Hidden Markov Model for Credit Risk Progression

Tran Quoc Thai
Student ID: 2370759

April 10, 2025

1. Introduction

A Hidden Markov Model (HMM) is a statistical model in which the system being modeled is assumed to follow a Markov process with unobservable (hidden) states. In this case, we model credit risk progression over time with states $S = \{\text{Low}, \text{Medium}, \text{High}\}$ and observed credit scores.

2. Model Specification

2.1. States and Transition Probabilities

Given states $S = \{L, M, H\}$ for Low, Medium, and High risk respectively, we define the transition probability matrix:

$$P = \begin{bmatrix} 0.7 & 0.3 & 0.0 \\ 0.0 & 0.6 & 0.4 \\ 0.0 & 0.2 & 0.8 \end{bmatrix}. \tag{1}$$

2.2. Emission Probabilities

Let $O = [710, 650, 680]$ be the sequence of observed credit scores. We assume a Gaussian distribution $P(o_t|s_i) \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for each state. Given the means and standard deviations:

$$\begin{aligned} \mu_L &= 720, \quad \sigma_L = 20, \\ \mu_M &= 650, \quad \sigma_M = 30, \\ \mu_H &= 600, \quad \sigma_H = 40, \end{aligned}$$

the emission probabilities are computed as:

$$P(o_t|s_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(o_t - \mu_i)^2}{2\sigma_i^2}\right). \tag{2}$$

3. Viterbi Algorithm for Decoding Most Likely Sequence

The Viterbi algorithm finds the most probable sequence of hidden states given the observations. Define:

- $\delta_t(i)$: probability of the most likely state sequence ending in state i at time t.
- $\psi_t(i)$: backpointer to reconstruct the optimal state sequence.

3.1. Initialization

Assuming equal initial probabilities $P(S_0 = i) = \frac{1}{3}$:

$$\delta_1(i) = P(S_0 = i)P(o_1|s_i). \tag{3}$$

3.2. Recursion

For each subsequent observation:

$$\delta_t(j) = \max_i \left[\delta_{t-1}(i)P(i \to j)\right] P(o_t|s_j), \tag{4}$$

$$\psi_t(j) = \arg\max_i \left[\delta_{t-1}(i)P(i \to j)\right]. \tag{5}$$

3.3. Termination and Backtracking

The most likely final state is:

$$S_T^* = \arg\max_i \delta_T(i), \tag{6}$$

and the sequence is reconstructed via backtracking:

$$S_t^* = \psi_{t+1}(S_{t+1}^*). \tag{7}$$

4. Conclusion

The Viterbi algorithm enables inferring the most probable sequence of credit risk states when analyzing observed credit scores. Using probability distributions as a soft decision approach enhances classification robustness because it outperforms those methods that use hard thresholding approaches.