Hidden Markov models

Source: Hastie et at. (2009), Daumé III. Thanks to D. Hsu.

Markov models

Markov model: a stochastic process $\{Y_t\}_{t\in\mathbb{N}}$ where, for each $t \in \mathbb{N}$, the conditional distribution of the next state Y_{t+1} given all previous states $\{Y_\tau : \tau \leq t\}$ only depends on the value of the current state Y_t.

Markov models

Markov model: a stochastic process $\{Y_t\}_{t \in \mathbb{N}}$ where, for each $t \in \mathbb{N}$, the conditional distribution of the next state Y_{t+1} given all previous states $\{Y_\tau : \tau \leq t\}$ only depends on the value of the current state Y_t.

Conditioned on present Y_t, past $\{Y_\tau\}_{\tau < t}$ and future $\{Y_\tau\}_{\tau > t}$ are independent.

$$\cdots \longrightarrow Y_{t-1} \longrightarrow \; Y_t \; \longrightarrow Y_{t+1} \longrightarrow \cdots$$

Markov models

Markov model: a stochastic process $\{Y_t\}_{t \in \mathbb{N}}$ where, for each $t \in \mathbb{N}$, the conditional distribution of the next state Y_{t+1} given all previous states $\{Y_\tau : \tau \leq t\}$ only depends on the value of the current state Y_t.

Conditioned on present Y_t, past $\{Y_\tau\}_{\tau < t}$ and future $\{Y_\tau\}_{\tau > t}$ are independent.

$$\cdots \longrightarrow Y_{t-1} \longrightarrow Y_t \longrightarrow Y_{t+1} \longrightarrow \cdots$$

Specifying a Markov chain (with discrete state space $[K] = \{1, 2, \ldots, K\}$):

- **Initial state distribution**: K-dimensional probability vector $\boldsymbol{\pi}$

$$\pi_i = \Pr(Y_1 = i).$$

- **Transition matrix**: $K \times K$ matrix $\boldsymbol{A}$

$$A_{i,j} = \Pr(Y_{t+1} = j \mid Y_t = i)$$

  (rows of $\boldsymbol{A}$ are probability vectors).

Hidden Markov models

Hidden Markov model (HMM): a Markov chain $\{(X_t, Y_t)\}_{t \in \mathbb{N}}$, where

Hidden Markov models

Hidden Markov model (HMM): a Markov chain $\{(X_t, Y_t)\}_{t \in \mathbb{N}}$, where

- $\{Y_t\}_{t \in \mathbb{N}}$ is also a Markov chain (with state space $[K] = \{1, 2, \ldots, K\}$) (hidden state sequence);

Hidden Markov models

Hidden Markov model (HMM): a Markov chain $\{(X_t, Y_t)\}_{t\in\mathbb{N}}$, where

- $\{Y_t\}_{t\in\mathbb{N}}$ is also a Markov chain (with state space $[K] = \{1, 2, \ldots, K\}$) (hidden state sequence);
- conditioned on Y_t, corresponding X_t is *independent of all other variables*;

Hidden Markov models

Hidden Markov model (HMM): a Markov chain $\{(X_t, Y_t)\}_{t\in\mathbb{N}}$, where

- $\{Y_t\}_{t\in\mathbb{N}}$ is also a Markov chain (with state space $[K] = \{1, 2, \ldots, K\}$) (hidden state sequence);
- conditioned on Y_t, corresponding X_t is *independent of all other variables*;
- the Y_t are *hidden*, and the X_t are *observed*.

Hidden Markov models

Hidden Markov model (HMM): a Markov chain $\{(X_t, Y_t)\}_{t\in\mathbb{N}}$, where

- $\{Y_t\}_{t\in\mathbb{N}}$ is also a Markov chain (with state space $[K] = \{1, 2, \ldots, K\}$) (hidden state sequence);
- conditioned on Y_t, corresponding X_t is *independent of all other variables*;
- the Y_t are *hidden*, and the X_t are *observed*.

$$
\begin{array}{ccccccccc}
\cdots & \longrightarrow & Y_{t-1} & \longrightarrow & Y_t & \longrightarrow & Y_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & X_{t-1} & & X_t & & X_{t+1} & &
\end{array}
$$

