GRAPHICAL MARKOV MODELS (WS2015)
3. SEMINAR

Assignment 1. Consider a fully specified HMM for sequence pairs (x, s) of length n given by

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

In addition, it is decided that the loss function (penalty for inference errors) is the Hamming distance between the true sequence of hidden states and the one inferred from the observation.

a) Describe the optimal inference strategy (i.e. the strategy which minimises the expected loss).

b) Some transition probabilities of the HMM model are zero. Therefore, some sequences of hidden states have zero probability. Can it happen that the strategy you have found in a) will infer a sequence of hidden states with zero probability? Give a simple example.

c*) Suppose we want to "repair" the problem in the following way. We modify the loss function by adding a term which ensures that inferring a sequence of hidden states with zero probability will cause a very high penalty:

$$c(s, s') = d_h(s, s') + \chi(s'),$$

where s denotes the true sequence, s' denotes the inferred sequence and χ is defined as follows

$$\chi(s) = \begin{cases} 0 & \text{if } p(s) > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Derive the optimal inference strategy for this modified loss.

Assignment 2*. Suppose that a regular language $\mathcal{L}$ of strings over the finite alphabet Σ is described by a non-deterministic finite-state machine. Given a string $y \notin \mathcal{L}$, the task is to find the string $x \in \mathcal{L}$ with smallest Hamming distance to y, i.e.

$$x^* = \arg\min_{x \in \mathcal{L}} d_h(x, y),$$

where d_h denotes the Hamming distance. Construct an efficient algorithm for this task.

Assignment 3. Derive an EM-algorithm for learning the following conditional independent probability model (sometimes called "naive Bayes probabilistic model")

$$p(x, y, k) = p(x \mid k) p(y \mid k) p(k),$$

where $x \in X$, $y \in Y$, $k \in K$ and all three sets are finite. The training data for learning are i.i.d. realisations of pairs (x, y) with empirical distribution $p^*(x, y)$ (the corresponding values of k are not observed).

Assignment 4. (Chow, Liu, 1968) Consider a collection of K-valued random variables s_i, $i \in V$. It is known that its joint probability distribution is a Markov model on a tree (V, E), i.e.

$$p(s) = \prod_{\{i,j\} \in E} p(s_i, s_j) \Big/ \prod_{i \in V} p(s_i)^{n_i - 1},$$

where n_i denotes the degree of vertex i. However, neither the edge set E nor the pairwise marginal probabilities $p(s_i, s_j)$ are known. Given an i.i.d. sample of realisations $\mathcal{T} = \{s^j \mid s^j \in K^V, j = 1, \dots, \ell\}$, the task is to estimate both the tree structure and the pairwise marginal probabilities on its edges.

a) Explain how to estimate the marginal pairwise probabilities of the model by the maximum likelihood estimator, provided that the tree structure is known.

b)* Using the result of a), express the likelihood of the training data as a function of the unknown tree structure. Prove that the optimal tree structure, i.e. the one which has maximal likelihood, can be found by a maximal spanning tree algorithm.