Parameter Learning in Markov Nets

Dhruv Batra,
10-708 Recitation

11/13/2008

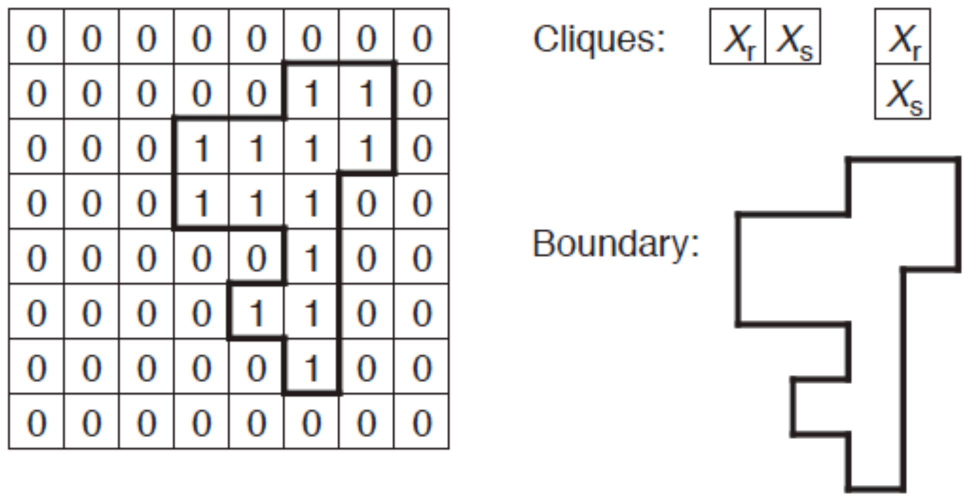
Contents

- MRFs
  - Parameter learning in MRFs
    - IPF
    - Gradient Descent
  - HW5 implementation

Semantics

- Priors on edges
  - Ising Prior / Potts Model
  - Metric MRFs

The Ising Model: A 2-D MRF[100]



- Potential functions are given by

$$V(x_r, x_s) = \beta\delta(x_r \neq x_s)$$

where β is a model parameter.

- Energy function is given by

$$\sum_{c \in \mathcal{C}} V_c(x_c) = \beta(\text{Boundary length})$$

- Longer boundaries $\Rightarrow$ less probable

Metric MRFs

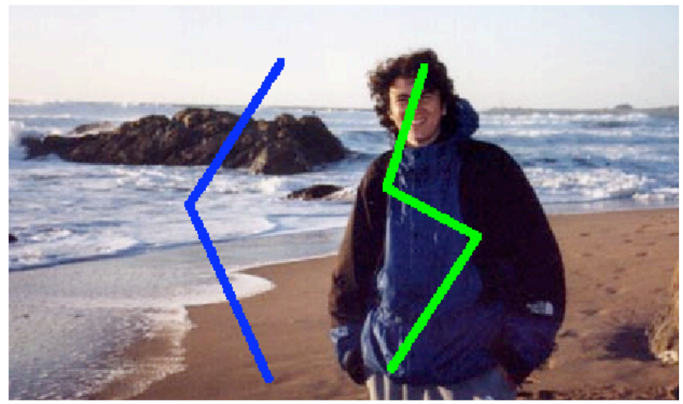
- Energies in pairwise MRFs

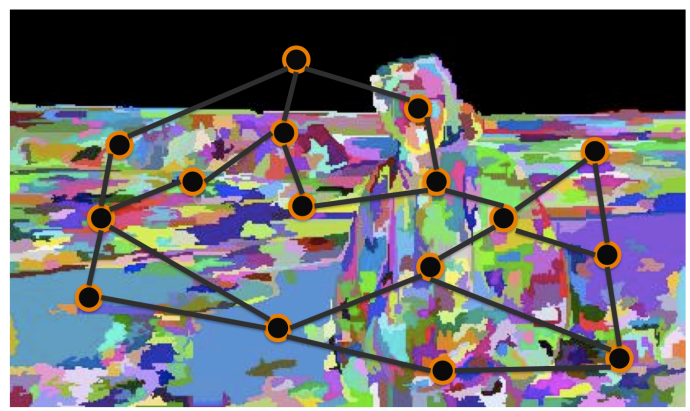
$$E(f) = \sum_{\{p,q\}\in\mathcal{N}} V_{p,q}(f_p, f_q) + \sum_{p\in\mathcal{P}} D_p(f_p),$$

$$\begin{aligned} V(\alpha,\beta) = 0 \quad &\Leftrightarrow \quad \alpha = \beta, \\ V(\alpha,\beta) \quad &= \quad V(\beta,\alpha) \geq 0, \\ V(\alpha,\beta) \quad &\leq \quad V(\alpha,\gamma) + V(\gamma,\beta), \end{aligned}$$

