Hidden Markov Model

- HMM Slides (./files/hmm14a.pdf) from Prof. Andrew W. Moore at CMU
- HMM Slides (./files/Lecture17.pdf) from Prof. AartiSingh at CMU
- HMM tutorial paper (./files/HMM Tutorial.pdf) from Lawrence R. Rabiner

Prof. Seungchul Lee
iSystems Design Lab
UNIST
http://isystems.unist.ac.kr/

Table of Contents

1. Markov Process

1.1. Sequential Data

- Most classifiers ignored the sequential aspects of data
- Consider a system which can occupy one of N discrete states or categories

$$q_t \in \{S_1, S_2, \cdots S_N\}$$

- We are interested in stochastic systems, in which state evolution is random
- Any joint distribution can be factored into a series of conditional distributions

$$p(q_0, q_1, \cdots, q_T) = p(q_0)\, p(q_1 \mid q_0)\, p(q_2 \mid q_1, q_0) \cdots$$

- But almost impossible to compute !!!

1.2. Hidden Markov Process

Markov chain

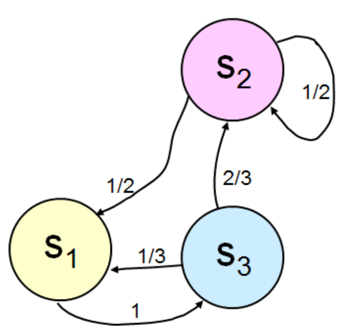
- For a Markov process, the next state depends only on the current state:

$$p(q_{t+1} \mid q_t, \cdots, q_0) = p(q_{t+1} \mid q_t)$$

- More clearly

$$\begin{aligned} p(q_{t+1} = s_j \mid q_t = s_i) &= p(q_{t+1} = s_j \mid q_t = s_i, \text{ any earlier history}) \\ p(q_0, q_1, \cdots, q_T) &= p(q_0)\, p(q_1 \mid q_0)\, p(q_2 \mid q_1, q_0)\, p(q_3 \mid q_2, q_1, q_0) \cdots \\ &= p(q_0)\, p(q_1 \mid q_0)\, p(q_2 \mid q_1)\, p(q_3 \mid q_2) \cdots \end{aligned}$$

- Now it is a kind of possible and tractable in computation



State Transition Matrix

- A stationary Markov chain with N states is described by an $N \times N$ transition matrix

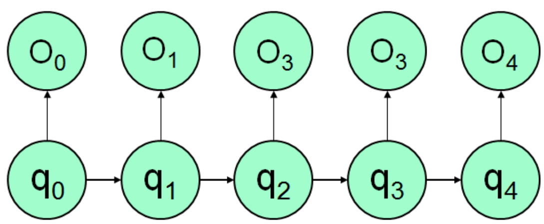
$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \\ 1/3 & 2/3 & 0 \end{bmatrix}$$

Hidden State

- Assumption
  - We can observe something that is affected by the true state
  - Natual way of thinking
- Limited sensors (incomplete state information)
  - But still partially related
- Noisy senors
  - Unreliable
- Observation emitted from q_t
  - O_t is noisily determined, depending on the current state q_t
  - Assume that O_t is conditionally independent of $\{q_{t-1}, q_{t-2}, \cdots, q_0, O_{t-1}, O_{t-2}, \cdots, O_0\}$ given q_t

Markov Property

1. a finite set of N states, $S = \{S_1, \cdots S_N\}$
2. a state transition probability, $P = \{a_{ij}\}_{N \times N},\ i \leq i, j \leq N$
3. an initial state probability distribution, $\pi = \{\pi_i\}$
4. an observation symbol probability distribution, $b_j(O_n)$



very simplified HMM

1.3. Three Questions in HMM

- Question 1: State estimation (most interesting to us)

  What is $p(q_t = s_i \mid O_1, O_2, \cdots, O_T)$

- Question 2: Most probable path

  Given $O_1 O_2 \cdots O_T$, what is the most probable path that I took? And what is that probability?

- Question 3: Learning HMMs

  Given $O_1 O_2 \cdots O_T$, what is the maximum likelihood HMM that could have produced this sequence of observations?

