mathcal{H}$

positive P factorizes over $\mathcal{H}$
$\Updownarrow$
$\mathcal{H}$ is an I-map of P



If X and Y are not separated given Z in $\mathcal{H}$,
then X and Y are dependent given Z in some distribution that factorizes over $\mathcal{H}$



$\phi_1(X,U) =$

X	U	ϕ_1
0	0	1
0	1	0
1	0	0
1	1	1

$\phi_2(U,Y) =$

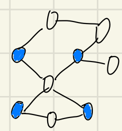
U	Y	ϕ_2
0	0	1
0	1	0
1	0	0
1	1	1

pairwise independencies: $I_p(\mathcal{H}) = \{ (X \perp Y \,|\, \mathcal{X} - \{X, Y\}) : X, Y \text{ are not directly connected} \}$

For a graph $\mathcal{H}$, the Markov blanket of X in $\mathcal{H}$ $MB_{\mathcal{H}}(X)$ is the neighbors of X

local independencies: $I_l(\mathcal{H}) = \{ (X \perp \mathcal{X} - \{X\} - MB_{\mathcal{H}}(X) \,|\, MB_{\mathcal{H}}(X)) : X \in \mathcal{X} \}$

a variable is independent of the rest given its immediate neighbors



given variables block the flow of influence

$P \models I_l(\mathcal{H}) \Leftrightarrow P \models I_p(\mathcal{H}) \Leftrightarrow P \models I(\mathcal{H})$ for a positive distribution

· From distribution to Graphs

2 approaches to encode independencies of a distribution with a graph structure

1° pairwise independencies

add edges between all pairs of X Y such that $P \not\models (X \perp Y \mid \mathcal{X} - \{X, Y\})$

2° local independencies

A set U is a Markov Blanket of X in a distribution P if

$$\begin{cases} X \notin U \\ U \text{ is a minimal set of node such that } (X \perp \mathcal{X} - \{X\} - U \mid U) \in \mathcal{I}(P) \end{cases}$$

define a graph $\mathcal{H}$ by introducing an edge $\{X, Y\}$ for all X and all $Y \in MB_P(X)$

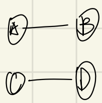
both of these 2 approaches produce the unique minimal I-map of P for a positive distribution

eg. a non positive distribution P over 4 binary variables A B C D,
$P(A, B, C, D) \neq 0$ only when $A = B = C = D$

A	B	C	D	
a^0	b^0	c^0	d^0	0.5
	⋮			
a^1	b^1	c^1	d^1	0.5

1° local independency construction

$P \models (A \perp C, D \mid B)$ → edge AB



$\mathcal{H}$

∴ B is a valid choice for $MB_P(A)$

$P \models (C \perp A, B \mid D)$ → edge $C - D$

$P \models (D \perp A, B \mid C)$ → edge $D - C$

$P \models (B \perp C, D \mid A)$ → edge $B - A$

but $\mathcal{H}$ is not a I-map since $(A \perp C) \in \mathcal{I}(\mathcal{H})$ $(A \perp C) \notin \mathcal{I}(P)$

2° pairwise independencies construction

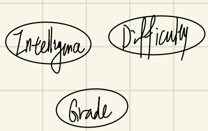
none of A, B, C, D satisfies $(X \perp Y \mid \mathcal{X} - \{X, Y\})$

∴ gives an empty graph

not an I-map for P

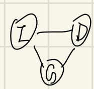
deterministic relations between variable can lead to failure in construction based on local and pairwise independence

eg.



$(D \perp G)$ $(I \perp G)$
$(I \perp D \mid G)$

$\xrightarrow{\text{minimal I-map}}$



does not capture $(I \perp D)$

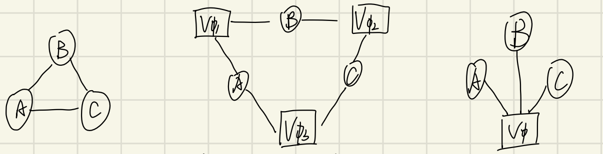
not every distribution P has a perfect map

Factor Graphs

A factor graph $\mathcal{F}$ is an undirected graph containing 2 types of nodes: variable nodes and factor nodes.

each edge connects two nodes of different types

each factor node V_ϕ has one factor ϕ whose scope is its neighbors



different graphs for the same Markov Network

Bayesian Network to Markov Network

Given a Bayesian network $\mathcal{B}$, how to represent the distribution $P_{\mathcal{B}}$ as a parameterized Markov network

finding a minimal I-map for distribution $P_{\mathcal{B}}$

Given a graph G, how to represent the independencies in G using a undirected graph $\mathcal{H}$

finding a minimal I-map for the independencies $\mathcal{I}(G)$

let $\mathcal{B}$ be a Bayesian Network over $\mathcal{X}$ and $E=e$ an observation. let $W = \mathcal{X} - E$

