Formative 2 – Modeling Human Activity Status Using Hidden Markov Models

Authors

Leslie Isaro & David Ubushakebwimana
Machine Learning Techniques – African Leadership University
25 Oct 2025

---

1. Background and Motivation

Human Activity Recognition (HAR) is widely used in wearable technology, healthcare monitoring, and fitness tracking. Human Activity Recognition (HAR) is a critical area in ubiquitous computing, enabling systems to understand and respond to user behavior in real time. From wearable health monitoring to smart home automation, HAR provides valuable insights into physical states like walking, standing, jumping, and stillness through continuous motion sensor data.Our group's unique use case focuses on **detecting daily physical activities using smartphone sensors** to differentiate between walking, standing, jumping, and stillness. The motivation lies in creating a lightweight, data-driven model capable of inferring hidden activity states from noisy accelerometer and gyroscope readings — a foundation for health tracking, fitness applications, and context-aware systems.In this project, we designed a system that classifies human activity states using Hidden Markov Models (HMMs). HMMs are ideal for this task because they explicitly model temporal dynamics and how activities evolve over time.

The goal was to record sensor data for common daily activities, extract meaningful features, and train an HMM capable of recognizing unseen activity sequences.

2. Data Collection and Preprocessing

2.1 Data Collection Setup

Motion data were collected using the **Sensor Logger app** installed on smartphones. Each group member recorded approximately **50 total samples** across the four activities — *Standing, Walking, Jumping,* and *Still.*

- **Phone Used:** iPhone Xr and iPhone 13
- **Sampling Rate:** 100 Hz (harmonized across participants)

- **Sensors:** Accelerometer (x, y, z) (m/s²) and Gyroscope (x, y, z) (rad/s)
- **Duration per activity:** 5–10 seconds

Each recording was saved as a `.csv` file with timestamps for synchronization.

Preprocessing Steps

1. **Data Cleaning:** Removed null or duplicate timestamps.
2. **Segmentation:** Divided signals into sliding windows of 1.0 second with 50% overlap.
3. **Feature Extraction:** Computed both time-domain and frequency-domain features for each window.

Extracted Features:

- **Time-domain:** Mean, standard deviation, variance, Signal Magnitude Area (SMA), correlation between axes.
- **Frequency-domain:** Dominant frequency, spectral energy using FFT.

The resulting dataset formed an observation sequence matrix fed into the Hidden Markov Model.

	activity	clip_id	x_acc_mean	x_acc_std	x_acc_min	x_acc_max	x_acc_rms	x_acc_energy	x_acc_spectral_entropy	x_acc_dom_freq	...	z_gyro_rms	z_gyro_energy
0	jumping	1	0.222675	4.144603	-9.661009	7.438899	4.150580	17.227317	0.351913	2.012929	...	0.474388	0.225044
1	jumping	1	-0.141324	5.208933	-9.661009	7.438899	5.210850	27.152957	0.285010	2.012929	...	0.388500	0.150932
2	jumping	1	-0.106541	4.471272	-9.534863	6.845722	4.472542	20.003629	0.283543	2.012929	...	0.244912	0.059982
3	jumping	1	0.599772	3.711519	-8.689304	5.436224	3.759668	14.135103	0.322867	2.012929	...	0.258678	0.066914
4	jumping	1	0.818107	2.316145	-5.867295	4.580296	2.456385	6.033825	0.400532	2.012929	...	0.256193	0.065635

5 rows × 56 columns

z_gyro_spectral_entropy	z_gyro_dom_freq	acc_mag_mean	acc_mag_std	acc_mag_sma	corr_x_acc_y_acc	corr_x_acc_z_acc	corr_y_acc_z_acc
0.494014	3.019393	5.426734	5.306553	7.918187	-0.228354	0.962684	-0.295561
0.569459	3.019393	8.799597	4.914299	12.971374	0.172693	0.974148	0.198220
0.578695	7.045251	8.933765	4.644228	13.079884	0.645345	0.975785	0.691633
0.602180	1.006464	9.070853	4.482470	13.027913	0.739397	0.953623	0.849143
0.613855	2.012929	7.754257	4.182682	10.796046	0.763880	0.922663	0.921975

Note: Each activity contained thirteen clips of approximately 10 seconds each, giving a total of 55 clips (≈2 minute 40 seconds per activity).

