

Hidden Markov Models

Guillaume Obozinski

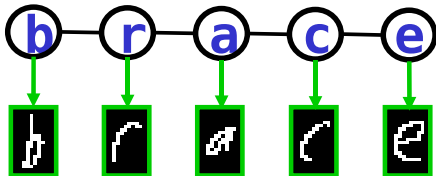
Swiss Data Science Center



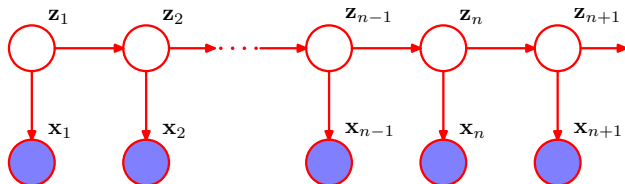
African Masters of Machine Intelligence, 2018-2019, AIMS, Kigali

Hidden Markov Model (HMM)

- voice recognition
- natural language processing
- handwritten character recognition
- modelling biological sequence (protein, DNA)



Hidden Markov Model (HMM)

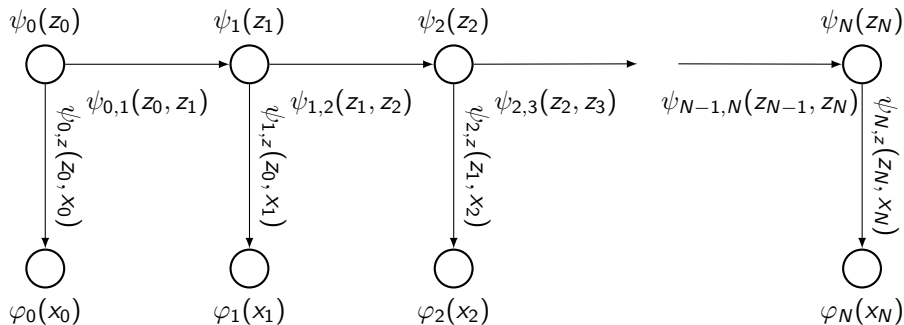


$$p(\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{z}_1, \dots, \mathbf{z}_N) = p(\mathbf{z}_1) \prod_{i=2}^N p(\mathbf{z}_i | \mathbf{z}_{i-1}) \prod_{i=1}^N p(\mathbf{x}_i | \mathbf{z}_i)$$

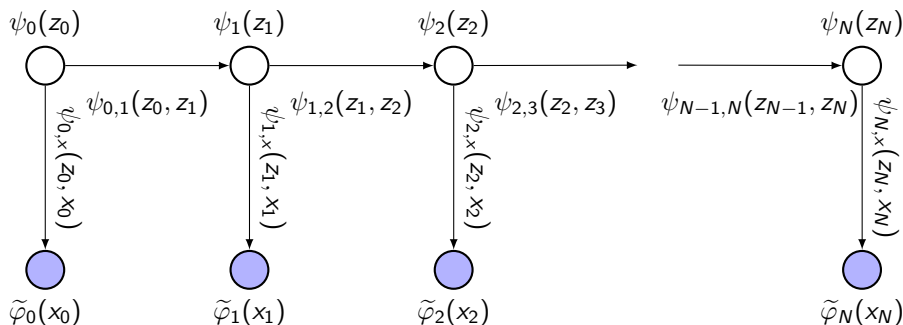
Homogeneous Markov chains

- $\mathbf{z}_i \in \{0, 1\}^K$ state indicator variable $(1, \dots, K)$
- *homogeneous* Markov chain: $\forall n, p(\mathbf{z}_i | \mathbf{z}_{i-1}) = p(\mathbf{z}_2 | \mathbf{z}_1)$
- **emitted symbol** \mathbf{x}_i ($\{0, 1\}^K$) / **observation** (\mathbb{R}^d)

