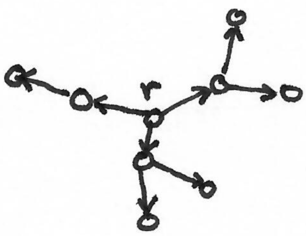
10. Extensions of Markov models: acyclic graphs, uncountable feature and state spaces

Markov models and HMMS discussed so far:

- graph: chain or „comb-like" (HMM)
- state/feature space: finite

A. Hidden Markov models on acyclic graphs

Let $T=(V,E)$ be an undirected, connected acyclic graph. Fixing an arbitrary node $r \in V$, we denote by $\vec{E}_r$ the edge set of the corresponding rooted digraph



Definition 1a Let $T=(V,E)$ be an undirected tree and $\{S_i \mid i \in V\}$ a collection of K-valued random variables. A p.d. for the random field $s \in K^V$ is a Markov model on T if

$$p(s) = p(s_r) \prod_{ij \in \vec{E}_r} p(s_j \mid s_i)$$

holds for <u>any</u> choice $r \in V$ (root). ■

Definition 1b A p.d. $p(s)$ for $s \in K^V$ is a Markov model on T if

$$p(s) = \prod_{\{ij\} \in E} g_{ij}(s_i, s_j)$$

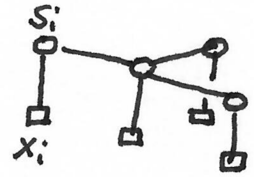
with some functions $g_{ij}: K^2 \to \mathbb{R}_+$. In particular,

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

where n_i denotes the degree of node $i \in V$. ■

An HMM on an undirected tree $T=(V,E)$ is a p.d. for pairs $s\in K^V$, $x\in F^V$ s.t.

- $p(s)$ is a Markov model on T
- $p(x\mid s) = \prod_{i\in V} p(x_i\mid s_i)$



Let us consider inference tasks for HMMs on trees

(a) Given an observation field $x\in F^V$, compute

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

Substituting the model, we get

$$p(x) = \sum_{s\in K^V} p(s_r)\,p(x_r\mid s_r) \prod_{ij\in \vec{E}_r} p(s_j\mid s_i)\,p(x_j\mid s_j),$$

which has the form (for fixed observation)

$$\sum_{s\in K^V} \prod_{i\in V} \varphi_i(s_i) \prod_{ij\in \vec{E}_r} f_{ij}(s_i,s_j)$$

with

$$f_{ij}(s_i,s_j) = p(s_j\mid s_i)\,p(x_j\mid s_j) \quad\text{and}\quad \varphi_i(s_i) = \begin{cases} p(s_r)\,p(x_r\mid s_r) & \text{if } i=r \\ 1 & \text{otherwise}\end{cases}$$

The algorithm recomputes the φ-s starting from an arbitrary <u>leaf</u> $j\in V$. Let $ij\in\vec{E}_r$ be its only incoming edge

$$\varphi_i(s_i) := \varphi_i(s_i) \sum_{s_j\in K} f_{ij}(s_i,s_j)\,\varphi_j(s_j)$$

The leaf is removed thereafter. This is repeated until only the root node r remains. Finally $p(x) = \sum_{s_r\in K} \varphi_r(s_r)$

<u>Complexity</u>: $\mathcal{O}(|K|^2|E|)$

Remark 1 The same approach is used for solving the task

$$s_* \in \operatorname*{argmax}_{s \in K^V} \log p(x,s).$$

Simply by replacing operations $\times \mapsto +$, $+ \mapsto \max$. ∎

(b) Computing marginal probabilities: Given an observation field $x \in F^V$, compute $p(x, s_i)$ $\forall i \in V$, $\forall s_i \in K$.

It follows from Def. 16 that

$$p(x, s_i) = p(s_i)\, p(x_i \mid s_i) \prod_{j \in N_i} p(x_{T_{ij}} \mid s_i),$$

where T_{ij} denotes the subtree given by

$$V(T_{ij}) = \{ m \in V \mid j \in \operatorname{path}(i,m) \}.$$

Let us denote $\varphi_{ij}(s_i) = p(x_{T_{ij}} \mid s_i)$. They fulfil the following system of equations

$$\varphi_{ij}(s_i) = \sum_{s_j \in K} p(s_j \mid s_i)\, p(x_j \mid s_j) \prod_{\substack{\ell \in N_j \\ \ell \neq i}} \varphi_{j\ell}(s_j).$$

Two passes through all edges of T suffice to compute all of them and, consequently all marginals.

Complexity: $\mathcal{O}(|K|^2 |E|)$

B. Uncountable feature space

All previously discussed inference & learning algorithms can be applied if the feature space F is uncountable infinite, provided that

- the conditional distribution densities $p(x_i \mid s_i) = p_{\theta_i}(x_i \mid s_i)$ are given in some parametric model (e.g. normal distribution)

- their parameters can be estimated from corresponding samples.

c. Uncountable state spaces

Inference ~~als~~ problems are not tractable anymore, except for special cases.

Kalman filter $s_i \in \mathbb{R}^n$, $x_i \in \mathbb{R}^m$

$$s_1 \sim \mathcal{N}(\mu_1, Q),$$
$$s_i \mid s_{i-1} \sim \mathcal{N}(A s_{i-1}, Q'),$$
$$x_i \mid s_i \sim \mathcal{N}(H s_i, R),$$

where $A: \mathbb{R}^n \to \mathbb{R}^n$ and $H: \mathbb{R}^n \to \mathbb{R}^m$ are linear mappings, Q, Q', R are covariance matrices and $\mathcal{N}(\mu, \Sigma)$ denotes multivariate normal distributions.

We know that

- product of normal p.d.f.s

$$\mathcal{N}(\mu, A) \cdot \mathcal{N}(\nu, B) = \mathcal{N}(\xi, C)$$

with $\xi = C(A^{-1}\mu + B^{-1}\nu)$, $C = (A^{-1} + B^{-1})^{-1}$

- convolution of normal p.d.f.s

$$\int_{\mathbb{R}^n} dx\, \mathcal{N}(x; \mu, A)\, \mathcal{N}(y-x; \nu, B) = \mathcal{N}(y; \xi, C)$$

with $C = A + B$, $\xi = \mu + \nu$

Hence, $p(s_i \mid x_{1:i})$ is normally distributed. Its parameters can be computed recursively.

Particle filters approximation for the general case

Approximate $p(s_i | x_{1:i})$ by sequential Monte Carlo sampling

1. generate an i.i.d. sample s_1^ℓ, $\ell = 1,..,L$ using
$$p(s_1 | x_1) \sim p(s_1)\, p(x_1 | s_1)$$

2. iterate: given a sample s_{i-1}^ℓ, $\ell = 1,..,L$ generated from $p(s_{i-1} | x_{1:i-1})$, sample s_i^ℓ from
$$s_i^\ell \sim p(x_i | s_i)\, p(s_i | s_{i-1} = s_{i-1}^\ell)$$

The obtained sample s_n^ℓ, $\ell = 1,..,L$ estimates $p(s_n | x_{1:n})$ and can be used to estimate the posterior expectation of a random variable $f(s_n)$ by
$$\mathbb{E}[f | x_{1:n}] = \int_{R^n} ds_n\, f(s_n)\, p(s_n | x_{1:n}) \approx \frac{1}{L} \sum_{\ell=1}^{L} f(s_n^\ell)$$