Classification of household devices based on electricity consumption using HMM

Julia Kiczka

June 9, 2024

1 Primary objective

The primary objective of this project is to develop a system that can automatically identify and categorize the devices responsible for the observed electricity consumption patterns in a household. We are given the electricity consumption data (time series) for 5 devices: lighting2, lighting5, lighting4, microwave and refrigerator. For each of the devices, we train separate Hidden Markov Model $\lambda_1 \dots \lambda_5$. Subsequently, we will use these trained models for classification: for each unknown device we will choose one of $\lambda_1 \dots \lambda_5$ with the highest likelihood. This corresponds to choosing the model for which it was most likely to generate such electricity consumption.

2 Hidden Markov Models (HMMs)

A Hidden Markov Model (HMM) is a statistical model used to model sequences of observable events, where the underlying process generating the events is assumed to be a Markov process with hidden states.

2.1 Components

An HMM consists of the following components:

- **Hidden States**: The system possesses a set of hidden states, which are not directly observable. Denote the set of hidden states as $S = \{s_1, s_2, \dots, s_N\}$.
- **Observations**: Each hidden state emits observable symbols or observations according to some probability distribution. Denote the set of possible observations as $V = \{v_1, v_2, \dots, v_M\}$.
- **State Transition Probabilities**: The system transitions between hidden states over time. These transitions are governed by a set of state transition probabilities a_{ij}, where a_{ij} represents the probability of transitioning from state s_i to state s_j.
- **Emission Probabilities**: Each hidden state emits observable symbols with certain probabilities. These emission probabilities are represented by $b_j(k)$, where $b_j(k)$ denotes the probability of emitting observation v_k while in state s_j.
- **Initial State Distribution**: The system starts in one of the hidden states with certain probabilities. The initial state distribution π_i represents the probability of starting in state s_i.

2.2 Model Parameters

The parameters of a Hidden Markov Model (HMM) include the state transition probabilities a_{ij}, the emission probabilities $b_j(k)$, and the initial state distribution π_i. For the purpose of this project, we make the assumption that each state emits observations according to a random variable with a Gaussian distribution. Our objective will be to determine the optimal number of hidden states for each model - the number that results in the best fit to the data.

2.3 Inference

Given a sequence of observations, generally the goal is to infer the most likely sequence of hidden states that generated the observations. This inference problem is typically solved using algorithms such as the Viterbi algorithm or the forward-backward algorithm.

Evaluation Metrics

Log-Likelihood

The log-likelihood measures the likelihood of observing a sequence of data given a particular HMM. Let $\boldsymbol{X} = x_1, x_2, \ldots, x_T$ be the observed sequence, and $\boldsymbol{Z} = z_1, z_2, \ldots, z_T$ be the corresponding hidden states. The log-likelihood of the observed sequence under the HMM is given by:

$$L(\boldsymbol{X}|\boldsymbol{\Theta}) = \prod_{t=1}^{T} P(x_t|z_t, \boldsymbol{\Theta})$$

where $\boldsymbol{\Theta}$ represents the parameters of the HMM, and $P(x_t|z_t, \boldsymbol{\Theta})$ is the probability of observing x_t at time t given the hidden state z_t and the model parameters $\boldsymbol{\Theta}$. Since we do not know the sequence of hidden states $z_1, \ldots, z_T$, we need to sum over all possible sequences weighted by their probabilities. Although there are exponentially many possible hidden sequences, the likelihood can be computed faster using **forward algorithm** which is based on dynamic programming idea.

Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is a measure of model quality that balances fit to the data with model complexity. It is calculated as:

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

where:

- $\hat{L}$ is the maximized value of the likelihood function of the model M, i.e., $\hat{L} = p(x|\hat{\theta}, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function and x is the observed data.
- k is the number of parameters estimated by the model.

Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC) is similar to AIC but penalizes model complexity more rigorously. It is calculated as:

$$\text{BIC} = k\ln(n) - 2\ln(\hat{L})$$

where:

- $\hat{L}$ is the maximized value of the likelihood function of the model M, i.e., $\hat{L} = p(x|\hat{\theta}, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function and x is the observed data.
- n is the number of data points in x, the number of observations, or equivalently, the sample size.
- k is the number of parameters estimated by the model.

