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
```

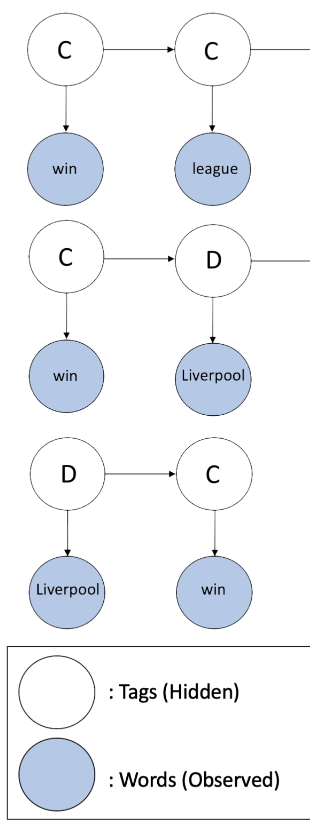

Figure 1: Visualization of Sequences

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 $\pi_j = P(Y_1 = s_j)$.

   - Find count matrix

$$\begin{matrix} & \textit{Count} \\ C & \\ D & \end{matrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \xrightarrow{\text{Pseudocount}} \begin{matrix} C \\ D \end{matrix} \overset{\textit{Count}}{\begin{bmatrix} 3 \\ 2 \end{bmatrix}}$$

   - Get probability matrix π:

$$\pi = \begin{matrix} C \\ D \end{matrix} \begin{bmatrix} 3/5 \\ 2/5 \end{bmatrix}$$

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

- Find count matrix

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \end{array} \xrightarrow{\text{Pseudocount}} \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix} \end{array}$$

- Get Transition Probability matrix **A**:

$$A = \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{bmatrix} 2/5 & 3/5 \\ 3/4 & 1/4 \end{bmatrix} \end{array}$$

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

- Find count matrix

$$\begin{matrix} & \texttt{win} & \texttt{league} & \texttt{Liverpool} \\ C \\ D \end{matrix} \begin{bmatrix} 3 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \xrightarrow{\text{Pseudocount}} \begin{matrix} & \texttt{win} & \texttt{league} & \texttt{Liverpool} \\ C \\ D \end{matrix} \begin{bmatrix} 4 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix}$$

- Get Emission Probability matrix **B**:

$$B = \begin{matrix} & \texttt{win} & \texttt{league} & \texttt{Liverpool} \\ C \\ D \end{matrix} \begin{bmatrix} 1/2 & 3/8 & 1/8 \\ 1/6 & 1/6 & 2/3 \end{bmatrix}$$

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

We want to find:

$$\begin{aligned} &P(X_1 = \texttt{Liverpool}, X_2 = \texttt{win}, X_3 = \texttt{league}) \\ &= \sum_{y_t \in C,D} P(x_1 = \texttt{Liverpool}, x_2 = \texttt{win}, x_3 = \texttt{league}, Y_t = y_t) \\ &= \sum_{y_t \in C,D} \alpha_T(y_t) \end{aligned}$$

$$\begin{aligned} \alpha_1 &= P(x_1|y_1) \cdot P(y_1) \\ &= B_{,3} \circ \pi \end{aligned}$$

$$= \begin{bmatrix} 1/8 \\ 2/3 \end{bmatrix} \circ \begin{bmatrix} 3/5 \\ 2/5 \end{bmatrix}$$
$$= \begin{bmatrix} 0.075 \\ 0.26667 \end{bmatrix}$$

$$\begin{aligned} \alpha_2 &= P(x_1, x_2, y_2) \\ &= P(x_2|y_2) \cdot \big(P(y_2|y_1) \cdot \alpha_1\big) \\ &= B_{,1} \circ (A^T \alpha_1) \end{aligned}$$

$$= \begin{bmatrix} 1/2 \\ 1/6 \end{bmatrix} \circ \begin{bmatrix} 0.2300025 \\ 0.1116675 \end{bmatrix}$$
