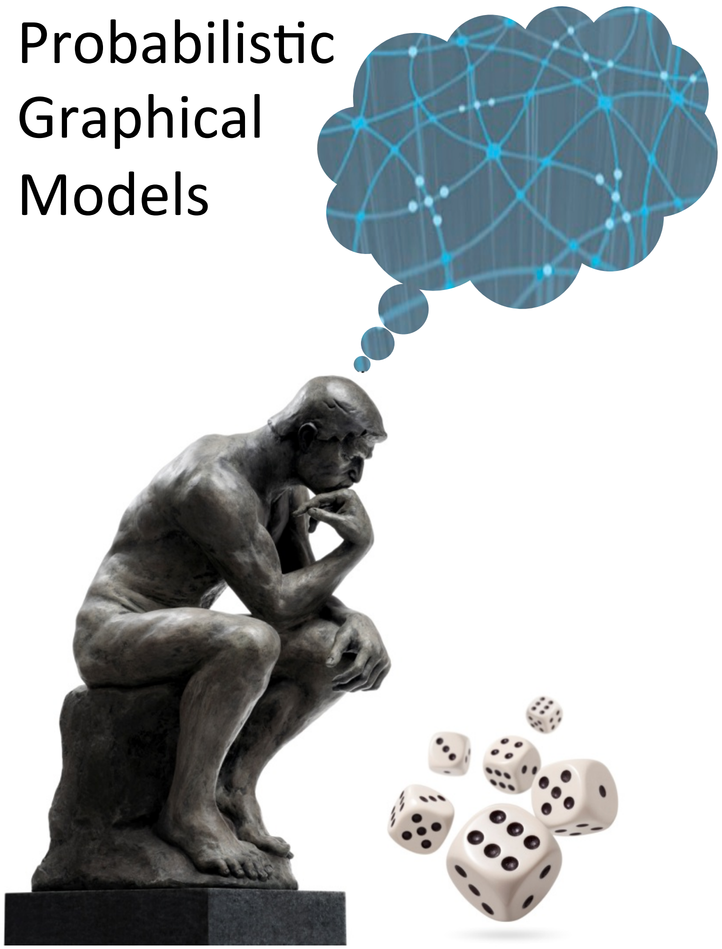
Probabilistic
Graphical
Models



Representation

Local Structure

Log-Linear Models

Log-Linear Representation

$$\boxed{\tilde{P} = \prod_i \phi_i(\boldsymbol{D}_i)} \Rightarrow \quad \tilde{P} = \exp\Big(\underbrace{-\sum_j \overset{\text{coefficient}}{w_j} f_j(\boldsymbol{D}_j)}_{\text{features}}\Big)$$

linear

$$\tilde{P} = \prod_j \underbrace{\exp\left(-w_j f_j(\boldsymbol{D}_j)\right)}_{\text{factor}}$$

