GRAPHICAL MARKOV MODELS
EXAM WS2017 (25P)

Assignment 1. (3p)
Consider a random walk on an undirected graph (V, E) with transition probabilities

$$p(s_t = v \mid s_{t-1} = u) = \begin{cases} \frac{1}{\deg(u)} & \text{if } \{u, v\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\pi_u = \frac{\deg(u)}{2|E|}$ is a stationary distribution.

Assignment 2. (6p)
Consider a Hidden Markov Model on a chain for pairs of sequences $x \in F^n$, $s \in K^n$, where $K = \{1, 2 \ldots, |K|\}$, i.e., the hidden states are integer numbers. Suppose that the loss function for inference is

$$\ell(s, s') = \sum_{i=1}^{n} |s_i - s'_i|.$$

a) Deduce the optimal inference strategy, i.e., the strategy that minimises the average loss.

b) Find an algorithm for the proposed inference strategy. What complexity has it?

Assignment 3. (7p)
Suppose you want to learn the transition probabilities of a homogeneous Markov chain model

$$p(s) = p(s_1) \prod_{i=2}^{n} p(s_i \mid s_{i-1})$$

with discrete valued states $s_i \in K$. You are given an i.i.d. sample $\mathcal{T}$ of sequences s as training data and decide to apply the maximum likelihood estimator. Unfortunately, you discover that the training data are corrupted in the following way. The j-th element is missing in all sequences $s \in \mathcal{T}$. Propose a remedy.

Assignment 4. (4p)
Consider the following clustering problem. N points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ in an Euclidean space should be grouped into K clusters ($K < N$). The objective function for the clustering is defined as the sum of all cluster costs. The cost of a cluster is the sum of squared distances $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2$ for all pairs (i, j) of its points. The task is to find the clustering with minimal cost. Formulate this task as a $(\min, +)$ problem on an appropriately chosen graph. Describe the graph, the label set and the objective function. Is this $(\min, +)$ problem submodular?

Assignment 5. (5p)
Consider an Ising model on an undirected graph (V, E), i.e. a random field with binary valued variables $s_i \in \{0, 1\}$, $i \in V$ and the probability distribution

$$p(s) = \frac{1}{Z} \exp\Big[-\alpha \sum_{ij \in E} |s_i - s_j|\Big].$$

Given a sample of i.i.d. realisations $\mathcal{T} = \{s^\ell \mid \ell = 1, 2 \ldots, m\}$, you want to estimate the parameter α by using the pseudo-likelihood estimator. Explain how to do that.