HW4

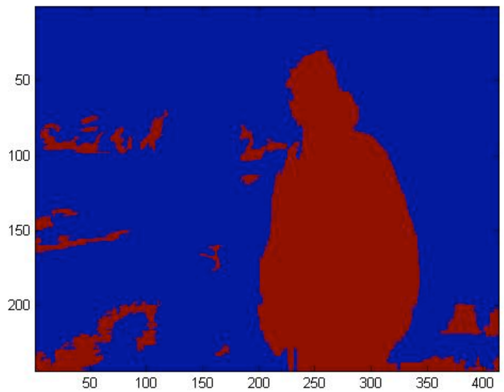
- Semi-supervised image segmentation



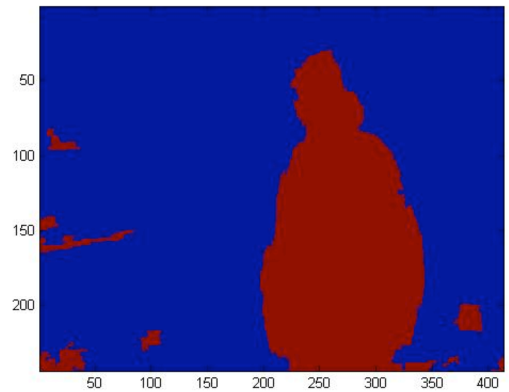


HW4

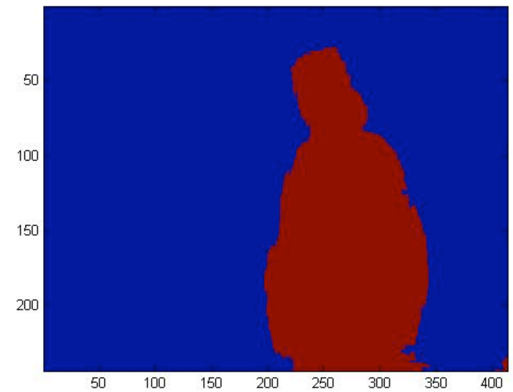
- Effect of varying beta



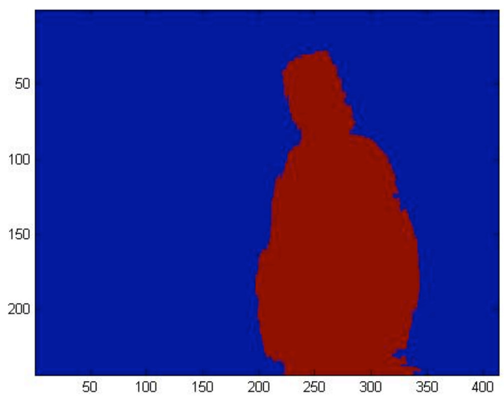
beta = 0



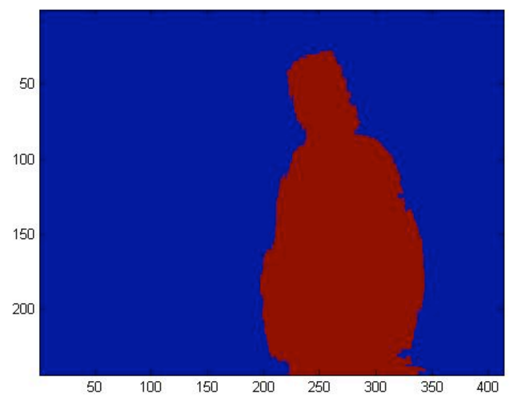
beta = 2



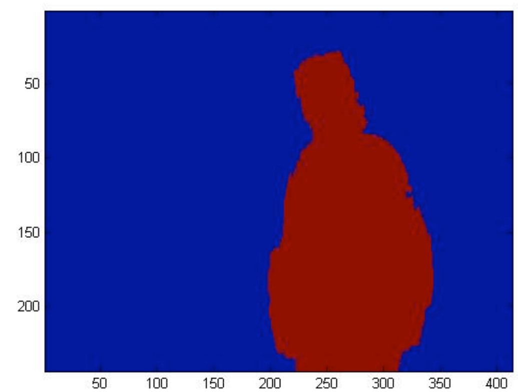
beta = 4



beta = 6



beta = 8



beta = 10

Segmentations by Congcong Li

HW 5

- Potts Model

$$\Psi(x_i, x_j) = \exp\left\{-\beta \times I\left(x_i \neq x_j\right)\right\},$$

- More general parameters

$$\Psi(x_i, x_j) \quad = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix}.$$