2. State Estimation

Given the observation sequence $O = O_1 O_2 \cdots O_T$, the probability of $q_T = S_i$

$$p(q_T = S_i \mid O_1 O_2 \cdots O_T, \lambda)$$

Then estiamted state $\hat{q}_T$

$$\hat{q}_T = \arg\max_i \{p(q_T = S_i \mid O_1 O_2 \cdots O_T, \lambda)\}$$

Bayes' rule with a conditional probabiliy

$$\begin{aligned} p(A, B \mid C) &= p(A \mid B, C)\, p(B \mid C) \\ \implies p(A \mid B, C) &= \frac{p(A, B \mid C)}{p(B \mid C)} \end{aligned}$$

- $A : q_T = S_i$
- $B : O_1 O_2 \cdots O_T$
- $C : \lambda$

$$p(q_T = S_i \mid O_1 O_2 \cdots O_T, \lambda) = \frac{p(q_T = S_i, O_1 O_2 \cdots O_T \mid \lambda)}{p(O_1 O_2 \cdots O_T \mid \lambda)}$$

Start with a wrong approach

HMM: λ

For one fixted state sequence $q = q_1 q_2 \cdots q_T$

$$\begin{aligned} p(O \mid q, \lambda) &= \prod_{t=1}^{T} p(O_t \mid q_t, \lambda) \\ &= b_{q_1}(O_1) \cdots b_{q_T}(O_T) \end{aligned}$$

Probability of such a sate sequence q

$$p(q \mid \lambda) = \pi_{q_1} a_{q_1 a_2} a_{q_2 a_3} \cdots a_{q_{T-1} a_T}$$

Then

$$\begin{aligned} p(O \mid \lambda) &= \sum_{\text{all } q} p(O, q \mid \lambda) \\ &= \sum_{\text{all } q} p(O \mid q, \lambda) p(q \mid \lambda) \end{aligned}$$

$\rightarrow$ require too much computation

Smarter way

$$p(q_T = S_i \mid O_1 O_2 \cdots O_T, \lambda) = \frac{p(q_T = S_i, O_1 O_2 \cdots O_T \mid \lambda)}{p(O_1 O_2 \cdots O_T \mid \lambda)}$$

Let

$$\alpha_t(i) \equiv p(O_1 O_2 \cdots O_t, q_t = S_i \mid \lambda)$$

Then

$$\begin{aligned} \alpha_1(i) &= p(O_1, q_1 = S_i \mid \lambda) = \pi_i b_i(O_1) \\ &\vdots \\ \alpha_t(i) &= p(O_1 O_2 \cdots O_t, q_t = S_i \mid \lambda) \\ \alpha_{t+1}(j) &= p(O_1 O_2 \cdots O_{t+1}, q_{t+1} = S_j \mid \lambda) \\ &= \left(\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right) b_j(O_{t+1}) \\ &\vdots \\ \alpha_T(j) &= p(O_1 O_2 \cdots O_T, q_T = S_j \mid \lambda) \quad : \text{recursive} \end{aligned}$$

Back to the problem

$$p(O_1 O_2 \cdots O_T \mid \lambda) = p(O \mid \lambda) = \sum_{j=1}^{N} \alpha_T(j)$$
$$p(q_T = S_i \mid O_1 O_2 \cdots O_T, \lambda) = \frac{p(q_T = S_i, O_1 O_2 \cdots O_T \mid \lambda)}{p(O_1 O_2 \cdots O_T \mid \lambda)} = \frac{\alpha_T(i)}{\sum_{j=1}^{N} \alpha_T(j)}$$

Then estimated state $\hat{q}_T$

$$\hat{q}_T = \arg\max_i \{p(q_T = S_i \mid O_1 O_2 \cdots O_T, \lambda)\}$$

Some thouhgts on HMM

- Sequence information
  - comes from state transition matrix
  - difficult to obtain it in practice
  - sequence might be useful in some applications
- Not easy to obtain observation symbol probability $b_j(O_n)$

Online lectures

- Lucy Yin at Caltech (https://www.youtube.com/watch?v=NebQx50u9gw (https://www.youtube.com/watch?v=NebQx50u9gw))
- Bert Huang at Virginia Tech (https://www.youtube.com/watch?v=9yl4XGp5OEg (https://www.youtube.com/watch?v=9yl4XGp5OEg))
- Nando de Freitas at UBC (https://www.youtube.com/watch?v=jY2E6ExLxaw (https://www.youtube.com/watch?v=jY2E6ExLxaw))
- By mathematicalmonk (https://www.youtube.com/watch?v=TPRoLreU9lA (https://www.youtube.com/watch?v=TPRoLreU9lA))