Hidden Markov Models: A Detailed Overview

1. Introduction

A **Hidden Markov Model (HMM)** is a probabilistic model used to describe systems that are assumed to be a Markov process with unobserved (hidden) states.

Applications: speech recognition, bioinformatics, finance, and natural language processing.

2. Components of an HMM

An HMM is defined by the following:

- A set of hidden states: $Z_1, Z_2, \dots, Z_T$, where each $Z_t \in \{1, 2, \dots, K\}$
- A set of observations: $X_1, X_2, \dots, X_T$, with $X_t \in \mathcal{O}$
- Initial state distribution: $\boldsymbol{\pi} = [\pi_1, \dots, \pi_K]$, where $\pi_k = P(Z_1 = k)$
- Transition probabilities: $A = [a_{ij}]$, where $a_{ij} = P(Z_{t+1} = j \mid Z_t = i)$
- Emission probabilities: $B = [b_j(o)]$, where $b_j(o) = P(X_t = o \mid Z_t = j)$

3. Joint Probability

The joint probability of a sequence of hidden states $\boldsymbol{z}$ and observations $\boldsymbol{x}$ is:

$$P(\boldsymbol{x}, \boldsymbol{z}) = P(Z_1) \prod_{t=2}^{T} P(Z_t \mid Z_{t-1}) \prod_{t=1}^{T} P(X_t \mid Z_t)$$

4. Inference Tasks

- **Evaluation:** Compute the likelihood of a sequence: $P(\boldsymbol{x})$
- **Decoding:** Find the most likely sequence of states: $\arg\max_{\boldsymbol{z}} P(\boldsymbol{z} \mid \boldsymbol{x})$
- **Learning:** Estimate the parameters $(\boldsymbol{\pi}, A, B)$ from data

5. Forward Algorithm (Evaluation)

Define forward probabilities:

$$\alpha_t(j) = P(X_1, \ldots, X_t, Z_t = j)$$

Recursion:

$$\alpha_1(j) = \pi_j b_j(x_1)$$
$$\alpha_t(j) = \left[\sum_{i=1}^{K} \alpha_{t-1}(i) a_{ij} \right] b_j(x_t)$$

Final probability:

$$P(\boldsymbol{x}) = \sum_{j=1}^{K} \alpha_T(j)$$

6. Viterbi Algorithm (Decoding)

Define:

$$\delta_t(j) = \max_{z_1, \ldots, z_{t-1}} P(Z_1, \ldots, Z_{t-1}, Z_t = j, X_1, \ldots, X_t)$$

Recursion:

$$\delta_1(j) = \pi_j b_j(x_1)$$
$$\delta_t(j) = \max_i [\delta_{t-1}(i) a_{ij}] \cdot b_j(x_t)$$

Backtrace to recover the best state sequence.

7. EM for HMM (Baum-Welch Algorithm)

We aim to maximize $P(\boldsymbol{x} \mid \theta)$ with latent states $\boldsymbol{z}$. Use EM:

E-step:

Compute:

$$\gamma_t(i) = P(Z_t = i \mid \boldsymbol{x}), \quad \xi_t(i, j) = P(Z_t = i, Z_{t+1} = j \mid \boldsymbol{x})$$

Use forward-backward algorithm:

$$\beta_t(i) = P(X_{t+1}, \ldots, X_T \mid Z_t = i)$$

M-step:

Update parameters:

$$\pi_i = \gamma_1(i)$$
$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$
$$b_j(o) = \frac{\sum_{t:x_t=o} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

8. Summary

Hidden Markov Models are powerful tools for modeling temporal or sequential data. They combine latent state modeling with efficient algorithms for inference and learning.