- Each feature f_j has a scope D_j
- Different features can have same scope

Representing Table Factors

$$\phi(X_1, X_2) = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}$$

$$f_{12}^{00} = \mathbb{1}\{X_1 = 0, X_2 = 0\}$$ 1 / 0

$$f_{12}^{01} = \mathbb{1}\{X_1 = 0, X_2 = 1\}$$

$$f_{12}^{10} = \mathbb{1}\{X_1 = 1, X_2 = 0\}$$

$$f_{12}^{11} = \mathbb{1}\{X_1 = 1, X_2 = 1\}$$

General representation

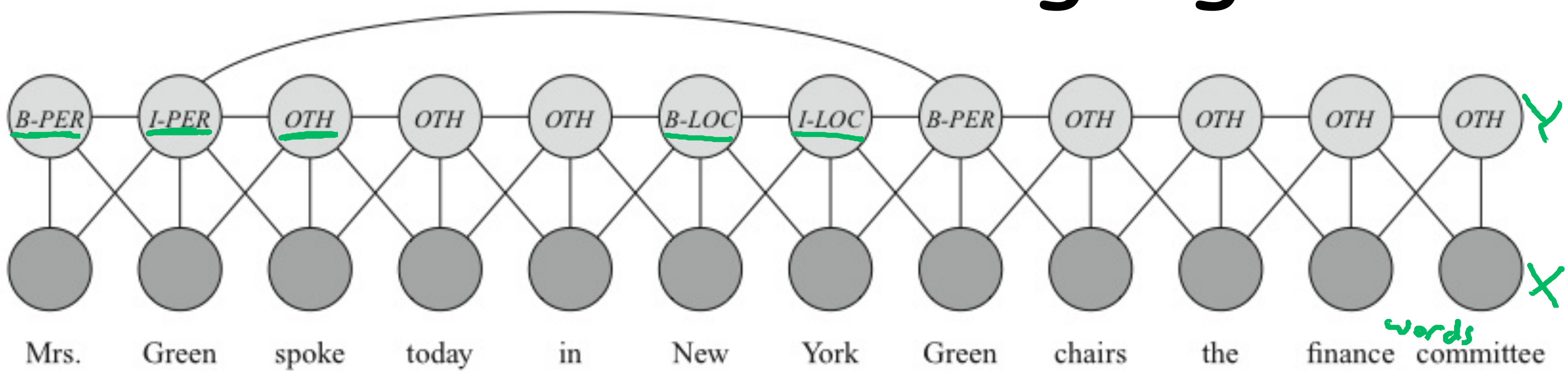
$$\phi(X_1, X_2) = \exp(-\textstyle\sum_{kl} w_{kl} f_{ij}^{kl}(X_1, X_2))$$

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

$\exp(-w_{00})$ when $x_1=0, x_2=0$

$\exp(-w_{01})$ when $x_1=0, x_2=1$

Features for Language



Features: word capitalized, word in atlas or name list, previous word is "Mrs", next word is "Times", …

$f(Y_i, X_i) = \mathbb{1}\{Y_i = \text{B-person}, X_i \text{ is capitalized}\}$

$\mathbb{1}\{Y_i = \text{B\_loc}, X_i \text{ appears in Atlas}\}$

Daphne Koller

Ising Model

pairwise

same / different

joint spins

$$E(x_1,\ldots,x_n) = -\sum_{i<j} w_{i,j}\, x_i x_j - \sum_i u_i x_i$$

$x_i \in \{-1,+1\}$ $\quad f_{i,j}(X_i, X_j) = X_i \cdot X_j$

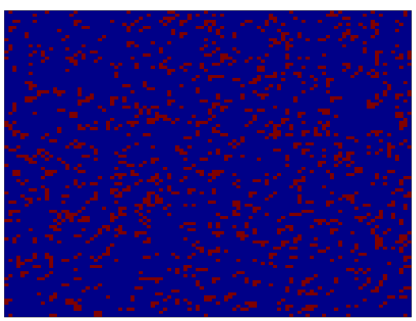
$$P(\boldsymbol{X}) \propto e^{-\frac{1}{T}E(\boldsymbol{X})}$$

