14. Computing marginal probabilities for GRFs

Consider a GRF for pairs (x,s) on a graph (V,E), where $x: V \to F$ is a field of features and $s: V \to K$ is a field of hidden variables

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

Computing its marginal probabilities on nodes and edges like

$$p_u(s_i),\ p_u(s_i \mid x),\ p_u(s_i, s_j),\ p_u(s_i, s_j \mid x)$$

for $i,j \in V$, $\{i,j\} \in E$ is needed for

(1) Inference with locally additive loss functions (e.g. Hamming distance) $\Rightarrow$ optimal decision is based on $p(s_i \mid x)$

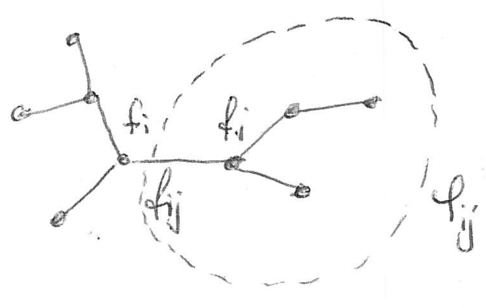
(2) Learning model parameters u_i, u_{ij}. This requires to compute node and edge marginals from u-s and vice versa. (see next section)

Computing the partition function $Z(u)$ and marginal prob's for general GRFs is <u>NP-hard</u>. We have to rely on approximation algorithms.

A. Belief propagation

Recall Sec. 10, computing marginal prob's of a Markov model on a tree:

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



$$= \prod_{i \in V} p(s_i) \prod_{ij \in E} \frac{p(s_i, s_j)}{p(s_i)\, p(s_j)}$$

Hence,
$$p(s_i) \propto f_i(s_i) \prod_{j \in \mathcal{N}_i} \varphi_{ij}(s_i)$$
$$p(s_i, s_j) \propto f_i(s_i) f_{ij}(s_i, s_j) f_j(s_j) \prod_{l \in \mathcal{N}_i / j} \varphi_{il}(s_i) \prod_{m \in \mathcal{N}_j / i} \varphi_{jm}(s_j)$$

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

- recall that φ-s are defined on oriented edges
- the above formula is an equivalent transformation in the $(+, \times)$-domain.

The recursive definition of the φ-s is
$$\varphi_{ij}(s_i) = \sum_{s_j \in K} f_{ij}(s_i, s_j) f_j(s_j) \prod_{l \in \mathcal{N}_j \setminus i} \varphi_{jl}(s_j)$$

Belief propagation for random fields on general graphs (aka message passing):

(1) repeatedly recompute φ-s according to the formula above, until (hopefully) a fixpoint φ^* is reached

(2) estimate marginals from
$$p(s_i) \propto f_i(s_i) \prod_{j \in \mathcal{N}_i} \varphi^*_{ij}(s_i)$$
$$p(s_i, s_j) \propto f_{ij}(s_i, s_j) \frac{p(s_i)\, p(s_j)}{\varphi^*_{ij}(s_i)\, \varphi^*_{ji}(s_j)}$$

Remark 1

(1) Quite often BP gives reasonable estimates for unary marginals. However, it inherently fails to estimate pairwise marginals.

(2) In the log-domain (i.e. replacing $(+,\times)$ by $(\min,+)$) this gives an approximation algorithm for solving $(\min,+)$-problems ▣

B. Sampling

Let $S = \{S_i \mid i \in V\}$ be a K-valued random field with joint p.d. $p(s)$ and let $F: K^V \to \mathbb{R}$ be a random variable. How can we estimate its expectation

$$\mathbb{E}_p(F) = \sum_{s \in K^V} p(s) F(s)$$

by sampling?

- generate an i.i.d. sample $\{s^j \in K^V \mid j = 1, \ldots, \ell\}$ of realisations from $p(s)$
- estimate the expectation by $\mathbb{E}_p(F) \approx \frac{1}{\ell} \sum_{j=1}^{\ell} F(s^j)$

How to sample from $p(s)$? Theorem 1, Sec. 1 $\Rightarrow$ design a homogeneous Markov chain with transition probability $T(s \mid s')$, $s, s' \in K^V$ s.t.

(a) the chain is irreducible and a-periodic

(b) its stationary p.d. is $p(s)$

In practice:

- Design a set of simple (sparse) transition prob. matrices B_m, $m \in M$ s.t. $p(s)$ is stationary for all of them.
- Compose T by

$$T = \prod_{m \in M} B_m \quad \text{or} \quad T = \sum_{m \in M} \alpha_m B_m \quad \left(\alpha_m \geq 0, \sum_{m \in M} \alpha_m = 1\right)$$

- prove that T is irreducible and a-periodic.

Gibbs Sampler (for GRFs)

Design B_i, $i \in V$ by

$$B_i(s|s') = \begin{cases} 0 & \text{if } s_{V \setminus i} \equiv s'_{V \setminus i} \\ p(s_i | s'_{V \setminus i}) \overset{!}{=} p(s_i | s'_{\mathcal{N}_i}) & \text{otherwise} \end{cases}$$

Stationarity of $p(s)$

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

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

It is easy to see that $T = \prod_{i \in V} B_i$ and $T = \sum_{i \in V}' \alpha_i B_i$ are irreducible and a-periodic if $p(s)$ is strictly positive.

Remark 2

- A Gibbs sampler is easy to implement
- Gibbs samplers are very slow: long „burn-in" time and slow mixing.

C. Mean field approximation

If only unary marginals needed $\Rightarrow$ approximate $p(s)$ by an factorising distribution $q(s) = \prod_{i \in V} q_i(s_i)$ with smallest KL-divergence from p

$$D_{KL}(q \| p) = \sum_{s \in K^V}' q(s) \log \frac{q(s)}{p(s)} \rightarrow \min_q$$

For a GRF on a graph this reads

$$\sum_{i\in V}\sum_{s_i\in K} q_i(s_i)\log q_i(s_i) - \sum_{i\in V}\sum_{s_i\in K} q_i(s_i)\,u_i(s_i) -$$

$$-\sum_{ij\in E}\sum_{s_i,s_j\in K} q_i(s_i)\,q_j(s_j)\,u_{ij}(s_i,s_j) \to \min_{q\geq 0}$$

$$\text{s.t.} \quad \sum_{s_i\in K} q_i(s_i) = 1 \quad \forall\, i\in V$$

This can be solved approximately e.g. by block-coordinate ascent.