GRAPHICAL MARKOV MODELS (WS2023)
4. SEMINAR

Assignment 1. (breakpoint detection) Consider the following probabilistic model for real valued sequences $\boldsymbol{x} = (x_1, \ldots, x_n)$, $x_i \in \mathbb{R}$ of fixed length n. Each sequence is a combination of a leading part $i \leqslant k$ and a trailing part $i > k$. The boundary $k = 1, \ldots, n$ is random with some categorical distribution $\boldsymbol{\pi} \in \mathbb{R}^n_+$, $\sum_k \pi_k = 1$. The p.d.s for the leading and trailing parts of the sequence arise from two homogeneous HMM models:

$$p(x_{1:k}) = \sum_{s_{1:k}} p_1(x_{1:k}, s_{1:k}) \text{ and } p(x_{k+1:n}) = \sum_{s_{k+1:n}} p_2(x_{k+1:n}, s_{k+1:n})$$

The HMMs p_1 and p_2 and the distribution $\boldsymbol{\pi}$ are known. Find an algorithm for inferring the boundary k for a given sequence $\boldsymbol{x}$, assuming that the loss function is $\ell(k, k') = (k - k')^2$.

Hints:
(i) You will need to find an efficient algorithm for computing $p(k \mid \boldsymbol{x})$, $\forall k = 1, \ldots, n$ for a given (fixed) sequence $\boldsymbol{x}$. What complexity has it?
(ii) Suppose you have already computed the n probabilities $p(k \mid \boldsymbol{x})$. What is the optimal prediction k^* for a quadratic loss function $\ell(k, k') = (k - k')^2$?

Assignment 2. Let X be a discrete random variable taking values $x \in \mathcal{X}$ from a finite set $\mathcal{X}$. Let us consider the exponential family of distributions

$$p(x; u) = \exp\big[\langle \phi(x), u \rangle - \log Z(u)\big],$$

where $\phi(x) \in \mathbb{R}^n$ is the sufficient statistics and $u \in \mathbb{R}^n$ is the vector of natural parameters.

a) Find the geometrical condition on the sufficient statistics (in $\mathbb{R}^n$), which guaranties that $p(x; u) \not\equiv p(x; u')$ whenever $u \neq u'$.

b) Let us now consider the family of homogeneous Markov chain models on strings $s = (s_1, \ldots, s_n)$ with elements $s_i \in K$, in a finite alphabet. Furthermore, the family contains only models with strictly positive probabilities, i.e. $p(s) > 0$, $\forall s \in K^n$. This family is an exponential family with distributions given by

$$p(s; U) = \exp\big[\langle \Phi(s), U \rangle - \log Z(U)\big],$$

where the sufficient statistics $\Phi(s)$ is a $K \times K$ matrix defined by

$$\Phi_{kk'}(s) = \sum_{i=2}^{n} [\![s_{i-1} = k' \wedge s_i = k]\!]$$

and U is the $K \times K$ matrix of natural parameters. Is the condition you found in a), fulfilled for this family?