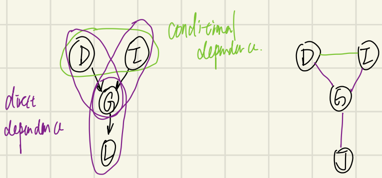
$P_{\mathcal{B}}(W \mid E=e)$ is a Gibbs distribution defined by $\Phi = \{\phi_{X_i}\}$

$$\phi_{X_i} = P_{\mathcal{B}}(X_i \mid Parents(X_i))[E=e]$$

The partition function for this Gibbs distribution is $P(e)$

the moral graph $\mathcal{M}[G]$ of a Bayesian network structure G over $\mathcal{X}$ is the undirected graph over $\mathcal{X}$ that contains an undirected edge between X and Y if

$$\begin{cases} \text{there's a directed edge between } X \text{ and } Y \\ \text{or} \\ X \text{ and } Y \text{ are both parents of the same node} \end{cases}$$



Let G be a Bayesian network structure, for any distribution P_B such that B is a parameterization of G, $M[G]$ is an I-map for P_B

$M[G]$ captures local dependencies (direct independencies and v-structure conditional dependencies)

let G be any Bayesian network graph. the moralized graph $M[G]$ is a minimal I-map

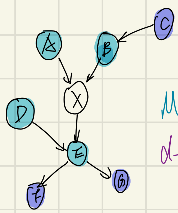
for any $X \in \mathcal{X}$, select the smallest set U such that $(X \perp \mathcal{X} - U - \{X\} \mid U)$

the Markov Blanket of X in a Bayesian Network G $MB_G(X)$ as

- X's parents
- X's children
- other parents of X's children

d-sep$(X; C \mid B)$

d-sep$(X; \{F, G\} \mid E)$



$MB_G(X)$

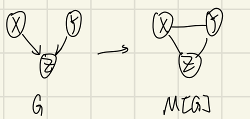
d-separated nodes

For a Bayesian Network graph G, $MB_G(X)$ d-separates X from all other variables

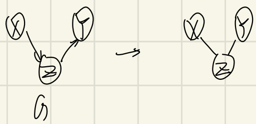
no subset of $MB_G(X)$ does

The addition of the moralizing edges to the Markov Network $\mathcal{H}$ leads to the loss of independency information

But independency informations is not always lost



$M[G]$ lose the information that X and Y are marginally independent



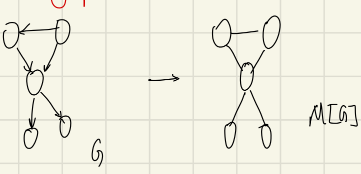
independency information is not lost

A Bayesian Network G is moral if it contains no immoralities (all v-structures have a covering edge)

covering edge

immorality

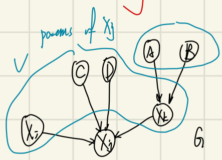
if a directed graph G is moral, then its moralized graph $M[G]$ is a perfect map of G



· Markov Network to Bayesian Network

let $\mathcal{H}$ be a markov network structure, G is any BN minimal I-map for $\mathcal{H}$.

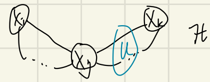
G can have no immoralities



suppose there is immorality in G

$\because$ $G \models (X_i \not\perp X_j \mid X_k)$ $(X_k \not\perp X_j \mid X_i)$

$\therefore$ $\mathcal{H} \models (X_i \perp X_j \mid X_k)$ $(X_k \not\perp X_j \mid X_i)$



$\therefore$ $\mathcal{H}$ contains one or more path $\begin{cases} \text{between } X_i \text{ and } X_j \text{ not cut by } X_k \\ \text{between } X_k \text{ and } X_j \text{ not cut by } X_i \end{cases}$

$\therefore$ there're one or more paths in $\mathcal{H}$ between X_i and X_k via X_j

$\because$ $G \models \{X_i \perp X_k \mid U\}$ $\therefore$ $\mathcal{H} \models \{X_i \perp X_k \mid U\}$

$\therefore$ U block either paths between X_i and X_j or paths between X_j and X_k in $\mathcal{H}$

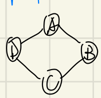
$\because$ $G \models (X_j \perp U \mid V)$ $\therefore$ $\mathcal{H} \models (X_j \perp U \mid V)$

$\therefore$ U cannot be in the path between X_j and X_k

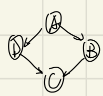
contradiction

A Markov network may not have a BN perfect map

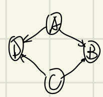
eg BN perfect Map of MN



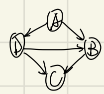
$(A \perp C \mid B, D)$
$(B \perp D \mid A, C)$



$(B \perp D \mid A)$
not a perfect map



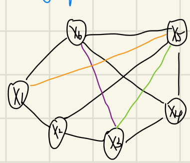
$(B \perp D \mid A, C)$
$(A \perp C)$
not a perfect map



$(A \perp C \mid B, D)$
$(D \not\perp B \mid A, C)$
not a perfect map

let $X_1 - X_2 - \cdots X_k$ be a loop in the graph; a chord in the loop is an edge connecting X_i and X_j for two nonconsecutive nodes X_i and X_j.