Gibbs model for $p(\mathbf{x}, \mathbf{z})$ in the HMM model

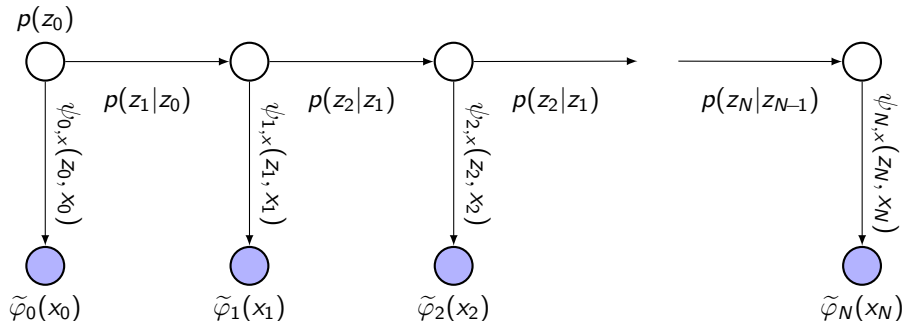


Gibbs model for $p(\mathbf{z} \mid \mathbf{x})$ in the HMM model



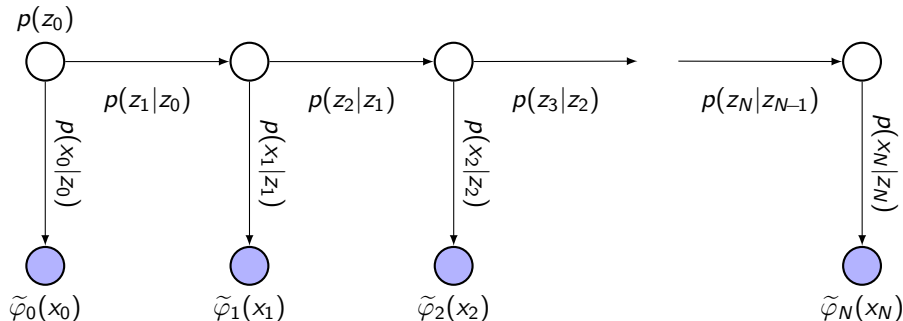
- with $\tilde{\varphi}_i(x_i) = \varphi_i(x_i) \delta(x_i, x_i^{\text{obs}})$

Gibbs model for $p(z \mid \mathbf{x})$ with explicit potentials



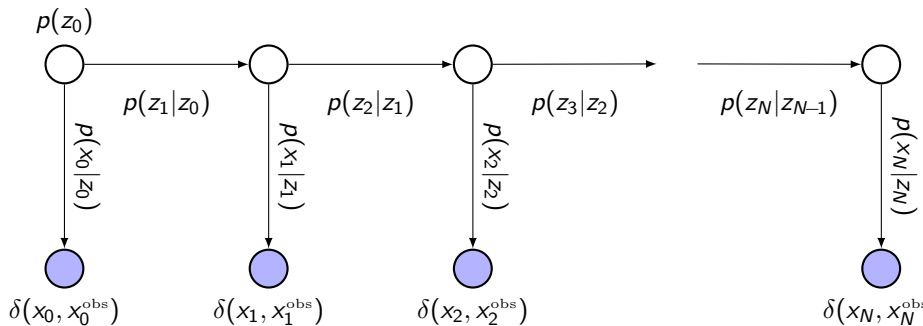
- with $\tilde{\varphi}_i(x_i) = \varphi_i(x_i) \delta(x_i, x_i^{\text{obs}})$

Gibbs model for $p(\mathbf{z} \mid \mathbf{x})$ with explicit potentials



- with $\tilde{\varphi}_i(x_i) = \varphi_i(x_i) \delta(x_i, x_i^{\text{obs}})$

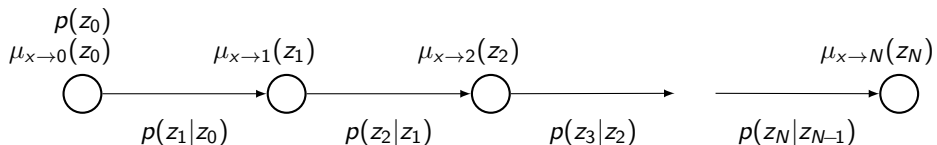
Gibbs model for $p(z | x)$ with explicit potentials



