Hidden Markov Model
Motivation

Sequential data often arise from measurements of time series

Goal

Predict new values given old observations

Weather Forecast

Person’s position over time
Markov Model

Idea
Restrict the connectivity of future states to previous states
Outline

- Markov Model
- Hidden Markov Model
- Inference using HMM
- Case Study: Speech Recognition
- Conclusion
Markov Model

- **Markov Models**: Only assume a certain number of previous states to be relevant

- *First*-order Markov Models: Only the last state is relevant for the current state
Markov Model

• Recall: Chain Rule

\[ p(x_1, x_2, x_3, \ldots, x_n) = \prod_{k=1}^{N} p(x_k \mid x_1, \ldots, x_{k-1}) \]

• First-order Markov Models simplify this to

\[ p(x_1, \ldots, x_K) = p(x_1) \prod_{k=2}^{K} p(x_k \mid x_{k-1}) \]

by only taking

the last state into account
Example

<table>
<thead>
<tr>
<th>$R_{k-1}$</th>
<th>$p(R_k)$</th>
<th>$p(-R_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>false</td>
<td>0.3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Transition model
State Space Model

**Input:** Human trajectory in 2d

**Output:** Classification
- Walking straight
- Right turn
- Left turn

**Problem:**
State space is not observation space
Hidden Markov Model
Is it raining outside??
Weather (Rain/ No rain) → Hidden

Umbrella (Yes/ No) → Observation

The rain **causes** the umbrella to appear!
Hidden Markov Model

- Weather
- Latent (Hidden) State Variable: Discrete
- Umbrella
- Observation Variable: Discrete or Continuous
An observation at time k is only dependent on the state of time k and independent of all states from the beginning to k-1.
Hidden Markov Model

\[ p(x_1, \ldots, x_K, z_1, \ldots, z_K) \]
Hidden Markov Model

\[ p(x_k | x_{k-1}) \]

\[ p(x_0) \]

\[ p(z_k | x_k) \]
Transition Model

\[ p(x_k \mid x_{0:k-1}) = p(x_k \mid x_{k-1}) \]

\[ A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \quad \sum_j A_{1j} = 1 \]

\( A_{32} \) represents the probability of a transition from state 3 to state 2.
Transition Model
Observation Model

\[ p(z_k \mid x_{0:k}, z_{0:k-1}) = p(z_k \mid x_k) \]

\[
E = \begin{pmatrix}
E_{11} & E_{12} & E_{13} & E_{14} & E_{15} \\
E_{21} & E_{22} & E_{23} & E_{24} & E_{25} \\
E_{31} & E_{32} & E_{33} & \boxed{E_{34}} & E_{35}
\end{pmatrix}
\]

\[ \sum_j E_{1j} = 1 \]

Emission probability of symbol 4 from state 3
Prior Probability
Hidden Markov Model

\[ p(x_{0:K}, z_{1:K}) = p(x_0) \prod_{k=1}^{K} p(x_k|x_{k-1}) p(z_k|x_k) \]

- **Prior**
- **Transition model**
- **Observation model**
Hidden Markov Model

Transition model

<table>
<thead>
<tr>
<th>$R_{k-1}$</th>
<th>$p(R_k)$</th>
<th>$p(\neg R_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>false</td>
<td>0.3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Observation model

<table>
<thead>
<tr>
<th>$R_k$</th>
<th>$p(U_k)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>0.9</td>
</tr>
<tr>
<td>false</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Important Note

- First-order Markov Models are often inaccurate, since you throw away any information about the past.
- Increasing the accuracy by more general models:
  - Increasing the order of the Markov model
  - Adding more state variables.
Inference
Hidden Markov Model

\[
\begin{array}{c|cc}
R_{k-1} & p(R_k) & p(\neg R_k) \\
\hline
\text{true} & 0.7 & 0.3 \\
\text{false} & 0.3 & 0.7 \\
\end{array}
\]

\text{Transition model}

Prior
\[
\begin{array}{c|cc}
p(R_0) & p(\neg R_0) \\
\hline
0.5 & 0.5 \\
\end{array}
\]

Observation model
\[
\begin{array}{c|c}
R_k & p(U_k) \\
\hline
\text{true} & 0.9 \\
\text{false} & 0.2 \\
\end{array}
\]
Inference Tasks

1. Filtering

\[ p(\mathbf{x}_k \mid \mathbf{z}_{1:k}) \]
Inference Tasks

1. Filtering
   \[ p(x_k \mid z_{1:k}) \]

2. Smoothing
   \[ p(x_t \mid z_{1:k}) \quad 0 \leq t < k \]

Past
Inference Tasks

1. Filtering
   \[ p(x_k \mid z_{1:k}) \]

2. Smoothing
   \[ p(x_t \mid z_{1:k}) \quad 0 \leq t < k \]

3. Prediction
   \[ p(x_{k+t} \mid z_{1:k}) \quad t > 0 \]
Inference Tasks

1. Filtering
   \[ p(x_k \mid z_{1:k}) \]

2. Smoothing
   \[ p(x_t \mid z_{1:k}) \quad 0 \leq t < k \]

3. Prediction
   \[ p(x_{k+t} \mid z_{1:k}) \quad t > 0 \]

4. Most Likely Sequence
   \[ \arg\max_{x_{1:k}} p(x_{1:k} \mid z_{1:k}) \]
Inference Example
What is the most likely weather sequence??
Most Likely Sequence

Recall, we want to find the state sequence $x^*$ that maximizes the probability along its path, i.e. $x^* = \arg \max_{x_{1:k}} p(x_{1:k} \mid z_{1:k})$.
Viterbi Algorithm

• Maximizing path for $k-1$

best path for $x_{k-1} = 1$

best path for $x_{k-1} = 3$

best path for $x_{k-1} = 2$
The algorithm computes
\[ \mu_k = \max_{x_{1:k-1}} p(x_{1:k}, z_{1:k}) \]
as
\[ \mu_k = p(z_k | x_k) \max_{x_{k-1}} \left( p(x_k | x_{k-1}) \mu_{k-1} \right) \]
Viterbi Algorithm

\[ U_k = \text{true} \]

\[
\begin{array}{c}
.8182 \\
.1818 \\
\end{array}
\]

\[
\begin{array}{c}
.5155 \\
.0491 \\
\end{array}
\]

\[
\begin{array}{c}
.0361 \\
.1237 \\
\end{array}
\]

\[
\begin{array}{c}
.0334 \\
.0173 \\
\end{array}
\]

\[
\begin{array}{c}
.0210 \\
.0024 \\
\end{array}
\]

Source [1]
Speech Recognition

- States?
- Observations?
- Transition Model?
- Observation Model?
- Prior?
**Strengths & Weaknesses of HMM**

- HMM provides better results than MM
- Principle of HMM can be adapted to many problems, e.g. finding alignment
- Computationally expensive, both in memory and time
- Need more training than classical Markov Models
Summary

Markov Model

Hidden MM

Transition Model

Observation Model

Prior
Summary

Inference using HMM

- Filtering
- Smoothing
- Prediction
- Most likely Sequence
  - Viterbi Algorithm
Thank you
Sources

Most important graphics from this presentation are taken from