HMM

Pumpkin

August 13, 2021

Contents

1 HMM

In Hidden Markov Model, current state only depends on last state.

$$P(x_{t+1}|x_t, x_{t-1}, \cdots, x_1, x_0) = P(x_{t+1}|x_t \cdots x_0) \tag{1}$$

A HMM Model can be written as $\lambda(S, O, A, B, \Pi)$
Where

S is State Space(Implicit)

O is Output Space(Explicit)

A is Trainsition Possibility

B is Emission Possibility

Π is Inital State Possibility

1.1 Trainsition Possibility

Trainsition Possibility $A = S \times S$ represents the Possibility of current state with refenrence to last state.

$$A_{ij} = P(s_{t+1} = i | s_t = j) \tag{2}$$

1.2 Emission Possibility

Emission Possibility $B = S \times O$ represents the Possibility of current output with refenrence to current state.

$$B_i(o) = P(o | s = j) \tag{3}$$

2 Forward

$$\begin{aligned}
\alpha_t(i) &= P(o_t, o_{t-1}, \cdots, o_1, o_0, s_t = i) \\
&= \sum_{j \in S} P(o_t, o_{t-1}, \cdots, o_1, o_0, s_{t-1} = j, s_t = i) \\
&= \sum_{j \in S} P(o_t, s_t = i | s_{t-1} = j) * P(o_{t-1}, \cdots, o_1, o_0, s_{t-1} = j) \\
&= \sum_{j \in S} P(s_t = i | s_{t-1} = j) * P(o_t | s_t = i) * \alpha_{t-1}(j) \\
&= \left(\sum_{j \in S} \alpha_{t-1}(j) * A_{ji} \right) * B_i(o_t)
\end{aligned} \tag{4}$$

$$\alpha_0(i) = P(s_0 = i) = \Pi_i \tag{5}$$

3 Backward

$$
\begin{aligned}
\beta_t(i) &= P(o_{t+1}, o_{t+2}, \cdots, o_{T-2}, o_{T-1}, s_t = i) \\
&= \sum_{j \in S} P(o_{t+1}, o_{t+2}, \cdots, o_{T-2}, o_{T-1}, s_{t+1} = j, s_t = i) \\
&= \sum_{j \in S} P(o_{t+1}, s_t = i | s_{t+1} = j) * P(o_{t+2}, \cdots, o_{T-2}, o_{T-1}, s_{t+1} = j) \\
&= \sum_{j \in S} P(s_t = i | s_{t+1} = j) * P(o_{t+1} | s_{t+1} = j) * \beta_{t+1}(j) \\
&= \sum_{j \in S} A_{ij} * B_j(o_{t+1}) * \beta_{t+1}(j)
\end{aligned} \tag{6}
$$

$$
\beta_{T-1}(i) = P(s_{T-1} = i) = 1 \tag{7}
$$

4 Single State Possibility

$$
\begin{aligned}
&P(s_t = i, O) \\
&= P(o_t, o_{t-1}, \cdots, o_1, o_0, s_t = i) * P(o_{t+1}, o_{t+2}, \cdots, o_{T-2}, o_{T-1}, s_t = i) \\
&= \alpha_t(i) * \beta_t(i)
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
\gamma_t(i) &= P(s_t = i | O) \\
&= \frac{P(s_t = i, O)}{\sum_{j \in S} P(s_t = j, O)} \\
&= \frac{\alpha_t(i) * \beta_t(i)}{\sum_{j \in S} \alpha_t(j) * \beta_t(j)}
\end{aligned} \tag{9}
$$