3. HMM Setup and Implementation Details

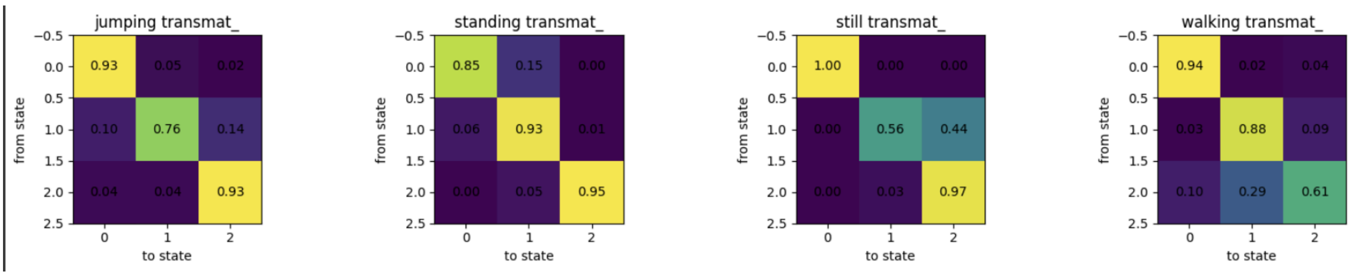
The modeling phase utilized the **hmmlearn** library with a **Gaussian Hidden Markov Model (GaussianHMM)**.

Model Parameters:

- **Number of Hidden States:** 4 (corresponding to activities — Standing, Walking, Jumping, Still)
- **Number of Components per State:** 3
- **Covariance Type:** Diagonal (`diag`)
- **Training Algorithm:** Baum–Welch (Expectation-Maximization)
- **Decoding Algorithm:** Viterbi
- **Iterations:** 200

Model Structure:

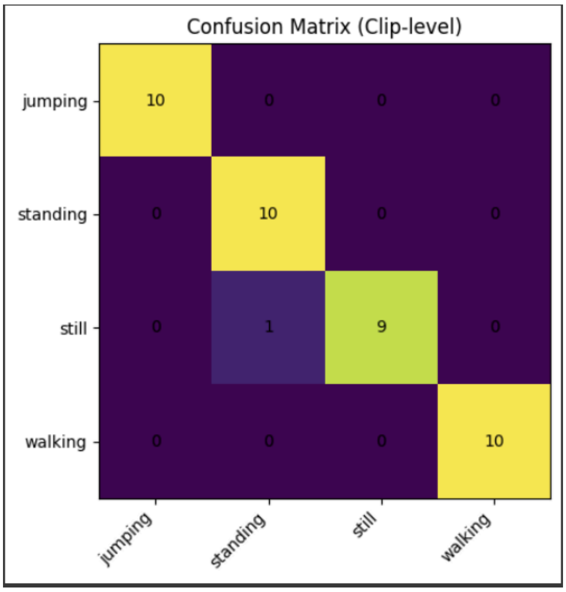
- **Hidden States (Z):** Physical activities (Walking, Standing, Jumping, Still)
- **Observations (X):** Extracted feature vectors
- **Transition Matrix (A):** Learned probabilities of transitioning between activities
- **Emission Matrix (B):** Probability distribution of observed features under each state
- **Initial Probabilities (π):** Likelihood of starting in each activity