3 Training results

Before training, we perform data splitting: the first 80% of the sequence will be considered as the training set, while the last 20% of each sequence will be considered as the test set. This allows us to evaluate the model's performance on unseen data, ensuring that it generalizes well and does not simply memorize the training data.

3.1 Inizialization

As stated in the 'hmmlearn' library documentation, the Expectation-Maximization (EM) algorithm used for training HMMs is a gradient-based optimization method, which means it can often get stuck in local optima. Therefore, it is generally advisable to run the 'fit' method with various initializations and select the model with the highest score. So in my parameter tuning, I will consider many initial random states and choose the one that allows to get the highest likelihood. Below are the results for each device and the loss functions describing the models $\lambda_1, \ldots, \lambda_5$ based on the number of hidden states (with the best random states chosen). We can observe that the highest likelihood in most cases corresponds to the lowest BIC and AIC (figures 1, 2, 3, 4, 5). Therefore, the number of hidden states at which the highest likelihood is achieved will be considered the optimal number of hidden states for each model. However, in some cases increasing number of hidden states doesn't change likelihood a lot, so it might be better to choose smaller number components to prevent overfitting the data. Table 1 presents the choice of the parameters.

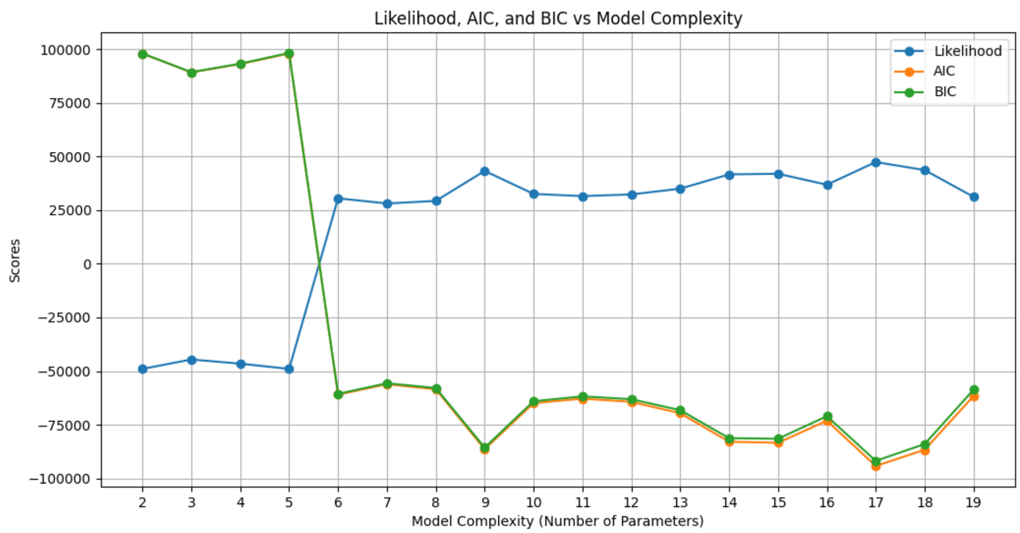

Figure 1: Model λ_1 for lighting2

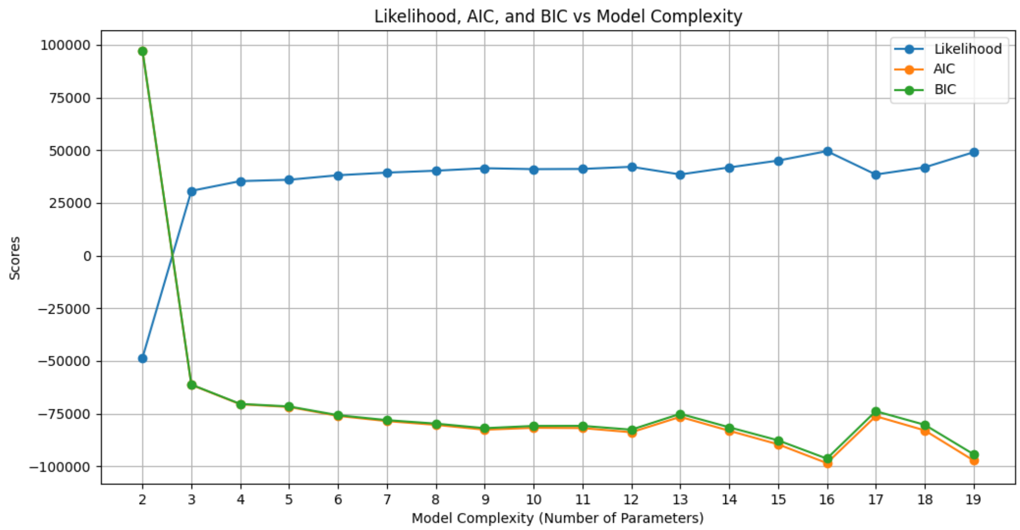

Figure 2: Model λ_2 for lighting5

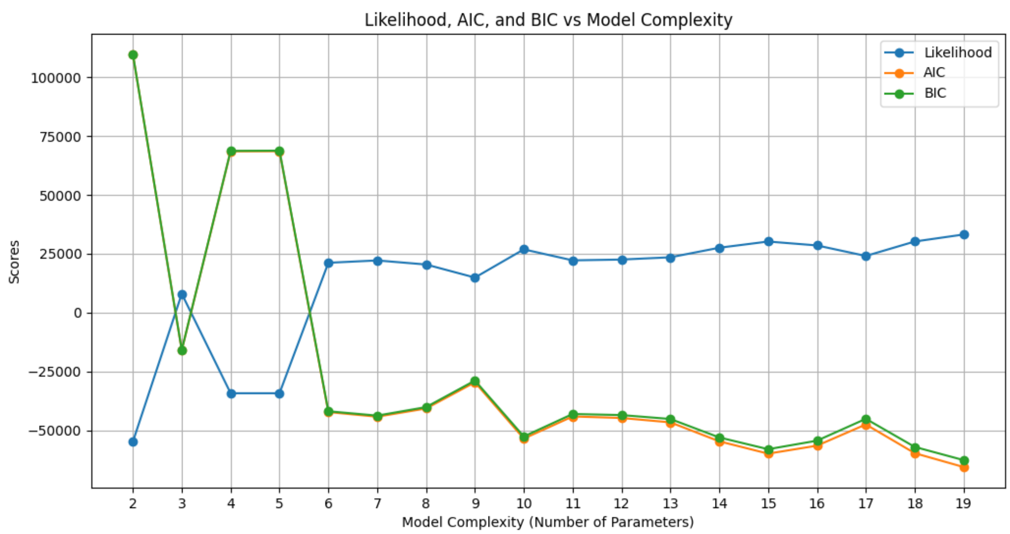

Figure 3: Model λ_3 for lighting4

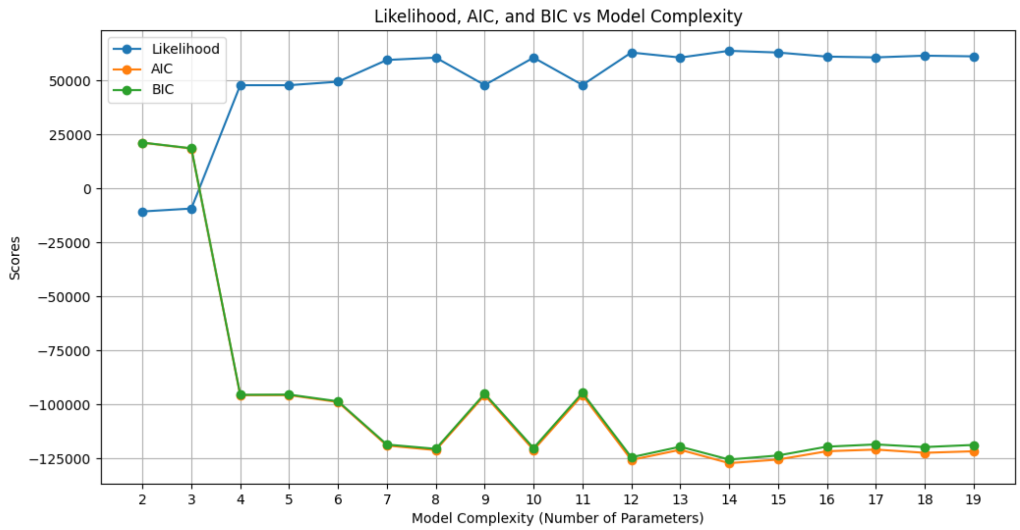

Figure 4: Model λ_4 for microwave

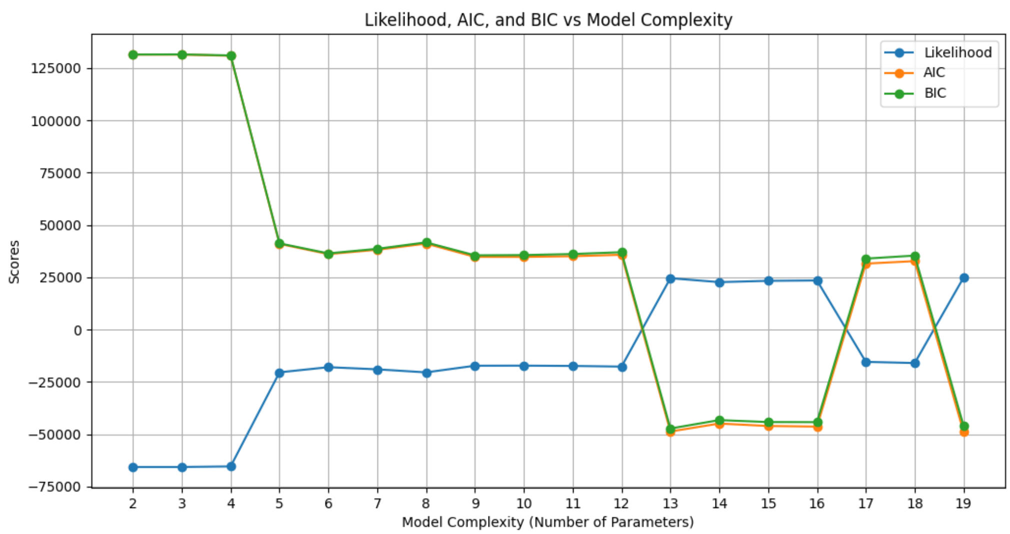

Figure 5: Model λ_5 for refrigerator

The tables 2, 3 present the likelihood of generating a given sequence by a given model for test sequences from each device. I consider original test sequences (20% of the original data) and shorter sequences (of length 100). We can observe that the highest value is obtained for model corresponding to the right device.

3.2 Summary and conclusions

Hidden Markov Models are well-suited for classifying time series related to household appliance power consumption because they effectively capture the sequential nature and temporal dependencies of the usage patterns. By modeling hidden states, HMMs