- Marginalize out all the variables x_i

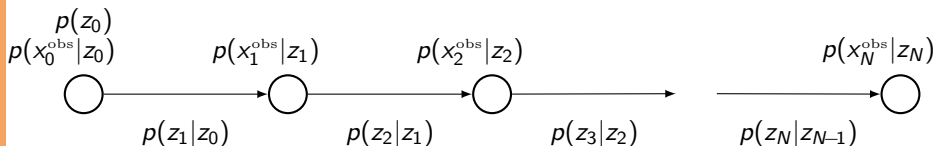
- $$\mu_{x \rightarrow i}(z_i) = \int_{x_i} \psi_{i,x}(z_i, x_i) \tilde{\varphi}_i(x_i) dx_i = \int_{x_i} p(x_i|z_i) \delta(x_i, x_i^{\text{obs}}) dx_i$$

Reduced Gibbs model for $p(\mathbf{z}|\mathbf{x}^{\text{obs}})$



- With $\mu_{x \rightarrow i}(z_i) = p(x_i^{\text{obs}}|z_i)$

Sum-product for the HMM



Messages for the sum-product algorithm

$$\mu_{i-1 \rightarrow i}(z_i) = \sum_{z_{i-1}} p(z_i | z_{i-1}) p(x_{i-1}^{\text{obs}} | z_{i-1}) \mu_{i-2 \rightarrow i-1}(z_{i-1})$$

$$\mu_{i+1 \rightarrow i}(z_i) = \sum_{z_{i+1}} p(z_{i+1} | z_i) p(x_{i+1}^{\text{obs}} | z_{i+1}) \mu_{i+2 \rightarrow i+1}(z_{i+1})$$

Rewriting the sum-product as the α and β recursions

Messages for the sum-product algorithm

$$\mu_{i-1 \rightarrow i}(z_i) := \sum_{z_{i-1}} p(z_i | z_{i-1}) p(x_{i-1}^{\text{obs}} | z_{i-1}) \mu_{i-2 \rightarrow i-1}(z_{i-1})$$

$$\mu_{i+1 \rightarrow i}(z_i) = \sum_{z_{i+1}} p(z_{i+1} | z_i) p(x_{i+1}^{\text{obs}} | z_{i+1}) \mu_{i+2 \rightarrow i+1}(z_{i+1})$$

Messages for the equivalent α and β recursions

$$\alpha_i(z_i) = \mu_{i-1 \rightarrow i}(z_i) p(x_i^{\text{obs}} | z_i) \quad \text{and} \quad \beta_i(z_i) = \mu_{i+1 \rightarrow i}(z_i)$$

$$\alpha_i(z_i) = p(x_i^{\text{obs}} | z_i) \sum_{z_{i-1}} p(z_i | z_{i-1}) \alpha_{i-1}(z_{i-1})$$

$$\beta_i(z_i) = \sum_{z_{i+1}} p(z_{i+1} | z_i) p(x_{i+1}^{\text{obs}} | z_{i+1}) \beta_{i+1}(z_{i+1})$$

Properties of α and β messages

$$\alpha_i(z_i) = p(x_i^{\text{obs}}|z_i) \sum_{z_{i-1}} p(z_i|z_{i-1}) \alpha_{i-1}(z_{i-1})$$

$$\beta_i(z_i) = \sum_{z_{i+1}} p(z_{i+1}|z_i) p(x_{i+1}^{\text{obs}}|z_{i+1}) \beta_{i+1}(z_{i+1})$$

$$\alpha_i(z_i) = p(x_1^{\text{obs}}, \dots, x_i^{\text{obs}}, z_i) \quad \beta_i(z_i) = p(x_{i+1}^{\text{obs}}, \dots, x_N^{\text{obs}}|z_i)$$

