CS725: Data Analysis and Machine Learning

Hidden Markov Model Learning

Jakramate Bootkrajang

Department of Computer Science
Chiang Mai University

HMM learning

- Learning = estimating the parameters
- Parameters of HMM
  - Transition probabilities matrix: A
  - Emission probabilities matrix: B
- Learning is done using Baum-Welch algorithm

The Baum-Welch algorithm

- Randomly initialising the values for A and B

1. Calculate
   - #times s follows s'
   - #times o is observed given s
2. Update
   - $\hat{A}(s', s) = \frac{\#\text{times } s \text{ follows } s'}{\#\text{ times anything follows } s'}$
   - $\hat{B}(s, o) = \frac{\#\text{times } o \text{ is observed given } s'}{\#\text{ times anything is observed given } s}$

How to count those quantities?

- Instead of actual counting we find psuedo count, i.e., normalised count
- #times s follows s' $\rightarrow P(s_{t-1} = s', s_t = s | O, A, B)$
- #times o is observed given s $\rightarrow \mathbb{1}[o_t = o] \cdot P(s_t = s | O, A, B)$
- Define
  - $\gamma[t, s] = P(s_t = s | O, A, B)$
  - $\xi[t, s', s] = P(s_{t-1} = s', s_t = s | O, A, B)$

Finding $\gamma[t, s]$

- By Bayes' rule

$$\gamma[t,s] = P(s_t = s | O, A, B) = \frac{P(s_t = s, o_{1:T} | A, B)}{P(O | A, B)} \tag{1}$$

$$= \frac{P(o_{1:t}, s_t = s | A, B) P(o_{t+1:T} | s_t = s, A, B)}{\sum_{s'} P(o_{1:t}, s_t = s' | A, B) P(o_{t+1:T} | s_t = s', A, B)} \tag{2}$$

Define

- $\alpha[t,s] = P(o_{1:t}, s_t = s | A, B) = P(o_1, o_2, \ldots, o_t, s_t = s | A, B)$
- $\beta[t,s] = P(o_{t+1:T} | s_t = s, A, B) = P(o_{t+1}, o_{t+2}, \ldots, o_T | s_t = s, A, B)$

Finding $\xi[t, s', s]$

$$\xi[t, s', s] = P(s_{t-1} = s', s_t = s | O, A, B)$$

$$= \frac{P(s_{t-1} = s', s_t = s, o_{1:T} | A, B)}{P(O|A, B)} \quad (3)$$

$$= \frac{P(o_{1:t-1}, s_{t-1} = s' | A, B) \cdot P(s_t = s | s_{t-1} = s' | A, B) \cdot P(o_{t+1:T} | s_t = s, A, B)}{\sum_{s'} P(o_{1:t}, s_t = s' | A, B) P(o_{t+1:T} | s_t = s', A, B)} \quad (4)$$

$\gamma[t, s]$ and $\xi[t, s', s]$ in terms of α and β

$$\gamma[t, s] = \frac{\alpha[t, s] \cdot \beta[t, s]}{\sum_{s'} \alpha[t, s'] \cdot \beta[t, s']} \tag{5}$$

$$\xi[t, s', s] = \frac{\alpha[t-1, s] \cdot A(s', s) \cdot \beta[t, s]}{\sum_{s'} \alpha[t, s'] \cdot \beta[t, s']} \tag{6}$$

Compute $\alpha[t, s]$ and $\beta[t, s]$ with DP

$$\alpha[t, s] = \sum_{s'} \alpha[t-1, s'] \cdot A(s', s) \cdot B(s, o_t) \tag{7}$$

$$\beta[t, s] = \sum_{s'} \beta[t+1, s'] \cdot A(s, s') \cdot B(s', o_{t+1}) \tag{8}$$

The Baum-Welch algorithm summary

1 Randomly initialising the values for A and B and repeat step 2 until convergence

2.1 Compute

$$\alpha[t, s] = \sum_{s'} \alpha[t-1, s'] \cdot A(s', s) \cdot B(s, o_t) \tag{9}$$

$$\beta[t, s] = \sum_{s'} \beta[t+1, s'] \cdot A(s, s') \cdot B(s', o_{t+1}) \tag{10}$$

$$\gamma[t, s] = \frac{\alpha[t, s] \cdot \beta[t, s]}{\sum_{s'} \alpha[t, s'] \cdot \beta[t, s']} \tag{11}$$

$$\xi[t, s', s] = \frac{\alpha[t-1, s] \cdot A(s', s) \cdot \beta[t, s]}{\sum_{s'} \alpha[t, s'] \cdot \beta[t, s']} \tag{12}$$

2.2 Update

- $\hat{A}(s', s) = \frac{\sum_{t=1}^{T-1} \xi[t,s',s]}{\sum_{s''} \sum_{t=1}^{T-1} \xi[t,s',s'']}$
- $\hat{B}(s, o) = \frac{\sum_{t=1}^{T} \mathbb{1}[o_t=o]\gamma[t,s]}{\sum_{t=1}^{T} \gamma[t,s]}$

References

- 11-711: Notes on Hidden Markov Model – http://www.cs.cmu.edu/~tbergkir/11711fa17/recitation4_notes.pdf