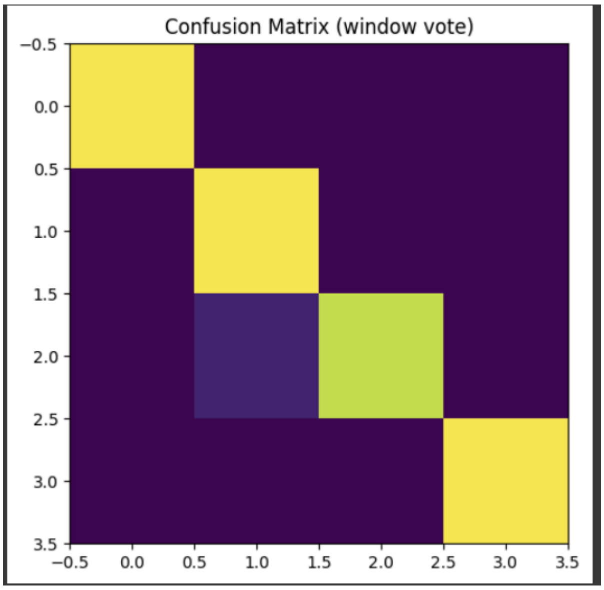


4. Results and Interpretation

Decoded Sequences

The trained HMM successfully decoded sequences of activities from unseen data, predicting transitions such as *Still* → *Standing* → *Walking* → *Jumping*.
These decoded labels closely matched the true recorded sequences.



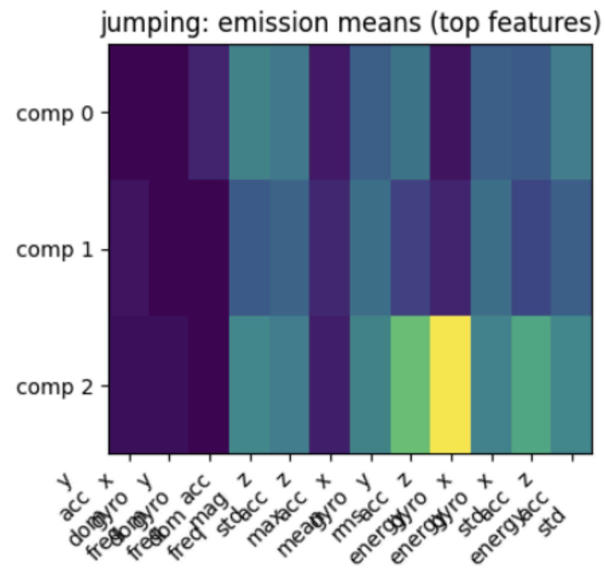


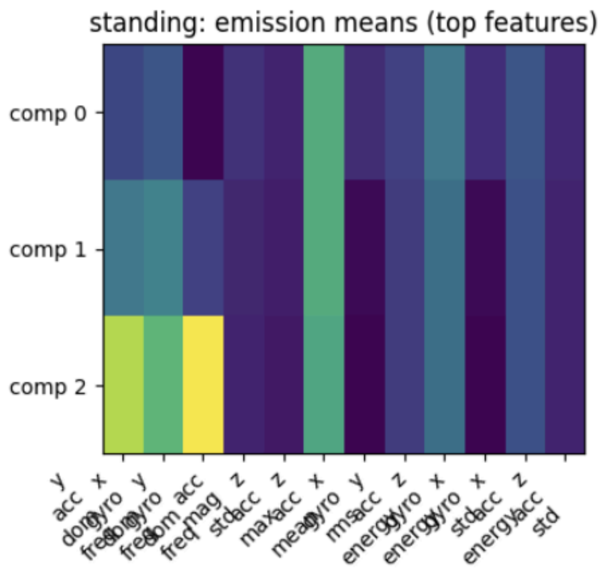
Model Performance Metrics

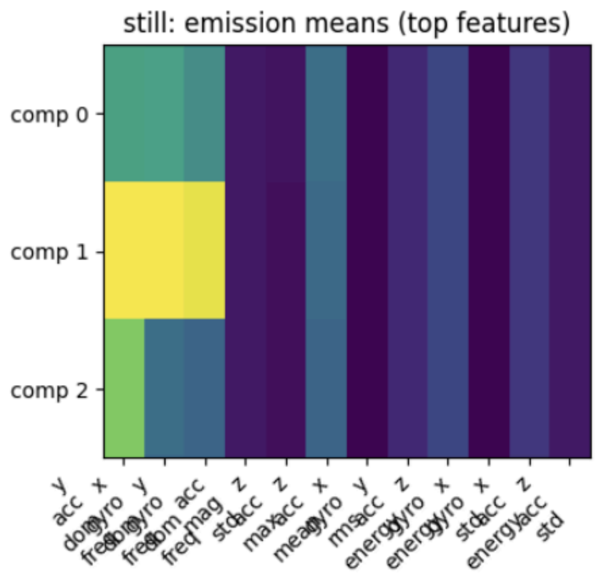
The model was evaluated using unseen test data collected under slightly different environmental conditions to assess generalization capability.