Hidden Markov models

Hidden Markov model (HMM): a Markov chain $\{(X_t, Y_t)\}_{t\in\mathbb{N}}$, where

- $\{Y_t\}_{t\in\mathbb{N}}$ is also a Markov chain (with state space $[K] = \{1, 2, \ldots, K\}$) (hidden state sequence);
- conditioned on Y_t, corresponding X_t is *independent of all other variables*;
- the Y_t are *hidden*, and the X_t are *observed*.

$$
\begin{array}{ccccccccc}
\cdots & \longrightarrow & Y_{t-1} & \longrightarrow & Y_t & \longrightarrow & Y_{t+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & X_{t-1} & & X_t & & X_{t+1} & &
\end{array}
$$

Time-homogeneous HMM: *conditional* distribution of X_t given Y_t does not depend on t. (We'll focus on these.)

Hidden Markov models

Hidden Markov model (HMM): a Markov chain $\{(X_t, Y_t)\}_{t \in \mathbb{N}}$, where

- $\{Y_t\}_{t \in \mathbb{N}}$ is also a Markov chain (with state space $[K] = \{1, 2, \ldots, K\}$) (hidden state sequence);
- conditioned on Y_t, corresponding X_t is *independent of all other variables*;
- the Y_t are *hidden*, and the X_t are *observed*.

$$\begin{array}{ccccccccc} \cdots & \longrightarrow & Y_{t-1} & \longrightarrow & Y_t & \longrightarrow & Y_{t+1} & \longrightarrow & \cdots \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & X_{t-1} & & X_t & & X_{t+1} & & \end{array}$$

Time-homogeneous HMM: *conditional* distribution of X_t given Y_t does not depend on t. (We'll focus on these.)

Useful subscript notation: $Y_{s:t} = (Y_s, Y_{s+1}, \ldots, Y_t)$ for $s \leq t$.

HMM parameters (discrete observations)

For time-homogeneous HMM where X_t takes values in $[D] = \{1, 2, \ldots, D\}$:

HMM parameters (discrete observations)

For time-homogeneous HMM where X_t takes values in $[D] = \{1, 2, \ldots, D\}$:

- **Initial state distribution**: K-dimensional probability vector $\boldsymbol{\pi}$

$$\pi_i = \Pr(Y_1 = i).$$

HMM parameters (discrete observations)

For time-homogeneous HMM where X_t takes values in $[D] = \{1, 2, \ldots, D\}$:

- **Initial state distribution**: K-dimensional probability vector $\boldsymbol{\pi}$

$$\pi_i \;=\; \Pr(Y_1 = i).$$

- **Transition matrix**: $K \times K$ matrix $\boldsymbol{A}$

$$A_{i,j} \;=\; \Pr(Y_{t+1} = j \mid Y_t = i)$$

  (rows of $\boldsymbol{A}$ are probability vectors).

HMM parameters (discrete observations)

For time-homogeneous HMM where X_t takes values in $[D] = \{1, 2, \ldots, D\}$:

- **Initial state distribution**: K-dimensional probability vector $\boldsymbol{\pi}$

$$\pi_i = \Pr(Y_1 = i).$$

- **Transition matrix**: $K \times K$ matrix $\boldsymbol{A}$

$$A_{i,j} = \Pr(Y_{t+1} = j \mid Y_t = i)$$

  (rows of $\boldsymbol{A}$ are probability vectors).

- **Emission matrix**: $K \times D$ matrix $\boldsymbol{B}$

$$B_{i,j} = \Pr(X_t = j \mid Y_t = i)$$

  (rows of $\boldsymbol{B}$ are probability vectors).

Solution: Viterbi algorithm and several other approaches

Connections to mixture models

Mixture model

$$\begin{array}{c} Y \\ \downarrow \\ \boldsymbol{X} \end{array}$$

(Y is hidden, $\boldsymbol{X}$ is observed.)

Hidden Markov model

$$\begin{array}{ccccc} Y_1 & \longrightarrow & Y_2 & \longrightarrow \cdots \longrightarrow & Y_\ell \\ \downarrow & & \downarrow & & \downarrow \\ X_1 & & X_2 & & X_\ell \end{array}$$

($Y_{1:\ell}$ is hidden, $X_{1:\ell}$ is observed.)