can differentiate between various operational modes of appliances, such as standby, active, and off states, based on observed power consumption data. Additionally, HMMs' probabilistic framework allows for handling the inherent variability and noise in real-world power usage data, leading to more accurate and robust classification.

Table 1: Chosen number of components and random states

Parameters	λ_1	λ_2	λ_3	λ_4	λ_5
components	17	12	15	14	13
random state	22	28	31	42	888

Table 2: Likelihood values on the original test sequences from different devices

Device	λ_1	λ_2	λ_3	λ_4	λ_5
Lighting2	10187.20	-129850.69	5732.35	-9476.91	6363.79
Lighting5	-7016.22	12236.60	-10807.75	-8034.52	-170.54
Lighting4	-940.46	-88858.16	9715.11	-10843.29	6528.19
Refrigerator	-3954.91	-93765.33	-7950.80	14500.64	-16798.85
Microwave	-13619.05	-248707.17	-5535.43	-10569.42	1850.68

Table 3: Likelihood values on the shorter test sequences from different devices

Device	λ_1	λ_2	λ_3	λ_4	λ_5
Lighting2	9639.79	-129837.84	5163.82	-9248.78	5851.18
Lighting5	-6845.83	11657.92	-10498.31	-7837.15	-239.78
Lighting4	-1487.87	-88845.31	9146.58	-10615.16	6015.58
Refrigerator	-3808.99	-90574.15	-7657.34	13932.08	-16195.86
Microwave	-13544.19	-246191.63	-5577.15	-10362.67	1708.31