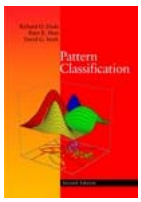
Chapter 3 (part 3): Maximum-Likelihood and Bayesian Parameter Estimation

- Hidden Markov Model: Extension of Markov Chains



Hidden Markov Model (HMM)

- Interaction of the visible states with the hidden states
  $\sum b_{jk} = 1$ for all j where $b_{jk} = P(V_k(t) \mid \omega_j(t))$.
- 3 problems are associated with this model
  - The evaluation problem
  - The decoding problem
  - The learning problem

■ The evaluation problem

It is the probability that the model produces a sequence V^T of visible states. It is:

$$P(V^T \mid \Theta) = \sum_{r=1}^{r_{max}} P(V^T \mid \omega_r^T) P(\omega_r^T) \; ; \quad r_{max} = c^T$$

parameters

where each r indexes a particular sequence of T hidden states $\omega_r^T = \{\omega(1), \omega(2), \ldots, \omega(T)\}$

(1) $$P(V^T \mid \omega_r^T) = \prod_{t=1}^{t=T} P(v(t) \mid \omega(t))$$ conditional independence

(2) $$P(\omega_r^T) = \prod_{t=1}^{t=T} P(\omega(t) \mid \omega(t-1))$$ Markov chain of order 1

Using equations (1) and (2), we can write:

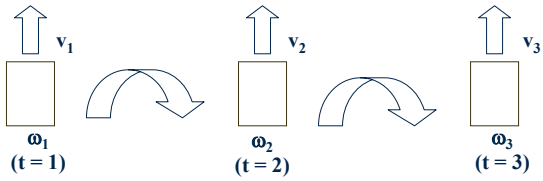
$$P(V^T \mid \Theta) = \sum_{r=1}^{r_{max}} \prod_{t=1}^{t=T} P(v(t) \mid \omega(t)) \; P(\omega(t) \mid \omega(t-1))$$

Interpretation: The probability that we observe the particular sequence of T visible states V^T is equal to the sum over all r_{max} possible sequences of hidden states of the conditional probability that the system has made a particular transition multiplied by the probability that it then emitted the visible symbol in our target sequence.

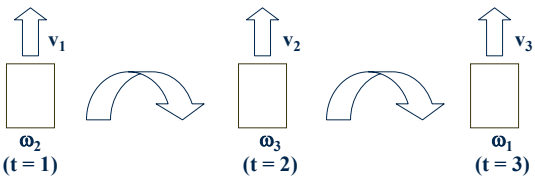
Example: Let ω_1, ω_2, ω_3 be the hidden states; v_1, v_2, v_3 be the visible states and $V^3 = \{v_1, v_2, v_3\}$ is the sequence of visible states

$P(\{v_1, v_2, v_3\}|\Theta) = P(\omega_1).P(v_1 \mid \omega_1).P(\omega_2 \mid \omega_1).P(v_2 \mid \omega_2).P(\omega_3 \mid \omega_2).P(v_3 \mid \omega_3)$
$+\ldots+$ (possible terms in the sum = all possible ($3^3 = 27$) cases !)

First possibility:



Second Possibility:



$P(\{v_1, v_2, v_3\}|\Theta) = P(\omega_2).P(v_1 \mid \omega_2).P(\omega_3 \mid \omega_2).P(v_2 \mid \omega_3).P(\omega_1 \mid \omega_3).P(v_3 \mid \omega_1) + \ldots +$

Therefore: $P(\{v_1, v_2, v_3\} \mid \Theta) = \sum_{\text{possible sequence of hidden states}} \prod_{t=1}^{t=3} P(v(t) \mid \omega(t)).P(\omega(t) \mid \omega(t-1))$

- The evaluation problem is solved using the forward algorithm

- **The decoding problem (optimal state sequence)**

  Given a sequence of visible states V^T, the decoding problem is to find the most probable sequence of hidden states.

  This problem can be expressed mathematically as:

  find the single "best" state sequence (hidden states)

$\hat{\omega}(1), \hat{\omega}(2), ..., \hat{\omega}(T)$ *such that* :

$$\hat{\omega}(1), \hat{\omega}(2), ..., \hat{\omega}(T) = \arg\max_{\omega(1),\omega(2),...,\omega(T)} P[\omega(1), \omega(2), ..., \omega(T), v(1), v(2), ..., V(T) \mid \Theta]$$

Note that the summation disappeared, since we want to find only one unique best case !

Where: $\Theta = [\pi, A, B]$

$\pi = P(\omega(1) = \omega)$ (initial state probability)

$A = a_{ij} = P(\omega(t+1) = j \mid \omega(t) = i)$

$B = b_{jk} = P(v(t) = k \mid \omega(t) = j)$

In the preceding example, this computation corresponds to the selection of the best path amongst:

$\{\omega_1(t = 1), \omega_2(t = 2), \omega_3(t = 3)\}$, $\{\omega_2(t = 1), \omega_3(t = 2), \omega_1(t = 3)\}$
$\{\omega_3(t = 1), \omega_1(t = 2), \omega_2(t = 3)\}$, $\{\omega_3(t = 1), \omega_2(t = 2), \omega_1(t = 3)\}$
$\{\omega_2(t = 1), \omega_1(t = 2), \omega_3(t = 3)\}$

- The decoding problem is solved using the Viterbi Algorithm

- **The learning problem (parameter estimation)**

This third problem consists of determining a method to adjust the model parameters $\Theta = [\pi, A, B]$ to satisfy a certain optimization criterion. We need to find the best model

$$\hat{\Theta} = [\hat{\pi}, \hat{A}, \hat{B}]$$

Such that to maximize the probability of the observation sequence:

$$\underset{\Theta}{Max}\, P(V^T \mid \Theta)$$

We use an iterative procedure such as Baum-Welch (Forward-Backward) or Gradient to find this local optimum

Parameter Updates:

Forward-Backward Algorithm

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T} \gamma_{ij}(t)}{\sum_{t=1}^{t=T} \sum_{k} \gamma_{ik}(t)}$$

$$\gamma_{ij}(t) = \frac{\alpha_i(t-1) a_{ij} b_{jk} \beta_j(t)}{P(V^T|\Theta)}$$

$$\hat{b}_{jk} = \frac{\sum_{\substack{t=1 \\ v(t)=v_k}}^{T} \sum_{l} \gamma_{jl}(t)}{\sum_{t=1}^{T} \sum_{l} \gamma_{jl}(t)}$$

- $\alpha_i(t)$= P(model generates visible sequence up to step t given hidden state $\omega_i(t)$)
- $\beta_i(t)$= P(model will generate the sequence from t+1 to T given $\omega_i(t)$)

Parameters Learning Algorithm

```
Begin initialize

 a_ij, b_jk, training sequence V^T, conv. criterion (cc),
  z=0
 Do z=z+1
    compute â(z) from a(z-1) and b(z-1)
     compute b̂(z) from a(z-1) and b(z-1)
     a_ij(z)= â_ij(z − 1)
     b_jk(z)= b̂_jk(z − 1)
  Until max{a_ij(z)-a_ij(z-1), b_jk(z)-b_jk(z-1)}< cc
  Return a_ij=a_ij(z); b_jk=b_jk(z)
End
```