STATISTICAL MACHINE LEARNING (WS2017)
SEMINAR 6

Assignment 1. Consider the class of Naive Bayes Models for feature vectors $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}$ and hidden labels $k \in K$ given by

$$p(\boldsymbol{x}, k) = p(k) \prod_{i=1}^{n} p_i(x_i \mid k).$$

Here, K is a finite set and, for simplicity, we assume that the set of vectors $\mathcal{X}$ is also finite. You may assume e.g. that $\mathcal{X} = V^n$ for some finite set V. Derive an EM-algorithm for this model class given training data $\mathcal{T}^m = \{\boldsymbol{x}^j \mid j = 1, \ldots, m\}$.
a) Notice, that the model is given by the $n|K|$ conditional distributions $p_i(x_i \mid k)$ and the prior distribution for the hidden label $p(k)$. These are the unknown model parameters.

b) Give a formula for computing the posterior probability $p(k \mid \boldsymbol{x})$, which is needed in the E-step of the algorithm.

c) Deduce that the optimisation task in the M-step of the algorithm decomposes into independent optimisation tasks for the conditional distributions $p_i(x_i \mid k)$, $i = 1, \ldots, n$ and for the prior distribution of the hidden label $p(k)$.

d) Complete the derivation by showing how to solve these independent optimisation tasks.

Assignment 2. Let us consider a Markov chain model for sequences $\boldsymbol{s} = (s_1, \ldots, s_n)$ of length n with states $s_i \in K$ from a finite set K. Its joint probability distribution is given by

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

The conditional probabilities $p(s_i \mid s_{i-1})$ and the marginal probability $p(s_1)$ for the first element are known.

Let $A \subset K$ be a subset of states and let $\mathcal{A} = A^n$ denote the set of all sequences $\boldsymbol{s}$ with $s_i \in A$ for all $i = 1, \ldots, n$. Find an efficient algorithm for computing the probability $p(\mathcal{A})$ of the event $\mathcal{A}$.

Assignment 3. Consider the same Markov model as in the previous assignment. You are given its most probable sequence $\boldsymbol{s}^* \in \arg\max_{\boldsymbol{s} \in K^n} p(\boldsymbol{s})$. The task is to find the most probable sequence $\boldsymbol{s}$ differing from $\boldsymbol{s}^*$ in all positions, i.e. $s_i \neq s_i^* \; \forall i = 1, \ldots, n$. Give an algorithm for solving this task.

Assignment 4. (*Gambler's ruin*) Consider a random walk on the set $L = \{0, 1, 2, \ldots, a\}$ starting in some point $x \in L$. The position jumps by either ± 1 in each time period (with

equal probabilities). The walk ends if either of the boundary states $0, a$ is hit. Compute the probability $u(x)$ to finish in state a if the process starts in state x.
Hints:

(1) What are the values of $u(0)$ and of $u(a)$?
(2) Find a difference equation for $u(x)$, $0 < x < a$ by relating it with $u(x-1)$ and $u(x+1)$.
(3) Translate the difference equation into a relation between the successive differences $u(x+1) - u(x)$ and $u(x) - u(x-1)$.
(4) Deduce that the solution is a linear function of x and find its coefficients from the boundary conditions $u(0)$ and $u(a)$.