HIDDEN MARKOV MODELS

David Talbot, Yandex Translate

Autumn 2020

Yandex School of Data Analysis

NON-INDEPENDENT HIDDEN PROCESS

You see a single sequence of coin flips and are told they were generated by two different coins but it's not clear which observations are associated with which coin.

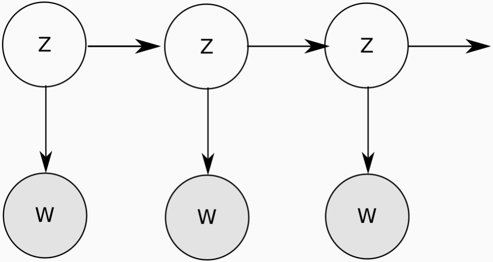
H, H, H, H, H, H, T, H, T, T, T, H, T, T, H, T, T, H, H, H, H, H, H, ...

How could you model this data?

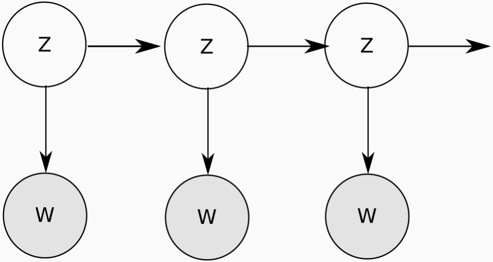
What parameters would you need?

What are the sufficient statistics?

How could you estimate the parameters?



Each state emits one symbol then state transition occurs.



What independence assumptions does it make?

Parameters:

$$\theta = (\pi, A, O)$$

Parameters:

$$\theta = (\pi, A, O)$$

Probability of starting in state i:

$$\pi_i = \Pr(Z_0 = i)$$

Parameters:

$$\theta = (\pi, A, O)$$

Probability of starting in state i:

$$\pi_i = \Pr(Z_0 = i)$$

Probability of moving from state i to j:

$$A_i(j) = \Pr(Z_t = j | Z_{t-1} = i)$$

Parameters:

$$\theta = (\pi, A, O)$$

Probability of starting in state i:

$$\pi_i = \Pr(Z_0 = i)$$

Probability of moving from state i to j:

$$A_i(j) = \Pr(Z_t = j | Z_{t-1} = i)$$

Probability of emitting x given we're in state i:

$$O_i(x) = \Pr(X_t = x | Z_t = i)$$

Parameters:

$$\theta = (\pi, A, O)$$

Probability of starting in state *i*:

$$\pi_i = \Pr(Z_0 = i)$$

Probability of moving from state *i* to *j*:

$$A_i(j) = \Pr(Z_t = j | Z_{t-1} = i)$$

Probability of emitting *x* given we're in state *i*:

$$O_i(x) = \Pr(X_t = x | Z_t = i)$$

What are the sufficient statistics?

In the observed case, we need the following statistics:

$$\#(Z_0 = i)$$

$$\#(Z_{t-1} = i, Z_t = j)$$

$$\#(X_t = x, Z_t = i)$$

In the hidden case, we need expectations for each sample:

$$\#(Z_0 = i) \rightarrow \Pr(Z_0 = i | X_{0:T} = x_{0:T}, \theta)$$

$$\#(Z_{t-1} = i, Z_t = j) \rightarrow \Pr(Z_{t-1} = i, Z_t = j | X_{0:T} = x_{0:T}, \theta)$$

$$\#(X_t = x, Z_t = i) \rightarrow \Pr(Z_t = i | X_{0:T} = x_{0:T}, \theta) \#(X_t = x)$$

We want to compute:

$$\Pr(Z_t = z | X_{0:T} = x_{0:T}, \theta) = \frac{\Pr(Z_t = z, X_{0:T} = x_{0:T})}{\Pr(X_{0:T} = x_{0:T})}$$

We want to compute:

$$\Pr(Z_t = z | X_{0:T} = x_{0:T}, \theta) = \frac{\Pr(Z_t = z, X_{0:T} = x_{0:T})}{\Pr(X_{0:T} = x_{0:T})}$$

But the computation looks exponential in the length T ...

$$\Pr(X_{0:T} = x_{0:T}) = \sum_{z_0, z_1, \ldots, z_T} \Pr(x_{0:T}, z_0, z_1, \ldots, z_T | \theta)$$

Use HMM independence assumptions to factorize

$$\Pr(x_0, \ldots, x_t, z_t, x_{t+1}, \ldots, x_T | \theta) = \Pr(x_0, \ldots, x_t, z_t | \theta) \Pr(x_{t+1}, \ldots, x_T | z_t, \theta)$$