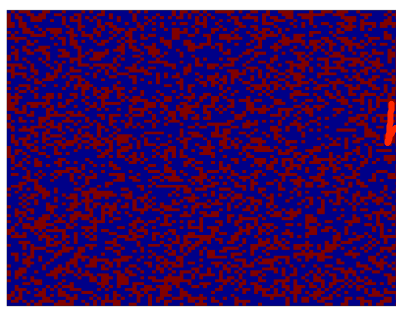
$\frac{w_{ij}}{T} \to 0$

T grows

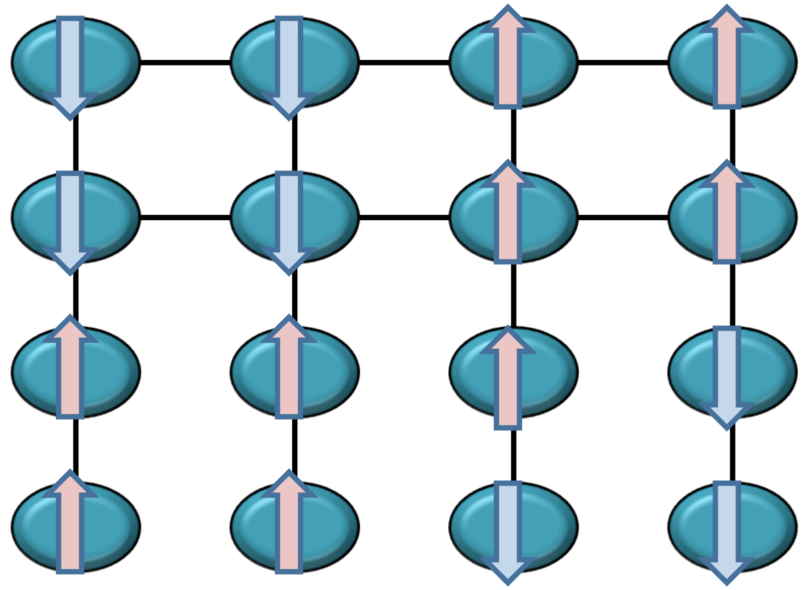
T decreases

low





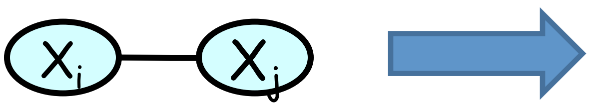
high T



Daphne Koller

Metric MRFs

- All X_i take values in label space V



want X_i and X_j to take "similar" values

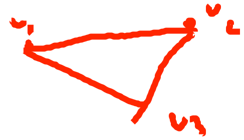
- Distance function $\mu : V \times V \rightarrow R^+$

metric

- Reflexivity: $\mu(v,v) = 0$ for all v
- Symmetry: $\mu(v_1,v_2) = \mu(v_2,v_1)$ for all v_1, v_2

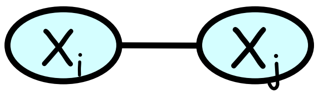
semi metric

- Triangle inequality: $\mu(v_1,v_2) \leq \mu(v_1,v_3) + \mu(v_3, v_2)$ for all v_1, v_2, v_3



Metric MRFs

- All X_i take values in label space V



want X_i and X_j to take "similar" values

- Distance function $\mu : V \times V \rightarrow R$

$$\underbrace{f_{i,j}(X_i, X_j)} = \mu(X_i, X_j)$$

lower distance

higher

$$\underbrace{\exp\left(-w_{ij} f_{ij}(X_i, X_j)\right)} \qquad w_{ij} > 0$$

w_{ij}

higher probability

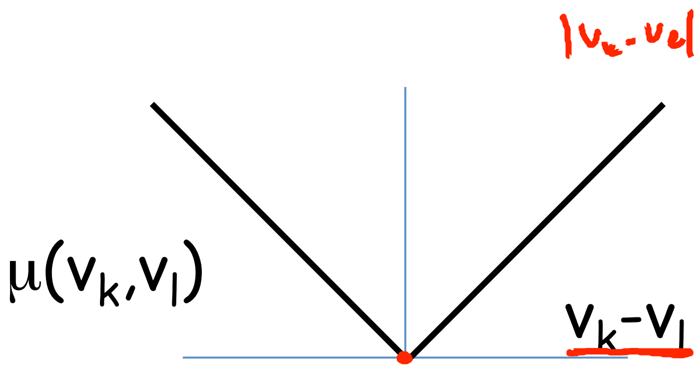
values of X_i and X_j far in μ ⇒ lower probability

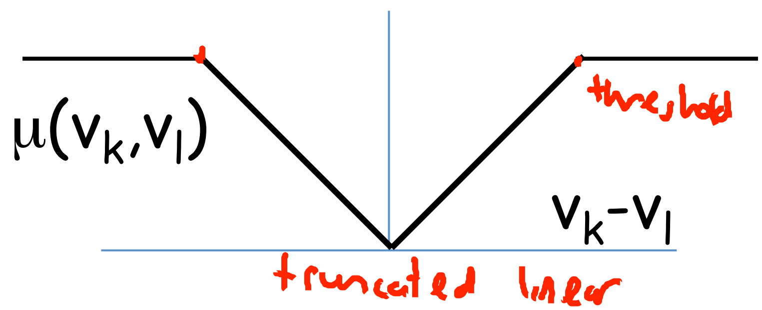
Daphne Koller

Metric MRF Examples

$$\mu(v_k, v_l) = \begin{cases} 0 & v_k = v_l \\ 1 & \text{otherwise} \end{cases}$$