A undirected graph $\mathcal{H}$ is said to be chordal if any loop $X_1 - X_2 \cdots - X_k - X_1$ for $k \geq 4$ has a chord



chord for loop $X_1 - X_2 - X_5 - X_6$

chord for loop $X_2 - X_3 - X_4 - X_5$

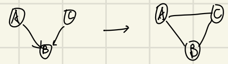
chord for loop $X_1 - X_2 - X_3 - X_4 - X_6$

let $\mathcal{H}$ be a nonchordal Markov network, then there is no Bayesian network G which is a perfect map for $\mathcal{H}$ $(I(G) = I(\mathcal{H}))$

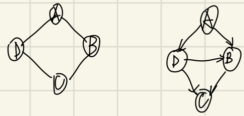
let $\mathcal{H}$ be a chordal Markov network, there is a Bayesian network G such that $I(\mathcal{H}) = I(G)$

Converting Bayesian Network ⟷ Markov Network loses independencies

BN → MN: lose independencies in V-structure



MN → BN: add triangulating edges to loops



· Conditional Random Field

A conditional Random Field is an undirected graph $\mathcal{H}$ whose nodes correspond to $X \cup Y$, the network is annotated with a set of factors $\phi_1(D_1), \dots, \phi_m(D_m)$ such that each $D_i \subseteq X$

the network encodes a conditional distribution as

$$P(Y|X) = \frac{1}{Z(X)} \tilde{P}(Y, X)$$

$$= \frac{1}{\sum_Y \prod_{i=1}^m \phi_i(D_i)} \prod_{i=1}^m \phi_i(D_i)$$

Two variables in $\mathcal{H}$ are connected by an undirected edge whenever they appear together in the same scope of a factor

instead of modeling a Joint Distribution $P(X, Y)$, CRF models a conditional distribution $P(Y|X)$ (not caring the distribution of X)

eg. logistic regression

X_i	Y	ϕ_i
0	0	$\exp(0)=1$
0	1	$\exp(0)=1$
1	0	$\exp(0)=1$
1	1	$\exp(w_i)$

$$\phi_i(X_i, Y) = \exp\left(w_i \cdot \mathbb{1}\{X_i=1, Y=1\}\right)$$

$$\phi_i(X_i, Y=1) = \begin{cases} \exp(w_i) & \text{if } X_i=1 \\ \exp(0)=1 & \text{if } X_i=0 \end{cases} \Rightarrow \exp(w_i \cdot X_i)$$ (how much X_i contribute to $Y=1$)

$$\tilde{P}_\theta(X, Y=1) = \prod_i \exp(w_i X_i) = \exp\left(\sum_i w_i X_i\right)$$

$$\tilde{P}_\theta(X, Y=0) = 1$$

$$P_\theta(Y=1|X) = \frac{\tilde{P}_\theta(X, Y=1)}{\tilde{P}_\theta(X, Y=1) + \tilde{P}_\theta(X, Y=0)} = \frac{\exp(w^T x)}{\underset{\sim\sim\sim}{1+\exp(w^T x)}}$$

the sigmoid function

· Log-Linear Models

rewrite a factor $\phi(D)$ as

$$\phi(D) = \exp(-\epsilon(D))$$

$$\epsilon(D) = -\ln(\phi(D))$$

$\epsilon(D) = -\ln\phi(D)$ is called an energy function,

in statistical physics, the probability of a physical state (eg. configuration of set of electrons) depends inversely on its energy.

Any Markov network parameterized using positive factors can be converted to a logarithmic representation

$$P(X_1 \cdots X_n) \propto \exp\left[-\sum_i \epsilon_i(D_i)\right]$$

eg.