Use HMM independence assumptions to factorize

$\Pr(x_0, \ldots, x_t, z_t, x_{t+1}, \ldots, x_T|\theta) = \Pr(x_0, \ldots, x_t, z_t|\theta)\Pr(x_{t+1}, \ldots, x_T|z_t, \theta)$

If we can compute this, then the denominator is easy

$$\Pr(x_0, \ldots, x_T|\theta) = \sum_{z_t} \Pr(x_0, \ldots, x_t, z_t|\theta)\Pr(x_{t+1}, \ldots, x_T|z_t, \theta).$$

HIDDEN MARKOV MODEL: BAUM-WELCH FORWARD PASS

Compute $\Pr(x_0, \ldots, x_t, z_t|\theta)$ from $\Pr(x_0, \ldots, x_{t-1}, z_{t-1}|\theta)$ as,

HIDDEN MARKOV MODEL: BAUM-WELCH FORWARD PASS

Compute $\Pr(x_0, \ldots, x_t, z_t|\theta)$ from $\Pr(x_0, \ldots, x_{t-1}, z_{t-1}|\theta)$ as,

$$\begin{aligned}
\Pr(x_0, \ldots, x_t, z_t|\theta) &= \sum_{z_{t-1}} \Pr(x_0, \ldots, x_{t-1}, x_t, z_{t-1}, z_t|\theta) \\
&= \sum_{z_{t-1}} \Pr(x_0, \ldots, x_{t-1}, z_{t-1}|\theta)\Pr(z_t|z_{t-1})\Pr(x_t|z_t) \\
&= \sum_{z_{t-1}} \Pr(x_0, \ldots, x_{t-1}, z_{t-1}|\theta)A_{z_{t-1}}(z_t)O_{z_t}(x_t)
\end{aligned}$$

HIDDEN MARKOV MODEL: BAUM-WELCH FORWARDS PASS

Definition:

$$\alpha_t(z) \equiv \Pr(x_0, \ldots, x_t, z_t | \theta)$$

Initialization:

$$\alpha_0(i) = \pi_i O_i(x_0)$$

Recursion:

$$\alpha_{t+1}(i) = \sum_j \alpha_t(j) A_j(i) O_i(x_t)$$

Gives us the probability of observed sequence since,

$$\Pr(x_0, \ldots, x_T | \theta) = \sum_{z_T} \Pr(x_0, \ldots, x_T, z_T | \theta) = \sum_i \alpha_T(i).$$

Definition:

$$\beta_t(z) \equiv \Pr(x_{t+1}, \ldots, x_T | z_t, \theta)$$

Initialization:

$$\beta_T(i) = 1$$

Recursion:

$$\beta_t(i) = \sum_j \beta_{t+1}(j) A_i(j) O_i(x_{t+1})$$

Posterior probabilities over single states

$$\begin{aligned}\Pr(Z_t = i | x_0, \ldots, x_T; \theta) &= \frac{\Pr(Z_t = i, x_0, \ldots, x_T | \theta)}{\Pr(x_0, \ldots, x_T | \theta)} \\ &= \frac{\alpha_t(i)\beta_t(i)}{\sum_j \alpha_t(j)\beta_t(j)}\end{aligned}$$

Posterior probabilities over state transitions

$$\begin{aligned}\Pr(Z_t = i, Z_{t+1} = j | x_0, \ldots, x_T; \theta) &= \frac{\Pr(Z_t = i, Z_{t+1} = j, x_0, \ldots, x_T | \theta)}{\Pr(x_0, \ldots, x_T | \theta)} \\ &= \frac{\alpha_t(i) A_j(i) O_i(x_{t+1}) \beta_{t+1}(i)}{\sum_i \sum_j \alpha_t(i) A_j(i) O_i(x_{t+1}) \beta_{t+1}(i)}\end{aligned}$$

You're given some texts that have been encrypted. Each was encoded using a different character substitution cipher.

You're given some texts that have been encrypted. Each was encoded using a different character substitution cipher.

You need to identify the author of each text.

You’re given some texts that have been encrypted. Each was encoded using a different character substitution cipher.

You need to identify the author of each text.

You are also provided with some samples of English and Russian writers.

You're given some texts that have been encrypted. Each was encoded using a different character substitution cipher.

You need to identify the author of each text.

You are also provided with some samples of English and Russian writers.

How could you use an HMM to solve this problem?