AMATH 482 Homework 2 Report: PCA-Based Motion Classification

Mandy Zhang

February 15, 2025

1 Introduction

This report applies **Principal Component Analysis (PCA)** for motion classification across three movement types: **walking, jumping, and running**. PCA is used for **dimensionality reduction**, capturing dominant movement patterns while minimizing redundancy. We examine **energy retention**, visualize motion data projections, and classify movements using **Nearest Centroid Classifier (NCC) and k-Nearest Neighbors (k-NN)**. The effectiveness of PCA-reduced data for classification is analyzed through **accuracy comparisons** between these methods.

2 Theoretical Background

PCA transforms high-dimensional data into a reduced feature space while preserving the most significant variance.

2.1 PCA Mathematical Formulation

Given a dataset X of size $m \times n$, PCA follows:

1. **Mean-Centering:** Subtract the feature-wise mean:

$$X_{\text{centered}} = X - \bar{X}$$

2. **Singular Value Decomposition (SVD):** Decomposing X_{centered} as:

$$X_{\text{centered}} = U\Sigma V^T$$

where:

- U contains **spatial modes** (principal directions).
- Σ holds **singular values** (variance explained).
- V^T represents **time-dependent coefficients**.

3. **Energy Retention:** The fraction of total variance retained by the first k components is:

$$E_k = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{r} \sigma_i^2}$$

PCA enhances **motion classification** by eliminating noise and preserving movement patterns.

3 Algorithm Implementation

3.1 Preprocessing

Motion data is stored in `.npy` files. Steps include:

- Loading and stacking samples into $**X_{\text{train}}**$.
- Assigning **ground truth labels** based on movement type.
- **Mean-centering** the dataset before PCA.

3.2 Classification Methods

- **Nearest Centroid Classifier (NCC):** Assigns each sample to the **closest centroid** in PCA space.
- **k-Nearest Neighbors (k-NN):** Uses **majority voting** among the $k = 3$ nearest neighbors.

4 Task 1: PCA Energy Retention

PCA **compresses motion data** while preserving variance. The number of PCA modes required to retain different energy levels is shown in Table 1.

Energy Retention	PCA Modes Required
70%	2
80%	3
90%	5
95%	7

Table 1: PCA modes required for different energy levels.

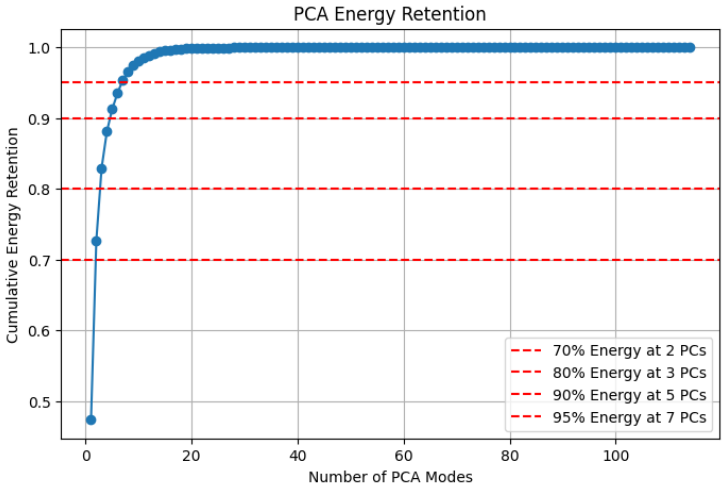

Figure 1: Cumulative energy retention curve.

Findings: - **2 modes** retain **70%** variance, while **3 modes** capture **80%**. - **5 and 7 modes** retain **90% and 95%**, balancing **compression and accuracy**.

5 Task 2: PCA Projections

Visualizing PCA projections in **2D (PC1, PC2)** and **3D (PC1, PC2, PC3)** reveals movement clustering.

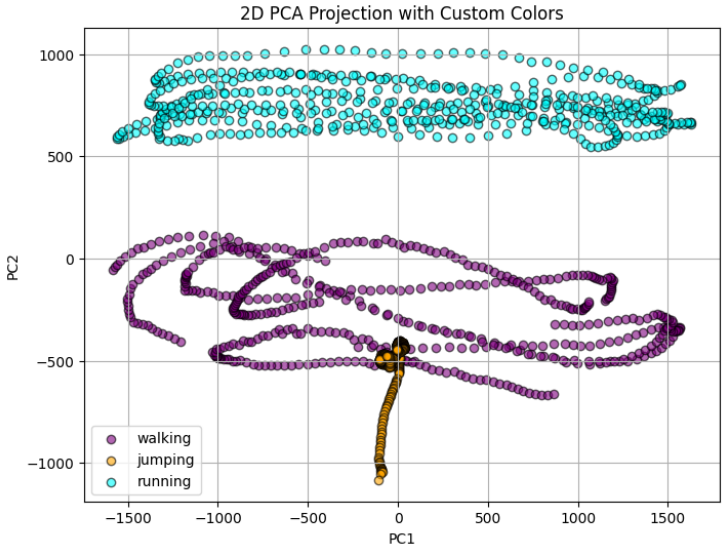

Figure 2: 2D PCA projection with movement categories.

