GRAPHICAL MARKOV MODELS (WS2013)
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 discussed 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 $x \in L$ 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. Consider an HMM for pairs of sequences (x, s), $x \in F^n$, $s \in K^n$ as in Assignment 1 with unknown transition probabilities and emission probabilities. It is however known, that the model is homogeneous. A sample of i.i.d. training data $\mathcal{T} = \{(x^j, s^j) \mid j = 1, \dots, \ell\}$ is given for learning. Modify the formulae for the maximum likelihood estimate (cf. sec. 7 in lecture notes) 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 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.

Assignment 5. 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).