$\phi_1(A,B)$			$\phi_2(B,C)$			$\phi_3(C,D)$			$\phi_4(D,A)$		
a^0	b^0	30	b^0	c^0	100	c^0	d^0	1	d^0	a^0	100
a^0	b^1	5	b^0	c^1	1	c^0	d^1	100	d^0	a^1	1
a^1	b^0	1	b^1	c^0	1	c^1	d^0	100	d^1	a^0	1
a^1	b^1	10	b^1	c^1	100	c^1	d^1	1	d^1	a^1	100
(a)			(b)			(c)			(d)		

$\Rightarrow$

$\epsilon_1(A,B)$			$\epsilon_2(B,C)$			$\epsilon_3(C,D)$			$\epsilon_4(D,A)$		
a^0	b^0	−3.4	b^0	c^0	−4.61	c^0	d^0	0	d^0	a^0	−4.61
a^0	b^1	−1.61	b^0	c^1	0	c^0	d^1	−4.61	d^0	a^1	0
a^1	b^0	0	b^1	c^0	0	c^1	d^0	−4.61	d^1	a^0	0
a^1	b^1	−2.3	b^1	c^1	−4.61	c^1	d^1	0	d^1	a^1	−4.61
(a)			(b)			(c)			(d)		

−4.61 when 2 var agree and 0 otherwise

let D be a set of variables. We define a feature $f(D)$ to be a function $f: Val(D) \mapsto \mathbb{R}$

$$\tilde{P}_\Phi = \prod_i \phi_i(D_i)$$

$$= \prod_j \exp(-w_j f_j(D_j))$$

f_j is a feature

eg.

$\phi_{12}(X_1, X_2) =$

X_1	X_2	ϕ_{12}
0	0	a_{00}
0	1	a_{01}
1	0	a_{10}
1	1	a_{11}

$$f^{00}(X_1, X_2) = 1\{X_1 = 0 \quad X_2 = 0\}$$

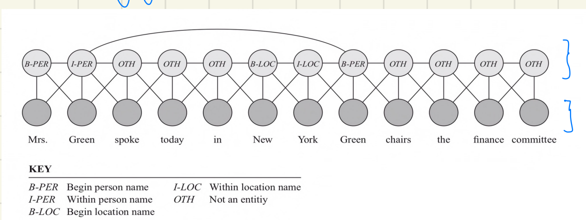
$$f^{01}(X_1, X_2) = 1\{X_1 = 0, X_2 = 1\}$$

$\vdots$

$$\phi_{12}(X_1, X_2) = \exp\left(-\sum_{kl} w_{kl} \cdot f^{kl}(X_1, X_2)\right)$$

$$w_{kl} = -\log a_{kl}$$

eg. Features of Language.



(a)

} Y: Category

} X: word

$$f_1(X_i, Y_i) = 1\{Y_i = \text{Person name}, X_i \text{ capitalized}\}$$

how important capitalization is in recognizing a person

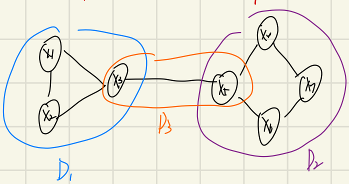
$$f_2(X_i, Y_i) = 1\{Y_i = \text{location name}, X_i \text{ appears in atlas}\}$$

A distribution P is a log-linear model over a Markov network $\mathcal{H}$ if P is associated with

$$\begin{cases} \text{a set of features } \mathcal{F} = \{f_1(D_1), \dots, f_k(D_k)\} \text{ where each } D_i \text{ is a complete subgraph in } \mathcal{H} \\ \text{a set of weights } w_1 \cdots w_k \end{cases}$$

such that

$$p(x_1, \dots x_n) = \frac{1}{Z} \exp\left[-\sum_{i=1}^{k} w_i f_i(D_i)\right]$$



$$p(x_1, \dots x_n) = \frac{1}{Z} \phi_1(D_1) \times \phi_2(D_2) \times \phi_3(D_3)$$

$$= \frac{1}{Z} \exp(-\epsilon_1(D_1)) \cdot \exp(-\epsilon_2(D_2)) \cdot \exp(-\epsilon_3(D_3))$$

$$= \frac{1}{Z} \exp\left[-\sum_{i=1}^{k} \epsilon_i(D_i)\right]$$

$$= \frac{1}{Z} \exp\left[-\sum_{i=1}^{k} w_i f_i(D_i)\right]$$

· Metric MRFs

All X_i takes value in same label space V $\quad$ X_i — X_j $\quad$ want X_i and X_j to take "similar" values

Distance function $\mu : V \times V \mapsto \mathbb{R}^{+}$

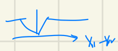
- reflexivity $\mu(x, x) = 0 \quad \forall x \in V$
- symmetry $\mu(x_1, x_2) = \mu(x_2, x_1)$
- triangle inequality: $\mu(x_1, x_3) \leq \mu(x_1, x_2) + \mu(x_2, x_3)$

$f_{ij}(X_i, X_j) = \mu(X_i, X_j)$

$\exp(-w_{ij} f_{ij}(X_i, X_j))$ where $w_{ij} > 0$ $\quad$ low distance $\rightarrow$ high probability

eg.

$\mu(x_1, x_2) = |x_1 - x_2|$

$\mu(x_1, x_2) =$ 

eg Image segmentation

no penalty when adjacent super pixels take same class

some penalty otherwise