$$= \begin{bmatrix} 0.11500125 \\ 0.01861125 \end{bmatrix}$$

$$\alpha_3 = B_{,2} \circ (A^T \alpha_2) = \begin{bmatrix} 0.02248460156 \\ 0.01227559375 \end{bmatrix}$$

Since

$$\begin{aligned} &P(x_1 = \texttt{Liverpool}, x_2 = \texttt{win}, x_3 = \texttt{league}) \\ &= \sum_{y_t \in C,D} \alpha_T(y_t), \text{where we set } T = 3 \end{aligned}$$

$\therefore P(x_1 = \texttt{Liverpool}, x_2 = \texttt{win}, x_3 = \texttt{league})$
$= 0.02248460156 + 0.01227559375$
$= 0.03476019531$

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

Remember that: $\beta_t(s_k) = P(x_{t+1:T}|Y_T = s_k)$ and $\beta_T(s_k) = 1$

Using matrix notation:

$$\beta_2 = A(B_{,x_3} \circ \beta_3) = A(B_{,2} \circ \beta_3)$$

Recall that:

$$B_{,2} = \begin{bmatrix} 3/8 \\ 1/6 \end{bmatrix} \text{ and } \beta_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ , since } T = 3$$

$$\therefore \beta_2 = \begin{bmatrix} 0.25 \\ 0.3229 \end{bmatrix}$$

Now, we go on to solve β_1

$$\beta_1 = A(B_{,x_2} \circ \beta_2) = A(B_{,1} \circ \beta_2)$$

Again, recall that:

$$B_{,1} = \begin{bmatrix} 1/2 \\ 1/6 \end{bmatrix} \text{ and } \beta_2 = \begin{bmatrix} 0.25 \\ 0.3229 \end{bmatrix}$$

$$\therefore \beta_1 = \begin{bmatrix} 0.08229 \\ 0.1072 \end{bmatrix}$$

Now, we have our α and β matrix:

$$\alpha = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{bmatrix} 0.0750 & 0.26667 \\ 0.1150 & 0.0186 \\ 0.0225 & 0.0123 \end{bmatrix} \end{array}$$

$$\beta = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{bmatrix} 0.0823 & 0.1072 \\ 0.2500 & 0.3229 \\ 1.0000 & 1.0000 \end{bmatrix} \end{array}$$

$$\begin{aligned} P(Y_1 = D|\vec{x}) &= \frac{\alpha_1(D)\beta_1(D)}{P(\vec{x})} \\ &= \frac{0.26667 \times 0.1072}{0.03476019531} \\ &= 0.8224068865 \end{aligned}$$

$$\begin{aligned} P(Y_2 = C|\vec{x}) &= \frac{\alpha_2(C)\beta_2(C)}{P(\vec{x})} \\ &= \frac{0.1150 \times 0.2500}{0.03476019531} \\ &= 0.8270954678 \end{aligned}$$

$$\begin{aligned} P(Y_3 = C|\vec{x}) &= \frac{\alpha_3(C)\beta_3(C)}{P(\vec{x})} \\ &= \frac{0.0225 \times 1}{0.03476019531} \\ &= 0.6472921052 \end{aligned}$$

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)

      i. Set up the matrices ω and b

$$\omega = \begin{array}{c} \\ \omega_0 \\ \omega_1 \\ \omega_2 \\ \omega_3 \end{array} \begin{array}{c} \begin{array}{ccc} \texttt{C} & \texttt{D} & \texttt{START} \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ - & - & - \\ - & - & - \\ - & - & - \end{bmatrix} \end{array}$$

      and

