MRF for Labels

binary labels — Ising model

multiple labels — Potts model ⇒ Will look at this case

Label MRF – X where labels are of form $k = 1, \dots, K$

Observed image – Y ; underlying parameters = θ

$$\text{optim} = \arg\max_{\cancel{x},\theta} P(x \mid y, \theta)$$

1) Fix θ, optimize X. (already discussed in image denoising!)

$$\begin{aligned}
P(x \mid y, \theta) &= P(x_i, x_{N_i} \mid y, \theta) \\
&= P(x_i \mid x_{N_i}, y, \theta)\, P(x_{N_i} \mid y, \theta) \\
&= P(x_i \mid x_{N_i}, y, \theta)\, P(x_{N_i} \mid y, \theta) \\
&= P(x_i \mid x_{N_i}, y_i, \theta)\, P(x_{N_i} \mid y, \theta)
\end{aligned}$$

$$\Rightarrow \arg\max_{x} P(x \mid y, \theta) = \arg\max_{x_i} P(x_i \mid x_{N_i}, y_i, \theta_i)$$

$$= \;'' \; \frac{P(y_i \mid x_{N_i}, x_i, \theta_i)\, P(x_i \mid x_{N_i}, \theta)}{P(y_i \mid x_{N_i}, \theta_i)}$$

$$= \;'' \; P(y_i \mid x_i, \theta_i)\, P(x_i \mid x_{N_i}, \theta_i) \quad //$$

2) Fix X, optimize θ :- If the noise is gaussian,

$$\hat{\theta}_i = (\hat{\mu}_i, \hat{C}_i) = (\text{Sample mean, Sample Covar.})$$

* Soft Segmentation

EM algorithm would have to be used, similar to FCM.

E step at t'th iteration

Recall $Q(\theta|\theta^i) = E_{P(x|y,\theta^t)}\left[\log P(x,y|\theta)\right]$

$$= E_{P(x|y,\theta^t)}\left[\log P(y|x,\theta)\, P(x|\theta)\right]$$

why θ not θ_i?

$$= E_{P(x|y,\theta^t)}\left[\sum_i \log P(y_i|x_i,\theta) + \log P(x|\theta)\right]$$

$$= E_{P(x|y,\theta^t)}\left[\log P(x|\theta)\right] + E_{P(x|y,\theta^t)}\left[\sum_i \log P(y_i|x_i,\theta)\right]$$

$$= \quad '' \quad + \sum_i E_{P(x|y,\theta^t)}\left[\log P(y_i|x_i,\theta)\right]$$

$$= \underbrace{E_{P(x|y,\theta^t)}\left[\log P(x|\theta)\right]}_{1} + \underbrace{\sum_i E_{P(x_i,x_{\sim i}|y,\theta^t)}\left[\ ''\ \right]}_{2}$$

- Handling 2 :- $E_{P(x|y,\theta^t)}\left[\log P(y_i|x_i,\theta)\right]$

Approximate as such :- $E_{P(x_i,x_{\sim i}|y,\theta^t)}\left[\log P(y_i|x_i,\theta)\right]$

$$\approx E_{P(x_i|x_{\sim i}, y_i, \theta^t)}\left[\log P(y_i|x_i,\theta)\right]$$

Sum over all labels of x_i for all possibilities of other labels

Keep other labels fixed and sum over all possible labels of x_i

$X_{\mathcal{N}_i}$ is usually fixed to MAP estimate given θ^t parameters.

$$\Rightarrow \quad Q(\theta|\theta^t) = ① + \sum_i E_{P(X_i|X_{\mathcal{N}_i}, Y, \theta^t)} \left[\log P(y_i|x_i, \theta)\right]$$

$$Q(\theta|\theta^t) - ① = \sum_i E_{P(X_i|X_{\mathcal{N}_i}, Y, \theta^t)} \left[\log P(y_i|x_i, \theta)\right]$$

$$= \sum_i E_{P(X_i|X_{N_i}, Y, \theta_t)} \left[\log P(y_i|x_i, \theta)\right]$$

$$= \sum_i \sum_{l=1}^{L} P\left(x_i = l \,\middle|\, X_{N_i}^{MAP}, Y, \theta^t\right) \log P\left(y_i | x_i = l, \theta\right)]$$

