2/27/14

MRF SEMANTICS

① I-map

→ Same concept & definition as BNs

$I(G) = \{$ Set of independence assumptions encoded in $G \}$

$= \{ (X \perp Y \mid Z) :$ X & Y are separated in G given $Z \}$

Now, G is an I-map of P if $I(G) \subseteq I(P)$

② P-MAP

G is a P-map of P if $\underbrace{I(G) = I(P)}$

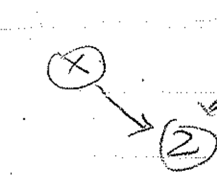
G encodes exactly the independencies in P
No more, no less

Do P-maps exist $\forall P$? No!

Consider $I(P) = \{ (X \perp Y) \}$

Note $(X \perp Y \mid Z) \notin I(P)$

Such a distribution has a P-map in BNs:



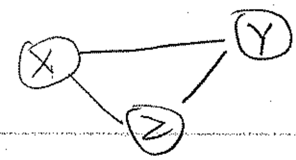
But can't represent this as a MN:

→ attempt #1 X Y Z but $X \perp Z \in I(G)$

$\notin I(P)$

G is not an I-map

→ attempt #2

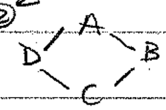


$I(G) = \phi$ so G is an I-map
but not a P-map

So overall, there are distributions P that have

→ P-maps in BNs but not in MNs e.g $X \rightarrow Z \leftarrow Y$

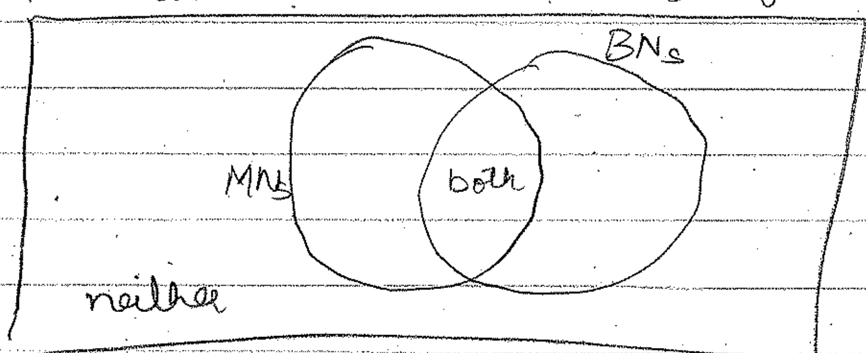
→ P-maps in MNs but not in BNs e.g



→ P-maps in both BNs & MNs e.g $X - Y$
$X \rightarrow Y$
$X \leftarrow Y$

→ P-maps in neither BNs Nor MN!
[Trickier to construct but exist; see HW]

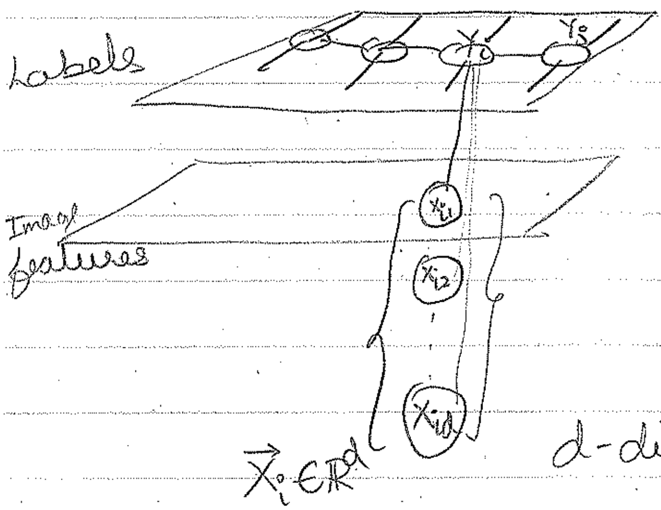
"Qualitative" Venn diagram



All distributions

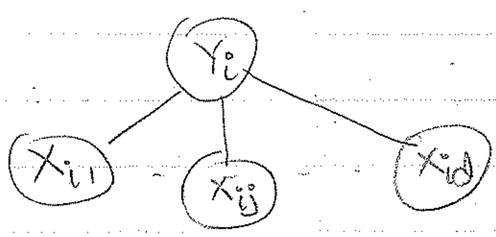
3 Conditional Random Fields

Consider an image labeling task



d-dim features at each pixel

If there was only 1 pixel in the image:



Looks like NB [Actually makes the same assumptions]

Recall that NB models $P(X_{i1} \ldots X_{id}, Y_i)$

& Logistic Regression models $P(Y_i \mid X_{i1}, \ldots, X_{id})$

The conversion from MRFs to CRFs is analogous

MRFs model the joint distribution

$$P(\vec{X}_1, \ldots \vec{X}_n, Y_1 \ldots Y_n) = \frac{1}{Z} \underbrace{\prod_i \prod_d \Phi_i(X_{id}, Y_i) \prod_{ij} \Phi_{ij}(Y_i, Y_j)}_{\text{Pairwise MRF}}$$

Notice $\Phi(\vec{X}_i, Y_i) \neq \underbrace{\Phi(X_{id}, Y)}_{\text{Pairwise-factor}}$

$$= \underbrace{\Phi(X_{i1}, \ldots X_{id}, Y_i)}_{\text{High-Order Factor}}$$

Notice $Z = \underbrace{\sum_{\substack{X_{i1} \ldots X_{id} \\ i \in \{1 \ldots n\}}}}_{\text{assuming discrete features}} \sum_{y_1 \ldots y_n} \prod_i \prod_d (\cdot) \prod_{ij} (\cdot)$

or $\int_{x_{i1}} \int \ldots \int_{x_{id}}$ if continuous

CRFs only compute

$$\underbrace{Z(\vec{x}_1, \ldots, \vec{x}_n)}_{\text{Z as a function of input}} = \sum_{y_1 \ldots y_n} \prod_i \prod_d \Phi_{id}(X_{id}, Y_i) \prod_{ij} \Phi_{ij}(Y_i, Y_j)$$

Different for each image!

