11. Hidden Markov Models on acyclic graphs

1A. **Definition 1a** Let $T = (V, E)$ be an undirected acyclic graph and s_i, $i \in V$ be K-valued discrete variables associated with its vertices. A p.d. on configurations $s \in K^{|V|}$ is a Markov model if the following holds. Fixing an arbitrary vertex $i_* \in V$ and denoting by $c(i)$ the next vertex in the path connecting $i \in V$ with i_*, the p.d. can be written as

$$p(s) = p(s_{i_*}) \prod_{i \neq i_*} p(s_i \mid s_{c(i)}) .$$

Definition 1b Equivalently, a p.d. $p(s)$ on $s \in K^{|V|}$ is a Markov model if

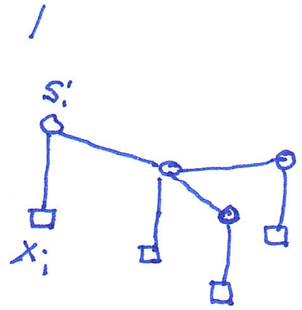
$$p(s) = \prod_{ij \in E} p(s_i, s_j) \Big/ \prod_{i \in V} p^{n_i - 1}(s_i)$$

holds, where n_i is the degree of vertex $i \in V$.

An HMM on an (undirected) acyclic graph $T(V, E)$ is defined on pairs $s \in K^{|V|}$, $x \in F^{|V|}$, where F is a feature space and $p(x, s) = p(x \mid s)\, p(s)$ where

$$p(x \mid s) = \prod_{i \in V} p(x_i \mid s_i) \quad \text{(conditional independent)}$$

$p(s)$ – Markov model on T



B. Consider the following task: Given an observation configuration $x \in F^{|V|}$, compute

$$p(x) = \sum_{s \in K^{|V|}} p(x, s)$$

Substitute the model (Def. 1a) and denote the chosen vertex by $r \in V$ (root).

$$p(x) = \sum_{s \in K^{|V|}} p(s_r)\, p(x_r | s_r) \prod_{j \in V \setminus r} p(s_j | s_{c(j)})\, p(x_j | s_j)$$

The observation x is fixed $\rightarrow$ the problem has the form

$$\sum_{s \in K^{|V|}} \prod_{i \in V} \varphi_i(s_i) \prod_{j \in V \setminus r} f_j(s_j, s_{c(j)})$$

where

$$\varphi_r(s_r) = p(s_r)\, p(x_r | s_r)$$
$$\varphi_i(s_i) \equiv 1 \quad \forall\ i \in V \setminus r$$
$$f_j(s_j, s_{c(j)}) = p(s_j | s_{c(j)}) \cdot p(x_j | s_j)$$

The algorithm re-calculates the φ-s beginning with an arbitrary leaf $b \in V$

$$\varphi_{c(b)}(s_{c(b)}) := \varphi_{c(b)}(s_{c(b)}) \cdot \sum_{s_b \in K} f_b(s_b, s_{c(b)}) \cdot \varphi_b(s_b)$$

and „removes" b. This is repeated until only the root remains. Finally

$$p(x) = \sum_{k \in K} \varphi_r(k)$$

Complexity: $|K|^2 \cdot |E|$

C. The most probable configuration of hidden states (MAP)

The same principle can be applied for the task

$$s^* = \arg\max_{s \in K^{|V|}} p(x,s) = \arg\max_{s \in K^{|V|}} \log p(x,s)$$

This is done by replacing operations

$$\times \to +$$
$$+ \to \max$$

D. Learning the tree structure

Suppose, we are given an i.i.d. sample of state configurations generated by a Markov model on a tree. Can we estimate its parameters and its structure?

$\beta(s)$ - occurence frequency of s in the sample

Log-likelihood for known structure (V, E)

$$\sum_{s \in K^{|V|}} \beta(s) \log p(s) \overset{\text{Def. 18}}{=}$$

$$= \sum_{s \in K^{|V|}} \beta(s) \Big[\sum_{ij \in E} \log p(s_i, s_j) - \sum_{i \in V} (n_i - 1) \log p(s_i) \Big]$$

$$= \sum_{ij \in E} \sum_{s_i, s_j} \beta(s_i, s_j) \log p(s_i, s_j) - \sum_{i \in V} (n_i - 1) \sum_{s_i} \beta(s_i) \log p(s_i)$$

$$\to \max_{p}$$

It can be proved that

$$p(s_i) = \beta(s_i), \quad p(s_i, s_j) = \beta(s_i, s_j)$$

solves the task. It remains to solve

$$\sum_{ij \in E} \sum_{s_i, s_j} \beta(s_i, s_j) \log \beta(s_i, s_j) - \sum_{i \in V} (n_i - 1) \sum_{s_i} \beta(s_i) \log \beta(s_i) \to \max_{E:\ \text{tree}}$$

Both terms depend on the unknown structure.
Rewriting the objective we get

$$\sum_{ij \in E} \left\{ \sum_{s_i, s_j} \beta(s_i, s_j) \log \beta(s_i, s_j) - \sum_{s_i} \beta(s_i) \log \beta(s_i) - \sum_{s_j} \beta(s_j) \log \beta(s_j) \right\} +$$

$$+ \sum_{i \in V} \sum_{s_i} \beta(s_i) \log \beta(s_i)$$

The last term does not depend on the edge structure E. Denoting the values of the curly brackets in the first term by h_{ij}, we get

$$\sum_{ij \in E} h_{ij} \to \max_{E:\ \text{tree}}$$

$\Rightarrow$ Maximum spanning tree problem, which can be easily solved.