Memberships are defined as $P\left(x_i = l \,|\, x_{N_i}, y, \theta^t\right)$ } ?

$$= P\left(x_i = l \,\middle|\, x_{N_i}^{MAP}, y_i, \theta^t\right)$$

$$= \frac{P\left(y_i | x_i = l, x_{N_i}^{MAP}, \theta^t\right) P\left(x_i = l \,|\, x_{N_i}^{MAP}, \theta^t\right)}{P\left(y_i | x_{N_i}^{MAP}, \theta^t\right)}$$

$$= \frac{G(y_i | \mu_l, \sigma_l)\, P\left(x_i = l \,|\, x_{N_i}^{MAP}, \theta^t\right)}{\langle normalization \rangle} \quad — \quad \sum_{l=1}^{L} (\quad)$$

- M Step

Let the updated memberships be γ_{nk}

$$\mu_k = \frac{\sum_n \gamma_{nk} y_n}{\sum_n \gamma_{nk}} \qquad C_k = \frac{\sum_n \gamma_{nk} (y_n - \mu_k)(y_n - \mu_k)^T}{\sum_n \gamma_{nk}}$$

* Hard Segmentation using S-T cuts (binary classification)

Hard segmentation was discussed already.

1 — $\max_\theta P(x|y,\theta) \rightarrow$ Sample mean, var if gaussian noise

2 — $\max_x P(x|y,\theta) \rightarrow$ Image denoising logic ...

$$\max_x P(x|y,\theta) = \max_x \log P(x|y,\theta)$$

$$= \max_x \left[\log P(y|x,\theta) + \log P(x|\theta)\right]$$

$$= \max_x \left[\log \prod_i P(y_i|x_i,\theta) + \log \frac{1}{Z}\exp\left(\frac{1}{2}\sum_{i,j}\beta_{ij}V(x_i,x_j)\right)\right]$$

X is MRF, let T = 2

$\beta_{ij} \geq 0$, Sign relative to defn of $V(x_i, x_j)$

$\beta_{ij} = \beta_{ji}$ — symmetric

$\beta_{ii} = 0$ — No self interaction

Our labels are binary (0,1) $\Rightarrow V(a,b) = ab + (1-a)(1-b)$

$$\Rightarrow P(y_i|x_i,\theta) = P(y_i|x_i=0,\theta)^{1-x_i}\, P(y_i|x_i=1,\theta)^{x_i}$$

$$\Rightarrow \max_x \left[\sum_i x_i P(y_i|x_i=1,\theta) + (1-x_i)P(y_i|x_i=0,\theta) + \frac{1}{2}\sum_{i,j}\beta_{ij}V(x_i,x_j)\right]$$

$$\Rightarrow \text{''} \left[\qquad \text{''} \qquad + \frac{1}{2}\sum_{i,j}\beta_{ij}\left(x_i x_j + (1-x_i)(1-x_j)\right)\right]$$

$$\Rightarrow \max_x \left[\sum_i \lambda_i x_i + \frac{1}{2}\sum_i\sum_j \beta_{ij}\left(2x_i x_j - x_j - x_i\right)\right]$$

We shall model this as an S-T graph problem.

Graph has n+2 nodes (Voxels + S + T)

- Add an edge between S, i if $\lambda_i > 0$ with cost λ_i
  T, i if $\lambda_i \leq 0$ $-\lambda_i$
- Add edge with cost β_{ij} between every neighbouring pair

Recall that a cut divides the set of nodes into two mutually exclusive and exhaustive parts P_1 P_2 where $S \in P_1$
$T \in P_2$

Capacity of a cut is the sum of costs of edges between P_1 and P_2.

★ A min. capacity cut in the above graph indicates MAP labelling ★