5. Marginal probabilities

GRF for pairs (x, s) on a graph (V, E), where

$x: V \to F$ image, $s: V \to K$ interpretation

$$p_u(s) = \frac{1}{Z(u)} \exp\Big[\sum_{i \in V} u_i(s_i) + \sum_{ij \in E} u_{ij}(s_i, s_j)\Big]$$

$$p(x|s) = \prod_{i \in V} p(x_i | s_i)$$

Marginal prob's: $p(s_i)$, $p(s_i, s_j)$, $p(s_i|x)$, $p(s_i, s_j|x)$ etc.

A. Why do we need them?

a) Inference with locally additive loss $\Rightarrow$ maximum marginal decision $\rightarrow$ $p(s_i|x)$ i.e. ~~???~~ posterior marginals

b) Supervised learning of model parameters u_i, u_{ij} of the prior p.d.

- training data $\mathcal{T} = \{s^1, \dots, s^\ell\} \Leftrightarrow \beta(s)$
- Maximum likelihood estimator

$$L(u) = \sum_{s \in K^{|V|}} \beta(s) \log p_u(s) \rightarrow \max_u$$

write $p_u(s) = \frac{1}{Z(u)} \exp\langle \vec{\varphi}(s), \vec{u} \rangle \Rightarrow$

$$L(\vec{u}) = \sum_{s \in K^{|V|}} \beta(s) \langle \vec{\varphi}(s), \vec{u} \rangle - \log Z(\vec{u}) \rightarrow \max_{\vec{u}}$$

$$L(\vec{u}) = \sum_{s \in K^{|V|}}' \beta(s) \langle \vec{\varphi}(s), \vec{u} \rangle - \log \sum_{s \in K^{|V|}}' \exp \langle \vec{\varphi}(s), \vec{u} \rangle$$

- $L(\vec{u})$ is concave in $\vec{u}$
- $$\nabla L(\vec{u}) = \sum_{s \in K^{|V|}}' \beta(s) \vec{\varphi}(s) - \sum_{s \in K^{|V|}}' p_u(s) \vec{\varphi}(s)$$

$$= \mathbb{E}_\beta(\vec{\varphi}) - \mathbb{E}_u(\vec{\varphi}) = 0$$

i.e. empirical marginals and model marginals must coincide

$$p_u(s_i, s_j) = \beta(s_i, s_j)$$

- problem: given u_i, u_{ij}, compute marginals $p_u(s_i, s_j)$ for all edges $ij \in E$.

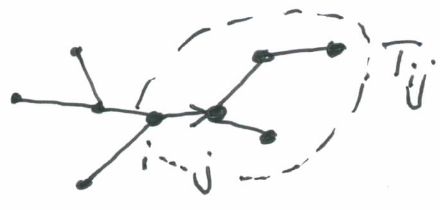
B. Belief propagation

Remember: ch I, sec. 11 computing marginals of an HMM on a tree $T = (V, E)$

$$p(s) = \frac{1}{Z} \prod_{i \in V} f_i(s_i) \prod_{ij \in E} f_{ij}(s_i, s_j) \overset{!}{=} \prod_{i \in V} p(s_i) \prod_{ij \in E} \frac{p(s_i, s_j)}{p(s_i) p(s_j)}$$

auxilliary numbers

$\varphi_{ij}(s_i) \iff$



$$p(s_i) \sim f(s_i) \prod_{j \in N_i} \varphi_{ij}(s_i)$$

$$p(s_i, s_j) \sim f_i(s_i) f_{ij}(s_i, s_j) f_j(s_j) \prod_{\ell \in N_i \setminus j} \varphi_{i\ell}(s_i) \prod_{m \in N_j \setminus i} \varphi_{jm}(s_i)$$

$$\hookrightarrow \frac{p(s_i, s_j)}{p(s_i)\,p(s_j)} = \frac{f_{ij}(s_i, s_j)}{\varphi_{ij}(s_i)\,\varphi_{ji}(s_j)}$$

- This is an equivalent transform in the $(+,\cdot)$-domain!
- recursive definition of the φ-s

$$\varphi_{ij}(s_i) = \sum_{s_j}{}' f_{ij}(s_i, s_j)\, f_j(s_j) \prod_{\ell \in \mathcal{N}_j \setminus i} \varphi_{j\ell}(s_j)$$

<u>BP for general graphs</u> (aka message passing)

- repeatedly recalculate φ-s according to the formula above until (hopefully) a fixpoint φ^* is reached
- estimate marginals by

$$p(s_i) \sim f_i(s_i) \prod_{j \in \mathcal{N}_i} \varphi_{ij}^*(s_i)$$

$$p(s_i, s_j) \sim p(s_i)\,p(s_j) \frac{f_{ij}(s_i, s_j)}{\varphi_{ij}^*(s_i)\,\varphi_{ji}^*(s_j)}$$

<u>Remarks</u>

1) BP quite often gives a reasonable estimate of unary marginals. However, it inherently fails to estimate pairwise marginals

2) In log-domain, when replacing $(+,\cdot)$ by $(\min,+)$ this leads to an approx. algorithm for solving $(\min,+)$-problems

C. Sampling

Let $F: K^{|V|} \to \mathbb{R}$ be a random variable. How can we estimate its expectation $\mathbb{E}_p[F] = \sum_{s \in K^{|V|}} p(s) F(s)$?

- generate an i.i.d sample s^j, $j = 1, \dots, \ell$ by $p(s)$
- $\mathbb{E}_p(F) \approx \frac{1}{\ell} \sum_{j=1}^{\ell} F(s^j)$

Sampling from $p(s)$? Theorem 1, sec. 1, ch I $\Rightarrow$ design a homogeneous Markov chain with transition matrix $T(s|s')$, $s, s' \in K^{|V|}$ s.t.

- the chain is irreducible and a-periodic
- its stationary p.d. is $p(s)$

Practically:

- Design a set of simpler trans. matr. B_m, $m \in M$ s.t. $p(s)$ is stationary for all of them
- compose T by

$$T = \prod_{m \in M} B_m \quad \text{or} \quad T = \sum_{m \in M} \alpha_m B_m$$

- prove that T is irreducible and a-periodic

Gibbs sampler Design B_i, $i \in V$ by

$$B_i(s|s') = \begin{cases} 0 & \text{if } s_{V \setminus i} \neq s'_{V \setminus i} \\ p(s_i | s'_{V \setminus i}) = p(s_i | s'_{N_i}) & \text{otherwise} \end{cases}$$

- stationarity of $p(s)$

$$\sum_{s' \in K^{|V|}} B_i(s|s')\,p(s') = \sum_{k \in K} p(s_i \mid s_{V\setminus i})\, p(s_i' = k, s_{V\setminus i})$$

$$= p(s_i \mid s_{V\setminus i})\, p(s_{V\setminus i}) = p(s)$$

- $T = \prod_{i \in V} B_i$ and $T = \sum_{i \in V} \alpha_i B_i$ are irreducible and a-periodic if $p(s) > 0,\ \forall s \in K^{|V|}$

Remarks

1) Gibbs sampler is easy to implement

2) Gibbs samplers are very slow: often long „burn in time" and „slow mixing".