5 Double State Possibility

$$
\begin{aligned}
&P(s_t = i, s_{t+1} = j, O) \\
&= P(o_t, o_{t-1}, \cdots, o_1, o_0, s_t = i) * P(o_{t+1}, s_{t+1} = j | s_t = i) \\
&* P(o_{t+2}, o_{t+2}, \cdots, o_{T-2}, o_{T-1}, s_{t+1} = j) \\
&= \alpha_t(i) * A_{ij} * B_j(o_{t+1}) * \beta_{t+1}(j)
\end{aligned} \tag{10}
$$

$$
\begin{aligned}
\xi_t(i,j) &= P(s_t = i, s_{t+1} = j|O) \\
&= \frac{P(s_t = i, s_{t+1} = j, O)}{\sum_{m\in S}\sum_{n\in S} P(s_t = m, s_{t+1} = n, O)} \\
&= \frac{\alpha_t(i) * A_{ij} * B_j(o_{t+1}) * \beta_{t+1}(j)}{\sum_{m\in S}\sum_{n\in S} \alpha_t(i) * A_{ij} * B_j(o_{t+1}) * \beta_{t+1}(j)}
\end{aligned}
\tag{11}
$$

$$
\xi_{T-1}(i,j) = P(s_{T-1} = i, s_T = j|O) = \alpha_{T-1}(i) * A_{ij} \tag{12}
$$

6 Issues

6.1 Evaluatation

Evaluate the probability of observation sequence

$$
\begin{aligned}
&P(O|\lambda) \\
&= \sum_{i\in S} P(o_{T-1}, o_{T-2}, \cdots, o_1, o_0, s_{T-1} = i) \\
&= \sum_{i\in S} \alpha_{T-1}(i) \\
&= \sum_{i\in S} P(o_0, s_0 = i, o_1, o_2, \cdots, o_{T-2}, o_{T-1}) \\
&= \sum_{i\in S} P(o_0|s_0 = i) * P(s_0 = i, o_1, o_2, \cdots, o_{T-2}, o_{T-1}) \\
&= \sum_{i\in S} \beta_0(i) * B_i(o_0)
\end{aligned}
\tag{13}
$$

6.2 Prediction/Decoding

Find the state sequence with the greatest possibility

$$
S' = \arg\max_S P(O|S,\lambda) \tag{14}
$$

6.3 Viterbi

$$\begin{aligned}
\delta_t(i) &= \max_{S' \in S^t} P(O|S', \lambda) \\
&= \left(\max_{s \in S} A_{si} * \delta_{t-1}(s) \right) * B_i(o_t) \\
\delta_0(i) &= \Pi_i * B_i(o_0) \\
\phi_t(i) &= \arg\max_{s \in S} A_{si} * \delta_{t-1}(s) \\
S'_{T-1} &= \arg\max_{s \in S} \delta_{T-1}(s) \\
S'_{t-1} &= \phi_t(S_t) \quad t \in [1, T-1]
\end{aligned} \tag{15}$$

6.4 Learning

Learning the parameters of HMM to fit the output sequences

$$\lambda' = \arg\max_{\lambda} \sum_{d \in D} P(O^{(d)}|\lambda) \tag{16}$$

$$A'_{ij} = \frac{\sum_{d \in D} \sum_{t=0}^{T-1} \xi_t^{(d)}(i,j)}{\sum_{d \in D} \sum_{t=0}^{T-1} \gamma_t^{(d)}(i)} \tag{17}$$

$$B'_i(j) = \frac{\sum_{d \in D} \sum_{t=0}^{T-1} I(o_t^{(d)} = j)\gamma_t^{(d)}(i)}{\sum_{d \in D} \sum_{t=0}^{T-1} \gamma_t^{(d)}(i)} \tag{18}$$

$$\Pi'_i = \frac{\sum_{d \in D} \gamma_0^{(d)}(i)}{D} \tag{19}$$

6.5 Maximum

Find the output sequence with the greatest probability
Output elements are independent of each other

$$O_t = \arg\max_{o_t \in O} P(o_t) \tag{20}$$

$$
\begin{aligned}
P(o_t) &= \sum_{i \in S} P(o_t | s_t = i) * P(s_t = i) \\
&= \sum_{i \in S} B_i(o_t) * P(s_t = i)
\end{aligned}
\tag{21}
$$

$$
\begin{aligned}
P(s_t = i) &= \sum_{j \in S} P(s_t = i | s_{t-1} = j) * P(s_{t-1} = j) \\
&= \sum_{j \in S} A_{ji} P(s_{t-1} = j) \\
P(s_0 = i) &= \Pi_i
\end{aligned}
\tag{22}
$$