Connections to mixture models

Mixture model

$$\begin{array}{c} Y \\ \downarrow \\ \boldsymbol{X} \end{array}$$

(Y is hidden, $\boldsymbol{X}$ is observed.)

For K component mixture model,
Y takes values in $[K]$.

Hidden Markov model

$$\begin{array}{ccccc} Y_1 & \longrightarrow & Y_2 & \longrightarrow \cdots \longrightarrow & Y_\ell \\ \downarrow & & \downarrow & & \downarrow \\ X_1 & & X_2 & & X_\ell \end{array}$$

($Y_{1:\ell}$ is hidden, $X_{1:\ell}$ is observed.)

Connections to mixture models

Mixture model

$$\begin{array}{c} Y \\ \downarrow \\ \boldsymbol{X} \end{array}$$

(Y is hidden, $\boldsymbol{X}$ is observed.)

For K component mixture model,
Y takes values in $[K]$.

Hidden Markov model

$$\begin{array}{ccccc} Y_1 & \longrightarrow & Y_2 & \longrightarrow \cdots \longrightarrow & Y_\ell \\ \downarrow & & \downarrow & & \downarrow \\ X_1 & & X_2 & & X_\ell \end{array}$$

($Y_{1:\ell}$ is hidden, $X_{1:\ell}$ is observed.)

For sequence of length ℓ,
$Y_{1:\ell}$ takes values in $[K]^\ell$.

Connections to mixture models

Mixture model

$$\begin{array}{c} Y \\ \downarrow \\ \boldsymbol{X} \end{array}$$

(Y is hidden, $\boldsymbol{X}$ is observed.)

For K component mixture model,
Y takes values in $[K]$.

Hidden Markov model

$$\begin{array}{ccccc} Y_1 & \longrightarrow & Y_2 & \longrightarrow \cdots \longrightarrow & Y_\ell \\ \downarrow & & \downarrow & & \downarrow \\ X_1 & & X_2 & & X_\ell \end{array}$$

($Y_{1:\ell}$ is hidden, $X_{1:\ell}$ is observed.)

For sequence of length ℓ,
$Y_{1:\ell}$ takes values in $[K]^\ell$.

Graphical diagram for HMM correctly suggests that every path—even ignoring arrow directions—is a Markov chain!

Connections to mixture models

Mixture model

$$
\begin{array}{c}
Y \\
\downarrow \\
\boldsymbol{X}
\end{array}
$$

(Y is hidden, $\boldsymbol{X}$ is observed.)

For K component mixture model,
Y takes values in $[K]$.

Hidden Markov model

$$
\begin{array}{ccccccc}
Y_1 & \longrightarrow & Y_2 & \longrightarrow \cdots \longrightarrow & Y_\ell \\
\downarrow & & \downarrow & & \downarrow \\
X_1 & & X_2 & & X_\ell
\end{array}
$$

($Y_{1:\ell}$ is hidden, $X_{1:\ell}$ is observed.)

For sequence of length ℓ,
$Y_{1:\ell}$ takes values in $[K]^\ell$.

Graphical diagram for HMM correctly suggests that every path—even ignoring arrow directions—is a Markov chain!

- $Y_1 \to Y_2 \to X_2$

Connections to mixture models

Mixture model

$$\begin{array}{c} Y \\ \downarrow \\ \boldsymbol{X} \end{array}$$

(Y is hidden, $\boldsymbol{X}$ is observed.)

For K component mixture model, Y takes values in $[K]$.

Hidden Markov model

$$\begin{array}{ccccccc} Y_1 & \longrightarrow & Y_2 & \longrightarrow \cdots \longrightarrow & Y_\ell \\ \downarrow & & \downarrow & & \downarrow \\ X_1 & & X_2 & & X_\ell \end{array}$$

($Y_{1:\ell}$ is hidden, $X_{1:\ell}$ is observed.)

For sequence of length ℓ, $Y_{1:\ell}$ takes values in $[K]^\ell$.

Graphical diagram for HMM correctly suggests that every path—even ignoring arrow directions—is a Markov chain!

- $Y_1 \to Y_2 \to X_2$
- $X_2 \to Y_2 \to Y_3 \to X_3$

Connections to mixture models

Mixture model

$$\begin{array}{c} Y \\ \downarrow \\ \boldsymbol{X} \end{array}$$

(Y is hidden, $\boldsymbol{X}$ is observed.)