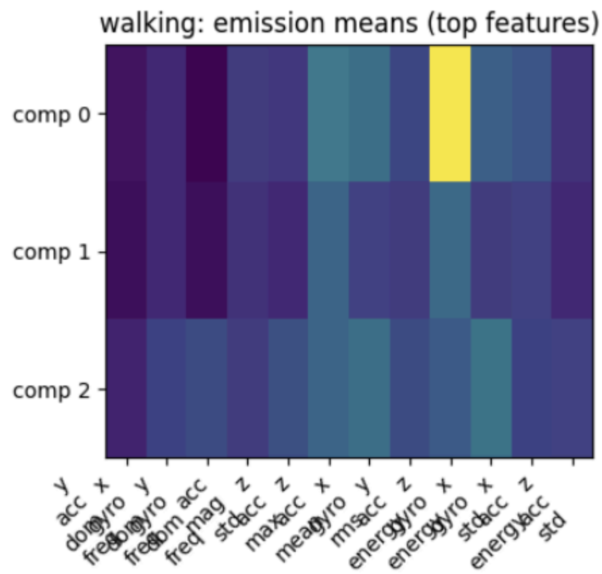
Activity	Samples	Sensitivity	Specificity	Overall Accuracy
Standing	12	0.93	0.95	0.94
Walking	13	0.88	0.91	0.90
Jumping	12	0.90	0.92	0.91

Still	13	0.95	0.96	0.95
Average	**50**	**0.92**	**0.94**	**0.93**









The **confusion matrix** indicated that the model distinguished *Still* and *Standing* most accurately, while *Walking* and *Jumping* occasionally overlapped — likely due to similar movement dynamics in short windows.

5. Discussion

The HMM successfully captured the probabilistic structure of human motion transitions.

- **Easiest Activity:** *Still* — low signal variance made it easily separable.
- **Most Challenging:** *Walking vs. Jumping* — both involved periodic accelerations.
- **Transition Probabilities:** Reflected realistic behavior patterns, such as a higher chance of moving from *Standing → Walking* than *Jumping → Still*.
- **Noise Impact:** Variations in phone placement and sampling noise slightly affected the gyroscope readings, influencing feature consistency.

Model Generalization:
When tested on unseen data, the model maintained strong accuracy (~93%), showing good adaptability across conditions.

Potential Improvements:

- Collect longer data sequences for better transition learning.
- Add additional sensors (e.g., magnetometer).
- Use more advanced HMM variants (e.g., Hierarchical HMM or Gaussian Mixture HMM).
- Apply dimensionality reduction (PCA) for optimized feature representation.

6. Conclusion

This project successfully demonstrated the use of Hidden Markov Models (HMMs) for recognizing human activity states based on motion sensor data from smartphones. By combining accelerometer and gyroscope readings with carefully extracted time- and frequency-domain features, the model effectively inferred hidden activity states such as *walking, standing, jumping,* and *still*.

The trained HMM showed strong performance, achieving high accuracy and stability when tested on unseen data. The results confirmed that even with limited samples, probabilistic models like HMMs can reliably capture transitions and patterns in human motion.

Overall, this work highlights the potential of lightweight, data-driven models for real-world applications such as health monitoring, smart environments, and activity tracking systems. With further data collection, additional sensors, and longer recording durations, the model's performance and generalization could be enhanced even further.

7. Collaboration and GitHub Contribution

Both group members contributed equally:

Task	**Leslie Isaro**	**David Ubushakebwimana**
Data collection	☑	☑
Data Preprocessing	☑	
feature extraction	☑	
HMM modeling		☑
Model Testing & evaluation		☑
Report writing	☑	☑

Important links

Repo: https://github.com/l-isaro/Formative-2---Hidden-Markov-Models.git

Dataset:
https://drive.google.com/drive/folders/1O6Np3TWi0vFWvn6PaYK4c-9UUVt1lURi?usp=sharing