Recitation 8
Hidden Markov Model

10-601: Introduction to Machine Learning

11/13/2020

1 HMMs

You are given the following training data:

```
win_C league_C Liverpool_D

win_C Liverpool_D league_C

Liverpool_D win_C
```

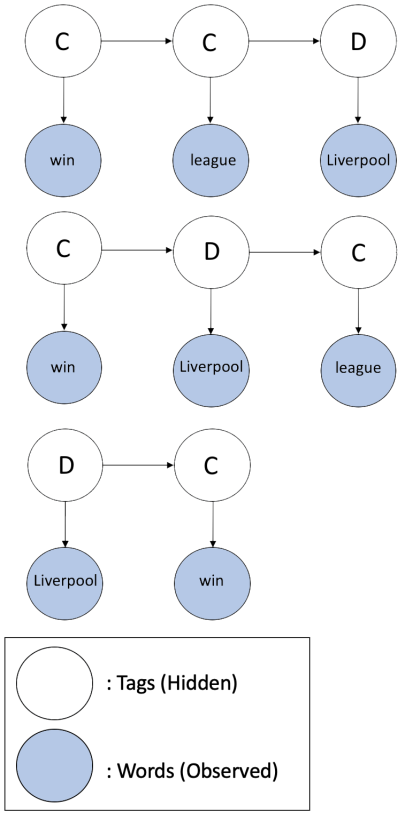

Figure 1: Visualization of Sequences

You are also given the following observed (validation) data:

```
Liverpool win league
```

In this question, let each observed state $x_t \in \{1, 2, 3\}$, where 1 corresponds to `win`, 2 corresponds to `league`, and 3 corresponds to `Liverpool`. Let each hidden state $Y_t \in \{C, D\}$, where $s_1 = C$ and $s_2 = D$.

1. First, we need to train our HMM by generating the initial probabilities: $\boldsymbol{\pi}$, the transition probability matrix: $\mathbf{A}$, the emission probability matrix: $\mathbf{B}$.

   (a) Find $\boldsymbol{\pi}$. Recall that $\pi_j = P(Y_1 = s_j)$.

(b) Find Transition Matrix: **A**. Recall that $A_{jk} = P(Y_t = s_k \mid Y_{t-1} = s_j)$

(c) Find Emission Matrix: **B**. Recall that $B_{jk} = P(X_t = k \mid Y_t = s_j)$.

2. What is the likelihood of observing this output?
   Recall that:

$$\alpha_t(k) = P(x_{1:t}, Y_t = s_k)$$
$$\beta_t(k) = P(x_{t+1:T}|Y_T = s_k)$$

   We also have the recursive procedure:

   (a) $\alpha_1(j) = \pi_j B_{jx_1}$.

   (b) For $t > 1$, $\alpha_t(j) = B_{jx_t} \sum_{k=1}^{J} \alpha_{t-1}(k) A_{kj}$

You are now told that the observed data has the following tags:

```
Liverpool_D win_C league_D
```

3. Given the observed sequence of words (denote $\vec{x} = [\text{Liverpool, win, league}]^T$), what is the probability of these assigned tags $P(Y_1 = D|\vec{x}), P(Y_2 = C|\vec{x}), P(Y_3 = D|\vec{x})$?

   Recall that:

   $$P(Y_t = s_k|\vec{x}) = \frac{\alpha_t(s_k)\beta_t(s_k)}{P(\vec{x})}$$

   So, we need to find β_T

   We also have a similar recursive procedure

   (a) $\beta_T(j) = 1$ (All states could be ending states)

   (b) For $1 \leq t \leq T-1$, $\beta_t(j) = \sum_{k=1}^{J} B_{kx_{t+1}}\beta_{t+1}(k)A_{jk}$ (Generate x_{t+1} from any state)

4. The sequence of words you observe is again the same:

   `Liverpool win league`

   However, you are only given the tag of the last word:

   `league_C`

   Using the Viterbi Algorithm, what is the most likely sequence of hidden states?

   Recall that:

$$\omega_t(s_k) = \max_{y_{1:t-1}} P(x_{1:t}, y_{1:t-1}, y_t = s_k)$$

$$b_t(s_k) = \arg\max_{y_{1:t-1}} P(x_{1:t}, y_{1:t-1}, y_t = s_k)$$

   (a) What is the most likely sequence of tags given the observed data? (Select `C` if there is a tie)