Now $\boxed{P(Y_1 \ldots Y_n \mid \vec{x}_1 \ldots \vec{x}_n) = \frac{1}{Z(\vec{x}_1 \ldots \vec{x}_n)} \prod (\cdot)}$

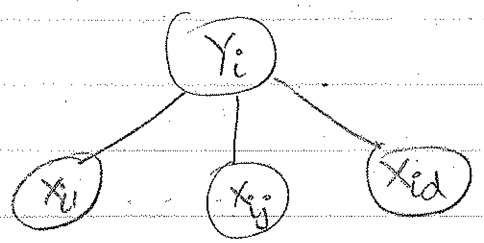
CRF!

So why should you care?

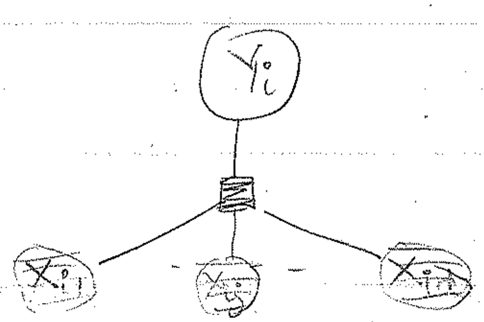
Because note: $\Phi_{id}(X_{id} = x_{id}, Y_i) = \underbrace{\begin{bmatrix} \\ \\ \end{bmatrix}_{k \times 1}}_{\text{Node potential over } Y_i}$

Moreover $\Phi_i(\vec{X}_i = \vec{x}_i, Y_i) \equiv$ also just a node potential

So an MRF (just like NB) was forced to make restrictive assumptions



But a CRF (just like a LR) can extract arbitrary information from $\vec{X}_i$



} Always observed

In fact $\Phi_i(Y_i)$ can look at features for the ENTIRE image

$$\underbrace{\Phi_i(Y_i)}_{\text{still a node potential}} \equiv \Phi(Y_i, \vec{X}_1, \ldots \vec{X}_n)$$

(4) Log-Linear Models

Consider an MRF/CRF: $P(Y_1 \ldots Y_n) = \frac{1}{Z} \prod_c \Phi_c(Y_c)$

Notice that $\Phi_c(Y_c)$ is always > 0
$\Rightarrow$ I can express it as $e^{\log \Phi_c(Y_c)}$

We define $\underbrace{E_c(Y_c)}_{\text{"energy" term}} = -\log \Phi_c(Y_c)$

So $$P(Y_1 \ldots Y_n) \propto e^{-\sum_c E_c(Y_c)}$$
$$= \frac{1}{Z} e^{-\sum_c E_c(Y_c)}$$
$$= e^{-\sum_c E_c(Y_c) - \log Z}$$

$$\underbrace{\sum_c E_c(Y_c)}_{} = \sum_i E_i(Y_i) + \sum_{ij} E_{ij}(Y_i, Y_j)$$

for pairwise MRFs

Energy function: term comes from statistical physics

Also $\operatorname*{argmax}_{y_1 \ldots y_n} P(Y_1 \ldots Y_n) = \underbrace{\operatorname*{argmin}_{y_1 \ldots y_n} \sum_c E_c(Y_c)}_{}$

MAP Inference $\equiv$ Energy Minimization

Now in log-space, it is common to represent

$$E_c(Y_c) = \underbrace{\vec{w}_c^{\,T}}_{\text{weigh}} \underbrace{\vec{f}(Y_c)}_{\text{feature}}$$

Log-linear model $P(Y_1 \dots Y_n) \propto e^{-\sum_c w_c^T f_c(Y_c)}$

Example of $\vec{w}_c$ & $\vec{f}_c$?

→ MRF: $\vec{f}_c(Y_c) \equiv$ indicator vector for state

ie $f_i(Y_i) = \begin{bmatrix} 1[[Y_i = 0]] \\ 1[[Y_i = 1]] \end{bmatrix}$

Now $E_i(Y_i = 0) = [w_0 \ \ w_1] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = w_0$

$E_i(Y_i = 1) = [w_0 \ \ w_1] \begin{bmatrix} 0 \\ 1 \end{bmatrix} = w_1$

Some feature of data

→ CRF: $f_i(Y_i) = \begin{bmatrix} 1[[Y_i = 0]] \cdot \Phi(\vec{x}) \\ 1[[Y_i = 1]] \cdot \Phi(\vec{x}) \end{bmatrix}_{2d \times 1}$ $\Phi(x) \in \mathbb{R}^d$

$$\vec{w}_i = \begin{bmatrix} \vec{w}_{i0} \in \mathbb{R}^d \\ \vec{w}_{i1} \in \mathbb{R}^d \end{bmatrix}$$

$E_i(Y_i = 0) = \vec{w}_{i0}^{\,T} \Phi(\vec{x})$ = Energy/Score of category 0

$E_i(Y_i = 1) = \vec{w}_{i1}^{\,T} \Phi(\vec{x})$ = Energy/Score of category 1