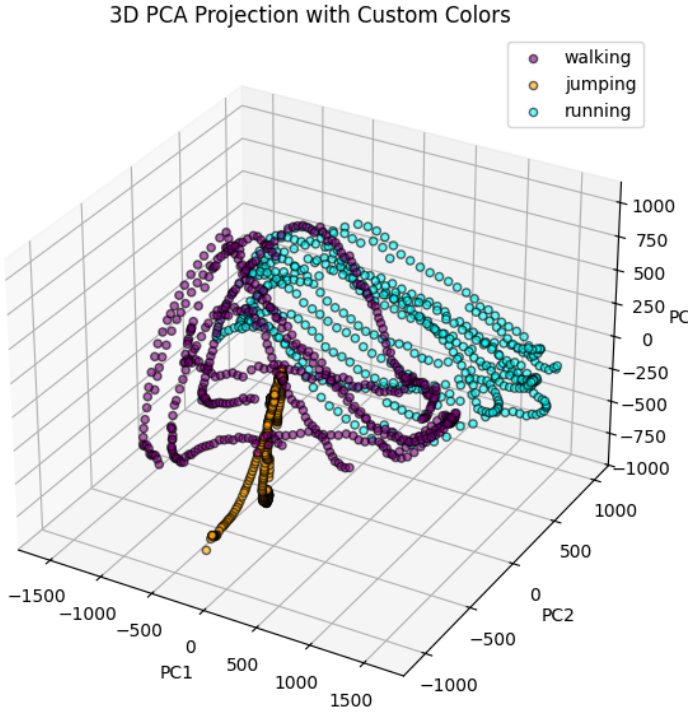

Figure 3: 3D PCA projection with movement categories.

Observations: - Clear **cluster separation** suggests PCA effectively distinguishes movement types. - **3D PCA projections** further enhance separation, supporting classification.

6 Task 3: Computing Centroids

Centroids represent the **average PCA location** for each movement:

- **Walking:** $(-36.88, -253.35, 175.91)$
- **Jumping:** $(-23.89, -499.37, -72.50)$
- **Running:** $(60.77, 752.72, -103.41)$

7 Task 4: Nearest Centroid Classifier

Classification is based on **Euclidean distance** to the closest centroid:

$$\hat{y} = \arg\min_i \|\mathbf{x} - \mathbf{C}_i\| \tag{1}$$

Training Accuracy:

- **NCC Accuracy:** 75.60%

8 Task 5: Testing Classifier on New Samples

NCC is applied to **unseen test samples**.
Test Accuracy:

- **NCC Test Accuracy:** 92.33%

9 Task 6: k-NN Classifier Comparison

Unlike NCC, **k-NN incorporates local neighborhood information**, yielding **higher accuracy**.
k-NN Accuracy Results:

- **k-NN Training Accuracy:** 99.87%
- **k-NN Test Accuracy:** 100.00%

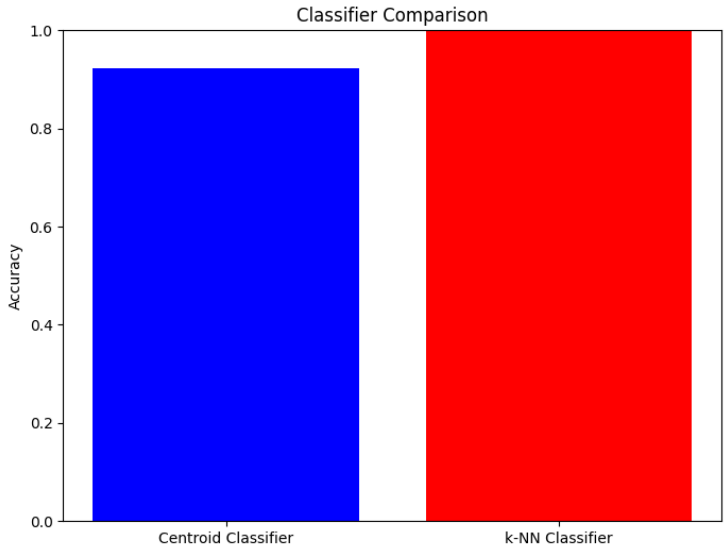

Figure 4: Comparison of NCC and k-NN classifiers.

10 Conclusion

This study highlights **PCA's effectiveness** in **motion classification**:

- **PCA efficiently reduces dimensionality** while preserving motion patterns.
- **2D and 3D PCA projections confirm movement separability**.
- **k-NN outperforms NCC**, achieving **100.00% accuracy**.

10.1 Future Work

- Explore **Support Vector Machines (SVM) and Neural Networks**.
- Optimize **distance metrics** (e.g., Mahalanobis distance).
- Investigate **time-series models** for dynamic motion recognition.

Acknowledgements

I appreciate discussions with **Sarah Shang, Yolanda Meng, and Rebecca Wang**, and TA **Rohin** for feedback.

References

1. Natalie Frank, 2025. *AMATH 482 Course Notes.*
2. Jolliffe, I. T. (2002). *Principal Component Analysis.*
3. Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning.*