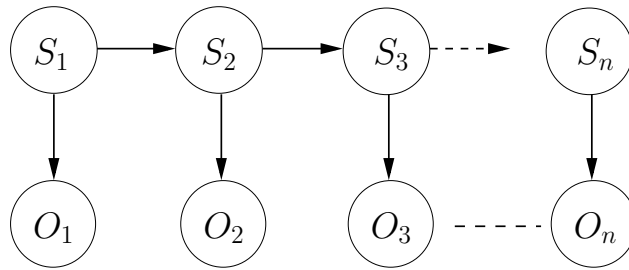


Probabilistic Models: Spring 2014

Gene Finding Example

We are given the following hidden Markov model describing the (simplified) behavior of DNA.



S_0	Θ_{S_0}	S_{t-1}	S_t	$\Theta_{S_t S_{t-1}}$	S_i	O_i	$\Theta_{O_i S_i}$
Genic	.5	Genic	Genic	.7	Genic	AT	.9
Intergenic	.5	Genic	Intergenic	.3	Genic	CG	.1
		Intergenic	Genic	.3	Intergenic	AT	.2
		Intergenic	Intergenic	.7	Intergenic	CG	.8

1. Use the forward algorithm to calculate the predictive posterior distribution over $S_1 \dots S_4$ given the following observations: AT, AT, CG, AT .
2. Use the backward algorithm to calculate the smoothed posterior distribution over S_1 given the observations: AT, AT .
3. Use the Viterbi algorithm to find the most likely instantiation of $S_1 \dots S_4$ given the observations: AT, AT, CG, CG .

Some useful equations

The forward algorithm

$$P(\text{next state}|\text{observations so far}, \text{next observation}) \propto P(\text{next observation}|\text{next state}) \sum_{\text{current state}} P(\text{next state}|\text{current state})P(\text{current state}|\text{observations so far})$$

$$P(S_{t+1}|O_1, O_2, \dots, O_{t+1}) \propto P(O_{t+1}|S_{t+1}) \sum_{S_t=s_t} P(S_{t+1}|S_t)P(S_t = s_t|O_1, \dots, O_t)$$

The backward algorithm

$$P(S_k|O_1, \dots, O_t) \propto \text{forward}(t)P(O_{k+1}, \dots, O_t | S_k)$$

$$P(\text{remaining observations}|\text{current state}) = \sum_{\text{next state}} P(\text{next state}|\text{current state})P(\text{next observation}|\text{next state})P(\text{further observations}|\text{next state})$$

$$P(O_{k+1}, \dots, O_t | S_k) = \sum_{S_{k+1}=s_{k+1}} P(S_{k+1} = s_{k+1}|S_k)P(O_{k+1}|S_{k+1})P(O_{k+2}, \dots, O_t | S_{k+1} = s_{k+1})$$

The Viterbi algorithm

$$\max_{\text{path so far}} P(\text{path so far}, \text{next state in path}|\text{observations so far}, \text{next observation})$$

$$\propto P(\text{next observation} | \text{next state}) \left\{ \max_{\text{current state}} P(\text{next state}|\text{current state}) \left\{ \max_{\text{previous states}} P(\text{previous states}, \text{current state}|\text{observations so far}) \right\} \right\}$$

$$\max_{s_1 \dots s_t} P(s_1 \dots s_t, S_{t+1}|O_1 \dots O_{t+1}) \propto P(O_{t+1}|S_{t+1}) \left\{ \max_{s_t} P(S_{t+1}|s_t) \left\{ \max_{s_1 \dots s_{t-1}} P(s_1 \dots s_{t-1}, s_t | O_1 \dots O_t) \right\} \right\}$$