Finally one gets the marginals:

$$p(z_i|\mathbf{x}^{\text{obs}}) = \frac{\alpha_i(z_i)\beta_i(z_i)}{p(\mathbf{x}^{\text{obs}})}$$

and

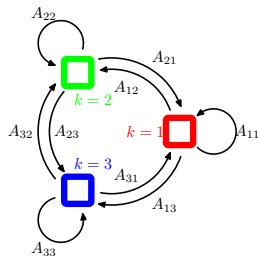
$$p(z_{i-1}, z_i|\mathbf{x}^{\text{obs}}) = \frac{\alpha_{i-1}(z_{i-1})p(x_i^{\text{obs}}|z_i)p(z_i|z_{i-1})\beta_i(z_i)}{p(\mathbf{x}^{\text{obs}})}$$

Hidden Markov Model (HMM)

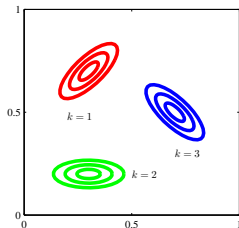
Parameterization

distribution of the initial state	$p(\mathbf{z}_1; \pi) = \prod_{\ell=1}^K \pi_k^{z_{1\ell}}$
transition matrix	$p(\mathbf{z}_i \mathbf{z}_{i-1}; A) = \prod_{k=1}^K \prod_{\ell=1}^K A_{k\ell}^{z_{i-1}, j} z_{i\ell}$
emission probabilities	$p(\mathbf{x}_i \mathbf{z}_i; \phi)$ e.g. Gaussian Mixture

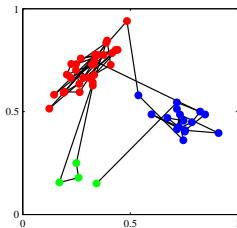
Interpretation



Transitions of \mathbf{z}_i



$p(\mathbf{x}_i | \mathbf{z}_i)$



Path of \mathbf{x}_i

Maximum likelihood for HMMs

Application of the EM algorithm

$$\gamma(\mathbf{z}_i) = p(\mathbf{z}_i | \mathbf{X}, \boldsymbol{\theta}^t) \quad \xi(\mathbf{z}_{i-1}, \mathbf{z}_i) = p(\mathbf{z}_{i-1}, \mathbf{z}_i | \mathbf{X}, \boldsymbol{\theta}^t)$$

Expected value of the log-likelihood:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = \sum_{\ell=1}^K \gamma(z_{1\ell}) \log \pi_k + \sum_{i=2}^N \sum_{k=1}^K \sum_{\ell=1}^K \xi(z_{i-1,k}, z_{i\ell}) \log A_{k\ell} + \sum_{i=1}^N \sum_{\ell=1}^K \gamma(z_{i\ell}) \log p(\mathbf{x}_i | \phi_\ell)$$

Maximizing with respect to the parameters $\{\pi, A\}$, one gets

$$\pi_k^{t+1} = \frac{\gamma(z_{1k})}{\sum_{k'=1}^K \gamma(z_{1k'})}$$

$$A_{k\ell}^{t+1} = \frac{\sum_{i=2}^N \xi(z_{i-1,k}, z_{i\ell})}{\sum_{\ell'=1}^K \sum_{i=2}^N \xi(z_{i-1,k}, z_{i\ell'})}$$

If the emissions are Gaussian then we also have:

$$\boldsymbol{\mu}_k^{t+1} = \frac{\sum_{i=1}^N \gamma(z_{ik}) \mathbf{x}_i}{\sum_{i=1}^N \gamma(z_{ik})} \quad \boldsymbol{\Sigma}_k^{t+1} = \frac{\sum_{i=1}^N \gamma(z_{ik}) (\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^\top}{\sum_{i=1}^N \gamma(z_{ik})}$$