15. Supervised parameter learning for GRFs

A. Generative learning

- $S = \{S_i \mid i \in V\}$ is a K-valued Gibbs random field on a graph (V,E) with joint p.d.

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

- T is an i.i.d. training sample

$$T = \{ s^j \in K^V \mid j = 1, \dots, \ell \}$$

Task: Estimate unary and pairwise potentials (i.e. model parameters) u_i, u_{ij} from training data

Let us first consider the maximum likelihood estimator

$$L(u) = \frac{1}{\ell} \sum_{s\in T}' \log p_u(s) \to \max_u$$

Using the representation of GRFs as exponential families (see Sec. 6) we get

$$L(u) = \frac{1}{\ell} \sum_{s\in T}' \log \frac{1}{Z(u)} e^{\langle \varphi(s), u\rangle}$$

$$= \frac{1}{\ell} \sum_{s\in T}' \langle \varphi(s), u \rangle - \log \sum_{s\in K^V}' e^{\langle \varphi(s), u\rangle} \to \max_u$$

The task has the structure

$$\langle v, u \rangle - g(u) \to \max_u$$

and $g(u)$ is convex

Can we solve it by gradient ascent?

$$\nabla g(u) = \nabla \log Z(u) = \frac{1}{Z(u)} \sum_{s \in K^V} \varphi(s)\, e^{\langle \varphi(s), u \rangle} = \mathbb{E}_u(\varphi)$$

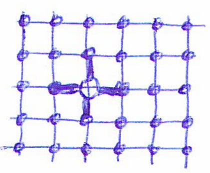
Computing the gradient of g thus requires to compute statistics of φ, i.e. to compute unary and pairwise marginal probabilities.

<u>Remark 1</u> Why is the task easy to solve if (V,E) is a chain or a tree? Because there is a closed form formula for computing the factors of the random field (i.e. e^{u_i}, $e^{u_{ij}}$) from given pairwise marginal statistics. ■

<u>Pseudo-likelihood estimator</u> (Besag, 1975)

Recall from Sec. 14: The Gibbs sampler for a GRF is defined by its conditional distributions

$$p_u(s_i \mid s_{\mathcal{N}_i}), \quad i \in V, \quad s_i \in K$$



and in turn defines the joint distribution $p_u(s)$.

<u>Idea:</u> use the pseudo-likelihood estimator

$$L_p(u) = \frac{1}{\ell} \sum_{s \in T} \sum_{i \in V} \log p_u(s_i \mid s_{\mathcal{N}_i}) \to \max_u$$

$$\nabla L_p(u) = \frac{1}{\ell} \sum_{s \in T} \sum_{i \in V} \nabla \log p_u(s_i \mid s_{\mathcal{N}_i})$$

where

$$\log p_u(s_i \mid s_{\mathcal{N}_i}) = \log \frac{\exp\left[u_i(s_i) + \sum_{j \in \mathcal{N}_i} u_{ij}(s_i, s_j)\right]}{\sum_{s_i \in K} \exp\left[\text{——— " ———} \right]}$$

$$= u(s_i) + \sum_{j \in N_i}' u_{ij}(s_i, s_j) - \log \sum_{k \in K}' \exp\Big[u_i(k) + \sum_{j \in N_i}' u_{ij}(k, s_j)\Big]$$

- is a concave function of u,
- its gradient is easy to compute

Theorem 1 (w10) proof

The pseudo-likelihood estimator is consistent for GRFs but has a higher variance than the MLE. ■

B. Discriminative learning

- X, S a pair of F-valued and K-valued random fields on a graph (V, E) with p.d.

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

- loss function $\ell(s, s') = \sum_{i \in V}' \mathbb{I}\{s_i \neq s_i'\}$

- i.i.d. training sample $\mathcal{T} = \{(x^j, s^j) \mid x^j \in F^V, s^j \in K^V, j = 1, .., m\}$

Task: Estimate unary and pairwise potentials by minimising the empirical risk on the training data

$$R(u, \mathcal{T}) = \frac{1}{m} \sum_{j=1}^{m}{}' \ell\big(s^j, \operatorname*{argmax}_{s' \in K^V} p_u(x, s')\big)$$

$$= \frac{1}{m} \sum_{j=1}^{m}{}' \ell\big(s^j, \operatorname*{argmax}_{s \in K^V} \langle \Phi(x^j, s), u \rangle\big) \to \min_u$$

The objective function is neither differentiable nor convex.
⇒ replace the true loss by some surrogate loss, e.g. margin rescaling loss

$$\tilde{R}(u, T) = \frac{1}{m} \sum_{j=1}^{m} \max_{s \in K^V} \left[\ell(s^j, s) - \langle \varphi(x^j, s^j), u \rangle + \langle \varphi(x^j, s), u \rangle \right]$$

- The objective function is now convex in u,
- Computing its subgradient amounts to solve a (Max,+)-problem for each example (x^j, s^j) in T.