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



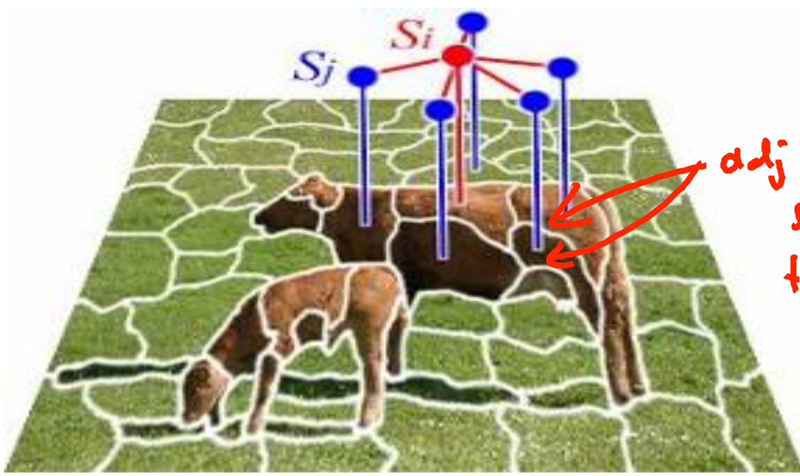


Daphne Koller

Metric MRF: Segmentation

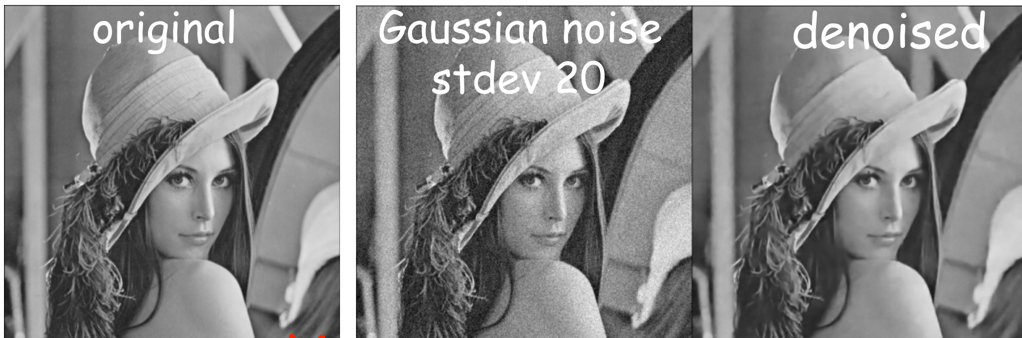
$$\mu(v_k, v_l) = \begin{cases} 0 & v_k = v_l \\ 1 & \text{otherwise} \end{cases}$$

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



adjacent superpixels take same class

Metric MRF: Denoising



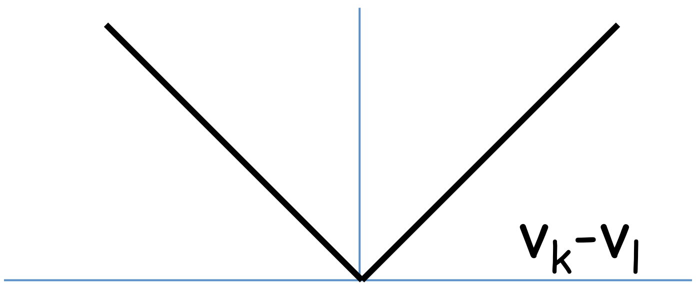
Y_i close to X_i

Y_i close to its neighbors

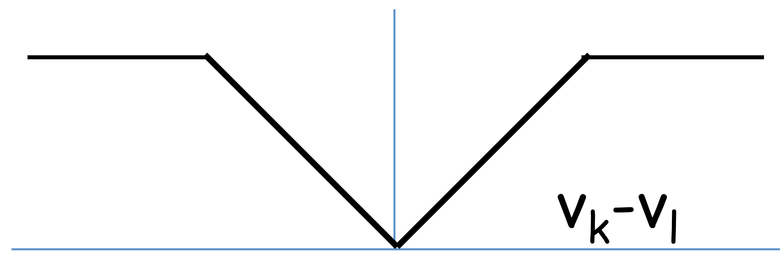
X noisy pixels

Y clean pixels

$\mu(v_k,v_l) = |v_k - v_l|$



$\mu(v_k,v_l) = \min(|v_k - v_l|, d)$



Similar idea for stereo reconstruction