GRAPHICAL MARKOV MODELS (WS2011)
4. SEMINAR

Assignment 1. Suppose that a regular language L of strings over the alphabet Σ is described by means of a non-deterministic finite-state machine. Given a string $y \notin L$, the task is to find the string $x \in L$ with smallest Hamming distance to y, i.e.

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

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

Assignment 2. Suppose we are given an HMM. 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 (the strategy which minimises the risk, i.e. the expected loss).
b) It turns out that some transition probabilities in the HMM model are zero. Some sequences of hidden states have zero probability as a consequence. Can it happen that the inference strategy discussed in a) will output 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 simply change the loss function so that 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

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

Derive the optimal inference strategy for this modified loss.

Assignment 3. Consider an HMM $p(x, s) = p(x \mid s)p(s)$ for pairs of sequences (x, s), where $x \in F^n$ and $s \in K^n$:

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

The transition probabilities and emission probabilities are unknown. It is however known, that the model is translational invariant. A sample of i.i.d. training data $\mathcal{T} = \{(x^j, s^j) \mid j = 1, \ldots, \ell\}$ is given for learning. Modify the formulae for supervised learning (unrestricted case) for this situation. Prove correctness.

Assignment 4. Consider the following variation of the previous assignment. The observable features are no longer discrete – instead they are real numbers, i.e., $F = \mathbb{R}$. It is however known, that the emission probabilities are Gaussian distributions. For each hidden state $k \in K$ there is a Gaussian p.d.

$$p(f \mid k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(f-\mu_k)^2}{2\sigma_k^2}}$$

with unknown mean and variance. Derive formulae for learning these model parameters.