$$b = \begin{array}{c} \\ b_1 \\ b_2 \\ b_3 \end{array} \begin{array}{c} \begin{array}{cc} \texttt{C} & \texttt{D} \end{array} \\ \begin{bmatrix} - & - \\ - & - \\ - & - \end{bmatrix} \end{array}$$

      Initialize $w_0(\text{START}) = 1$

      ii. Solve for matrix entries using Dynamic Programming:

$$\begin{aligned} \omega_1(C) &= \max_{s_j \in \texttt{C,D,START}} P(x_1 = \texttt{Liverpool} | Y_1 = \texttt{C}) \omega_0(s_j) P(Y_1 = \texttt{C}) \\ &= \frac{1}{8} \cdot 1 \cdot \frac{3}{5} \\ &= \frac{3}{40} \end{aligned}$$

$$b_1(\texttt{C}) = \texttt{START}$$

$$\begin{aligned}\omega_1(D) &= \max_{s_j \in \texttt{C,D,START}} P(x_1 = \texttt{Liverpool}|Y_1 = \texttt{D}) w_0(s_j) P(Y_1 = \texttt{D}) \\ &= \frac{2}{3} \cdot 1 \cdot \frac{2}{5} \\ &= \frac{4}{15}\end{aligned}$$

$$b_1(\texttt{D}) = \texttt{START}$$

$$\begin{aligned}\omega_2(C) &= \max_{s_j \in \texttt{C,D}} P(x_2 = \texttt{win}|Y_2 = \texttt{C}) \omega_1(s_j) P(Y_2 = \texttt{C}|Y_1 = s_j) \\ &= \max\left(\frac{1}{2} \cdot \frac{3}{40} \cdot \frac{2}{5}, \frac{1}{2} \cdot \frac{4}{15} \cdot \frac{3}{4}\right) \\ &= \frac{1}{10}\end{aligned}$$

$$b_2(\texttt{C}) = \texttt{D}$$

$$\begin{aligned}\omega_2(D) &= \max_{s_j \in \texttt{C,D}} P(x_2 = \texttt{win}|Y_2 = \texttt{D}) \omega_1(s_j) P(Y_2 = \texttt{D}|Y_1 = s_j) \\ &= \max\left(\frac{1}{6} \cdot \frac{3}{40} \cdot \frac{3}{5}, \frac{1}{6} \cdot \frac{4}{15} \cdot \frac{1}{4}\right) \\ &= \frac{1}{90}\end{aligned}$$

$$b_2(\texttt{D}) = \texttt{D}$$

$$\begin{aligned}\omega_3(C) &= \max_{s_j \in \texttt{C,D}} P(x_3 = \texttt{league}|Y_3 = \texttt{C}) \omega_2(s_j) P(Y_3 = \texttt{C}|Y_2 = s_j) \\ &= \max\left(\frac{3}{8} \cdot \frac{1}{10} \cdot \frac{2}{5}, \frac{3}{8} \cdot \frac{1}{90} \cdot \frac{3}{4}\right) \\ &= \frac{3}{200}\end{aligned}$$

$$b_3(\texttt{C}) = \texttt{C}$$

$$\begin{aligned}\omega_3(D) &= \max_{s_j \in \texttt{C},\texttt{D}} P(x_3 = \texttt{league}|Y_3 = \texttt{D})\omega_2(s_j)P(Y_3 = \texttt{D}|Y_2 = s_j) \\ &= \max\left(\frac{1}{6} \cdot \frac{1}{10} \cdot \frac{3}{5}, \frac{1}{6} \cdot \frac{1}{90} \cdot \frac{1}{4}\right) \\ &= \frac{1}{100}\end{aligned}$$

$$b_3(\texttt{D}) = \texttt{C}$$

Now, to figure out the order, we set $\hat{y}_t = b_{t+1}(\hat{y}_{t+1})$

$$\begin{aligned}y_{T+1} &= \texttt{END} \\ y_3 &= \texttt{C} \\ \hat{y}_2 &= b_3(\texttt{C}) \\ &= \texttt{C} \\ \hat{y}_1 &= b_2(\texttt{C}) \\ &= \texttt{D} \\ \hat{y}_0 &= b_1(\texttt{D}) \\ &= \texttt{START}\end{aligned}$$

So, the most likely sequence is `START-D-C-C-END`