16. Unsupervised learning for GRFs

Consider a graph $(V \cup H, E)$, where V are „visible" nodes and H are „hidden nodes. Let $S = \{S_i \in K \mid i \in V \cup H\}$ be a field of K-valued random variables with distribution

$$p_u(s) = \frac{1}{Z(u)} \exp \langle \varphi(s_V, s_H), u \rangle$$

such that it represents a GRF w.r.t. to the graph $(V \cup H, E)$. We want to learn its parameters u given a training sample of realisations of the variables indexed by the visible nodes

$$T = \{ s_V^j \in K^V \mid j = 1, \dots, m \}.$$

Let us consider the log-likelihood for one example $s_V \in T$

$$\log p_u(s_V) = \log \sum_{s_H \in K^H} p_u(s_V, s_H) =$$

$$= \log \sum_{s_H \in K^V} e^{\langle \varphi(s_V, s_H), u \rangle} - \log Z(u)$$

Computing its gradient we get

$$\nabla_u \log p_u(s_V) = \mathbb{E}_u [\varphi \mid s_V] - \mathbb{E}_u [\varphi]$$

This requires to compute marginal statistics both for the uncoditioned case and conditioned on s_V.

<u>Q</u> Does the EM-algorithm helps here?

$$\log p_u(s_V) = \log \sum_{s_H \in K^H} p_u(s_V, s_H) \geqslant \sum_{s_H \in K^H} \alpha(s_H \mid s_V) \log \frac{p_u(s_V, s_H)}{\alpha(s_H \mid s_V)}$$

Thus, the EM-algorithm reads

E-step: $\alpha^{(t)}(s_H | s_V) = p_{u^{(t)}}(s_H | s_V)$

M-step:

$$\frac{1}{|\mathcal{T}|} \sum_{s_V \in \mathcal{T}} \sum_{s_H \in K^H} \alpha(s_H | s_V) \log p_u(s_V, s_H) =$$

$$= \frac{1}{|\mathcal{T}|} \sum_{s_V \in \mathcal{T}} \sum_{s_H \in K^H} \alpha(s_H | s_V) \langle \Phi(s_V, s_H), u \rangle - \log Z(u) \to \max_u$$

Again, computing the gradient of the objective function in the M-step is not tractable.