For K component mixture model,
Y takes values in $[K]$.

Hidden Markov model

$$\begin{array}{ccccccc} Y_1 & \longrightarrow & Y_2 & \longrightarrow \cdots \longrightarrow & Y_\ell \\ \downarrow & & \downarrow & & \downarrow \\ X_1 & & X_2 & & X_\ell \end{array}$$

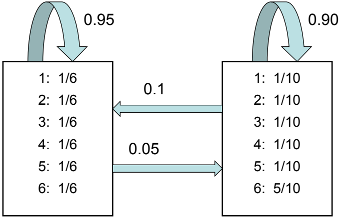
($Y_{1:\ell}$ is hidden, $X_{1:\ell}$ is observed.)

For sequence of length ℓ,
$Y_{1:\ell}$ takes values in $[K]^\ell$.

Graphical diagram for HMM correctly suggests that every path—even ignoring arrow directions—is a Markov chain!

- $Y_1 \to Y_2 \to X_2$
- $X_2 \to Y_2 \to Y_3 \to X_3$
- $X_1 \to Y_1 \to Y_{2:\ell} \to X_{2:\ell}$
- ...

Example: dishonest casino

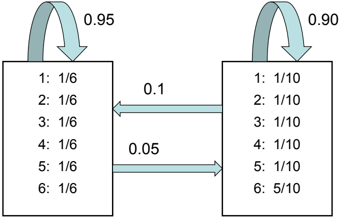


Casino die-rolling game:

Randomly switch between two possible dice: one is fair, the other loaded.

The dice are otherwise indistinguishable!

Example: dishonest casino



Casino die-rolling game:

Randomly switch between two possible dice: one is fair, the other loaded.

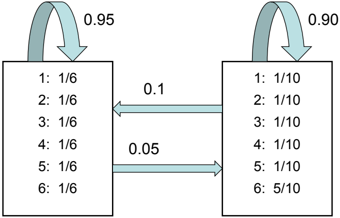
The dice are otherwise indistinguishable!

HMM parameters:

$$\boldsymbol{A} = \begin{array}{c c} & \begin{array}{cc} \text{fair die} & \text{loaded die} \end{array} \\ \begin{array}{c} \text{fair die} \\ \text{loaded die} \end{array} & \begin{pmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{pmatrix} \end{array}, \quad \boldsymbol{B} = \begin{array}{c c} & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{array}{c} \text{fair die} \\ \text{loaded die} \end{array} & \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{2} \end{pmatrix} \end{array},$$

and $\boldsymbol{\pi} = (1, 0)$ if the casino starts out with the fair die.

Example: dishonest casino



Casino die-rolling game:

Randomly switch between two possible dice: one is fair, the other loaded.

The dice are otherwise indistinguishable!

HMM parameters:

Transition matrix

Emision matrix

$$\boldsymbol{A} = \begin{array}{c} \\ \text{fair die} \\ \text{loaded die} \end{array} \begin{array}{c} \begin{array}{cc} \text{fair die} & \text{loaded die} \end{array} \\ \begin{pmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{pmatrix} \end{array}, \quad \boldsymbol{B} = \begin{array}{c} \\ \text{fair die} \\ \text{loaded die} \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{2} \end{pmatrix} \end{array},$$

and $\boldsymbol{\pi} = (1, 0)$ if the casino starts out with the fair die.

Problem: Based on a sequence of rolls, guess which die was used at each time.

HMM inference/learning problems

Conditional probabilities (e.g., filtering/smoothing)

- **Given**: parameters $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})$, observation sequence $x_{1:\ell} \in [D]^{\ell}$.
- **Goal**: conditional distribution of $Y_{s:t}$ given $X_{1:\ell} = x_{1:\ell}$ $(1 \leq s \leq t \leq \ell)$:

$$\Pr_{\boldsymbol{\theta}}\left(Y_{s:t} = y_{s:t} \,|\, X_{1:\ell} = x_{1:\ell}\right), \quad \text{for each } y_{s:t} \in [K]^{t-s+1}.$$

HMM inference/learning problems

Conditional probabilities (e.g., filtering/smoothing)