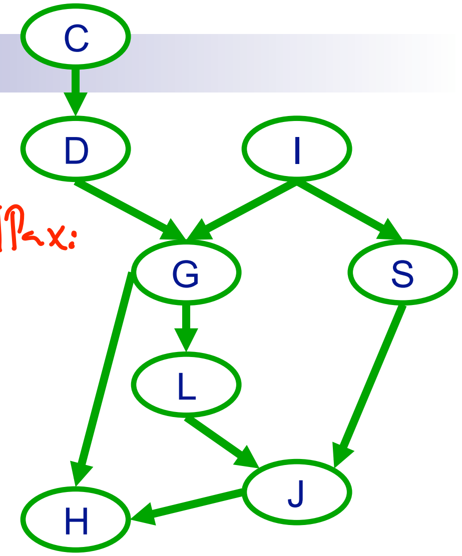
Learning Parameters of a BN

- Log likelihood decomposes:

$$\ell(\mathcal{D}:\theta) = \log P(\mathcal{D} \mid \theta) = m \sum_i \sum_{x_i, \mathbf{Pa}_{x_i}} \underbrace{\hat{P}(x_i, \mathbf{Pa}_{x_i})}_{\text{empirical dist}} \log \underbrace{P(x_i \mid \mathbf{Pa}_{x_i})}_{\theta_{x_i \mid Pa_{x_i}}}$$

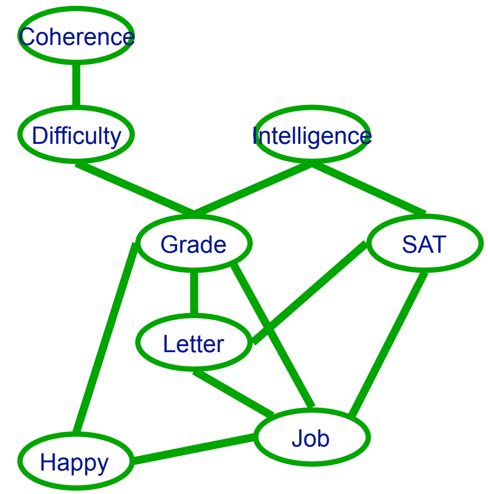
$$\theta_{X_i = t \mid Pa_{x_i} = u} = \frac{\text{Count}(X_i = t, Pa_{x_i} = u)}{\text{Count}(Pa_{x_i} = u)}$$

- Learn each CPT independently
- Use counts



$$\hat{P}(\mathbf{u}) = \frac{\text{Count}(\mathbf{U} = \mathbf{u})}{m}$$

Learning Parameters of a MN



Log-linear Markov network (most common representation)

- **Feature** is some function ϕ[**D**] for some subset of variables **D**
  - e.g., indicator function

$$\psi(G,I,D) = \begin{cases} 1; & \text{if } G=t, I=t, D=f \\ 0; & \text{otherwise} \ldots \end{cases}$$

- **Log-linear model** over a Markov network *H*:
  - a set of features $\phi_1[\mathbf{D}_1],\ldots,\phi_k[\mathbf{D}_k]$
    - each $\mathbf{D}_i$ is a subset of a clique in *H*
    - two ϕ's can be over the same variables
  - a set of weights $w_1,\ldots,w_k$
    - usually learned from data

learn these

  - $$P(X_1,\ldots,X_n) = \frac{1}{Z}\exp\left[\sum_{i=1}^{k} w_i \phi_i(\mathbf{D}_i)\right]$$

HW 5

$$\Psi(x_i, x_j) = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix}.$$

For convenience we denote $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$. This MRF is not decomposable, and therefore we cannot estimate the potentials in closed form.

(a) Input image (b) Ground Truth Segmentation

Figure 2: Training Data

(a) Assume that Φ is known (computed in a manner similar to HW 4, only now all pixels are labelled). Write down the IPF update equation for $\Psi(x_i, x_j)$. What is the cost of computing $\Psi^{(t+1)}(x_i, x_j)$?

(b) Using the equation from part (a) implement IPF for Ψ using the images and data provided. Report the final value of θ. Use loopy belief propagation to compute any

Questions?

Semantics

- Factorization

$$P(X_1, \ldots, X_n) = \frac{1}{Z} P'(X_1, \ldots, X_n)$$

where

$$P'(X_1, \ldots, X_n) = \pi_i[\boldsymbol{D}_1] \times \pi_2[\boldsymbol{D}_2] \times \cdots \times \pi_m[\boldsymbol{D}_m]$$

- Energy functions

$$\pi[\boldsymbol{D}] = \exp(-\epsilon[\boldsymbol{D}]),$$

- Equivalent representation

$$P(X_1, \ldots, X_n) \propto \exp\left[-\sum_{i=1}^{m} \epsilon_i[\boldsymbol{D}_i]\right].$$

Semantics

- Log Linear Models

Definition 5.6.9: A distribution P is a *log-linear model* over a Markov network $\mathcal{H}$ if it is associated with:

- a set of features $\phi_1[\boldsymbol{D}_1], \ldots, \phi_k[\boldsymbol{D}_m]$, where each $\boldsymbol{D}_i$ is a subclique in $\mathcal{H}$,
- a set of weights $w_1, \ldots, w_k$,

such that

$$P(X_1, \ldots, X_n) = \frac{1}{Z} \exp\left[-\sum_{i=1}^{k} w_i \phi_i[\boldsymbol{D}_i]\right].$$