GRAPHICAL MARKOV MODELS (WS2021)
3. SEMINAR

Assignment 1. Consider the task of predicting the sequence of hidden states $s = (s_1, \dots, s_n)$ for an HMM, given the sequence of observed features $x = (x_1, \dots, x_n)$. The loss function is the Hamming distance, i.e.

$$\ell(s, s') = \sum_{i=1}^{n} [\![s_i \neq s'_i]\!]$$

a) Give the optimal predictor and the algorithms needed for its implementation.
b) Can it happen that the predicted sequence s will have zero probability in the model? If yes, propose an augmented loss function that prevents this and derive the corresponding inference algorithm.

Assignment 2. We want to estimate the parameters of a Markov chain model

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

from a training set $\mathcal{T}^m = \{s^\ell \mid s^\ell \in K^n, \ell = 1, \dots, m\}$ by the maximum likelihood estimator. Give the formula for the log-likelihood $\log \mathbb{P}(\mathcal{T}^m)$, substitute the expression for the probabilities $p(s^\ell)$ and prove that the maximum is achieved at

$$p(s_i = k \mid s_{i-1} = k') = \frac{\alpha_i(k', k)}{\sum_k \alpha_i(k', k)},$$

where $\alpha_i(k', k)$ denotes the frequency of the event $s_{i-1} = k'$ and $s_i = k$ in the training data.

Assignment 3. Consider the following probabilistic model for real valued sequences $x = (x_1, \dots, x_n)$, $x_i \in \mathbb{R}$ of fixed length n. Each sequence is a combination of a leading part $i \leqslant k$ and a trailing part $i > k$. The boundary $k = 1, \dots, n$ is random with some categorical distribution $\pi \in \mathbb{R}^n_+$, $\sum_k \pi_k = 1$. The values x_i, in the leading and trailing part are statistically independent and distributed with some probability density function p_1 and p_2 respectively. Altogether the distribution for pairs (x, k) reads

$$p(x, k) = \pi_k \prod_{i=1}^{k} p_1(x_i) \prod_{j=k+1}^{n} p_2(x_j). \tag{1}$$

The densities p_1 and p_2 are known. Given an i.i.d. sample of sequences $\mathcal{T}^m = \{x^\ell \in \mathbb{R}^n \mid \ell = 1, \dots, m\}$, the task is to estimate the unknown boundary distribution π by the EM-algorithm.

a) The E-step of the algorithm requires to compute the values of auxiliary variables $\alpha_\ell^{(t)}(k) = p(k \mid x^\ell)$ for each example x^ℓ, given the current estimate $\pi^{(t)}$ of the boundary distribution. Give a formula for computing these values from model (1).

b) The M-step requires to solve the optimisation problem

$$\frac{1}{m}\sum_{\ell=1}^{m}\sum_{k=1}^{n}\alpha_\ell^{(t)}(k)\log p(x^\ell,k)\to\max_\pi .$$

Substitute the model (1) and solve the optimisation task.

Assignment 4. Let $s=(s_1,\ldots,s_n)$, be a sequence of K-valued random variables. Suppose that $v_i(k,k')$, $i=2,\ldots,n$, $k,k'\in K$ is a system of pairwise probabilities associated with consecutive pairs s_{i-1}, s_i. Consider the set $\mathcal{P}(v)$ of all joint probability distributions $p(s)$, which have v as pairwise marginals, i.e.

$$\sum_{s\in K^n}p(s)\delta_{s_{i-1}k}\delta_{s_ik'}=v_i(k,k')\quad\forall i=2,\ldots,n,\ \forall k,k'\in K.$$

We want to find the distribution with highest entropy

$$H(p)=-\sum_{s\in K^n}p(s)\log p(s)$$

in $\mathcal{P}(v)$. Prove that the unique maximiser is the Markov chain model defined by the pairwise marginals $\boldsymbol{v}$.
Hint: Formulate and solve the constrained optimisation task by using its Lagrange function.