- **Given**: parameters $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})$, observation sequence $x_{1:\ell} \in [D]^\ell$.
- **Goal**: conditional distribution of $Y_{s:t}$ given $X_{1:\ell} = x_{1:\ell}$ ($1 \leq s \leq t \leq \ell$):

$$\Pr_{\boldsymbol{\theta}}\bigl(Y_{s:t} = y_{s:t} \mid X_{1:\ell} = x_{1:\ell}\bigr), \quad \text{for each } y_{s:t} \in [K]^{t-s+1}.$$

Most probable state sequence (decoding)

- **Given**: parameters $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})$, observation sequence $x_{1:\ell} \in [D]^\ell$.
- **Goal**: $\displaystyle\operatorname*{arg\,max}_{y_{1:\ell} \in [K]^\ell} \Pr_{\boldsymbol{\theta}}\bigl(Y_{1:\ell} = y_{1:\ell} \mid X_{1:\ell} = x_{1:\ell}\bigr)$.

HMM inference/learning problems

Conditional probabilities (e.g., filtering/smoothing)

- **Given**: parameters $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})$, observation sequence $x_{1:\ell} \in [D]^\ell$.
- **Goal**: conditional distribution of $Y_{s:t}$ given $X_{1:\ell} = x_{1:\ell}$ ($1 \le s \le t \le \ell$):

$$\Pr_{\boldsymbol{\theta}}\bigl(Y_{s:t} = y_{s:t} \,|\, X_{1:\ell} = x_{1:\ell}\bigr), \quad \text{for each } y_{s:t} \in [K]^{t-s+1}.$$

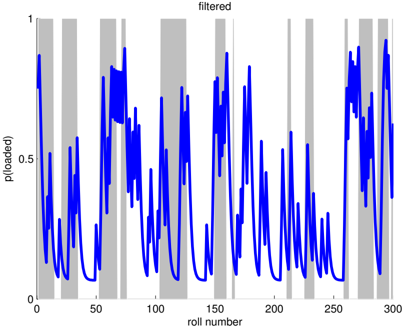
Most probable state sequence (decoding)

- **Given**: parameters $\boldsymbol{\theta} = (\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B})$, observation sequence $x_{1:\ell} \in [D]^\ell$.
- **Goal**: $\displaystyle\operatorname*{arg\,max}_{y_{1:\ell} \in [K]^\ell} \Pr_{\boldsymbol{\theta}}\bigl(Y_{1:\ell} = y_{1:\ell} \,|\, X_{1:\ell} = x_{1:\ell}\bigr)$.

Parameter estimation

- **Given**: n observation sequences $x_{1:\ell}^{(s)}$ for $s \in [n]$.
- **Goal**: parameter estimates $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\pi}}, \widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}})$.

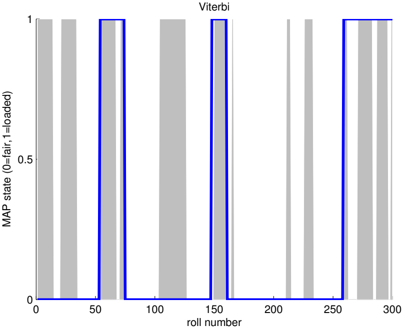
Example: dishonest casino



Conditional probability

Gray bars: Loaded dice used.

Blue: $\Pr_{\boldsymbol{\theta}}(Y_t = \text{loaded}|X_{1:\ell} = x_{1:\ell})$



Decoding

Gray bars: Loaded dice used.

Blue: Most probable state Z_t.

Some applications

- **Bioinformatics**

  Observations: amino acids in a protein

  Hidden states: indicators of evolutionary conservation

Some applications

- **Bioinformatics**

  Observations: amino acids in a protein

  Hidden states: indicators of evolutionary conservation

- **Natural language processing**

  Observations: words in a sentence

  Hidden states: words' part-of-speech or other word-type semantics

Some applications

- **Bioinformatics**

  Observations: amino acids in a protein

  Hidden states: indicators of evolutionary conservation

- **Natural language processing**

  Observations: words in a sentence

  Hidden states: words' part-of-speech or other word-type semantics

- **Speech recognition**

  Observations: recorded speech at various (discrete) times

  Hidden states: phonemes that the speaker intended to vocalize

- **Financial market cycles forecast:**
  Observations: index of stock market
  Hidden states: bull or bear market