# Unsupervised Learning Hand Postures Clustering

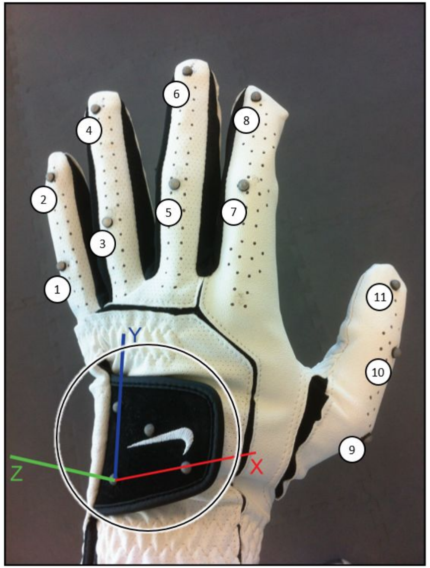
Michael Tillapaugh

# Intro

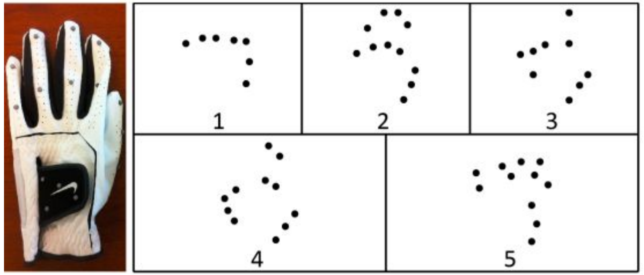
A Vicon motion capture camera system was used to record 12 users performing 5 hand postures with markers attached to a left-handed glove.

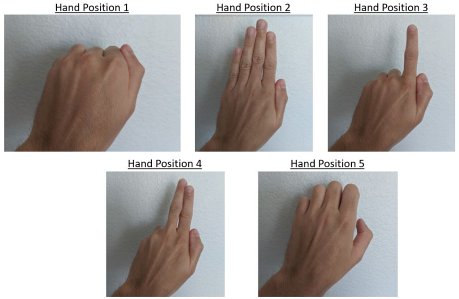
The glove used to capture data

1. Fist (with thumb out)
2. Stop (hand flat)
3. Point (with 1 finger)
4. Point (with 2 fingers)
5. Grab (fingers curled as if to grab)



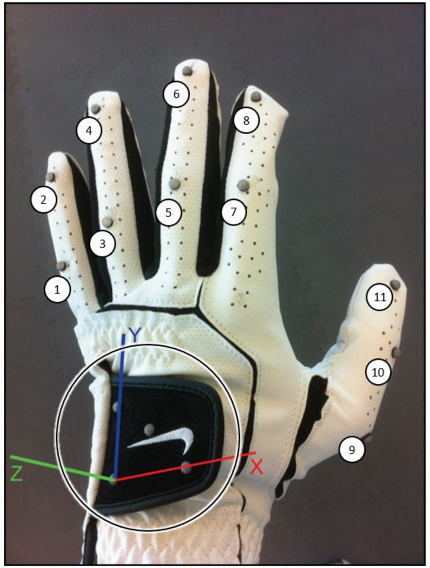

<u>Goal</u>: To classify hand postures together based on coordinate measurements taken



# Dataset

- 78,095 recordings
- 5 Classes (one for each hand posture)
- 12 Users (each person using the glove)
- 12 Coordinate Markers (X, Y, and Z for each marker)

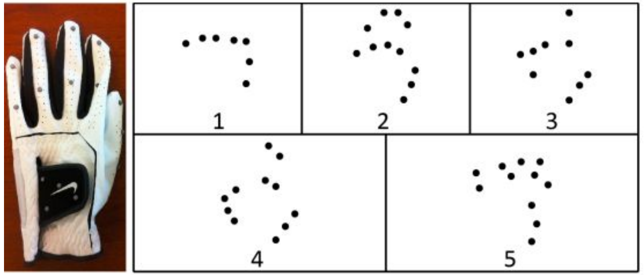



Goal: To classify hand postures together based on coordinate measurements taken

| | Class | User | X0 | Y0 | Z0 | X1 | Y1 | Z1 | X2 | Y2 | Z2 | X3 | Y3 | Z3 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0 | 0 | 0 |
| 1 | 1 | 0 | 54.263880 | 71.466776 | -64.807709 | 76.895635 | 42.462500 | -72.780545 | 36.621229 | 81.680557 | -52.919272 | 85.2322638852917 | 67.7492195028673 | -73.684130041833 |
| 2 | 1 | 0 | 56.527558 | 72.266609 | -61.935252 | 39.135978 | 82.538530 | -49.596509 | 79.223743 | 43.254091 | -69.982489 | 87.4508729469625 | 68.4008083028339 | -70.703990925959 |
| 3 | 1 | 0 | 55.849928 | 72.469064 | -62.562788 | 37.988804 | 82.631347 | -50.606259 | 78.451526 | 43.567403 | -70.658489 | 86.8353875680762 | 68.9079249764243 | -71.1383441365739 |

# EDA

- Only 690 of 78,095 entries had values for every column, therefore simply dropping null values would remove too much data.
- Two approaches:
  - Remove specific columns and keep null values
    - Dropping columns for markers 9-12, 8-12, 7-12 ... 4-12
  - Remove specific markers based on hand posture shapes
    - Only dropping markers on the thumb
    - Only dropping fingertip markers
- This creates 10 dataframes that we could attempt to cluster



# Mini-Batch K-Means

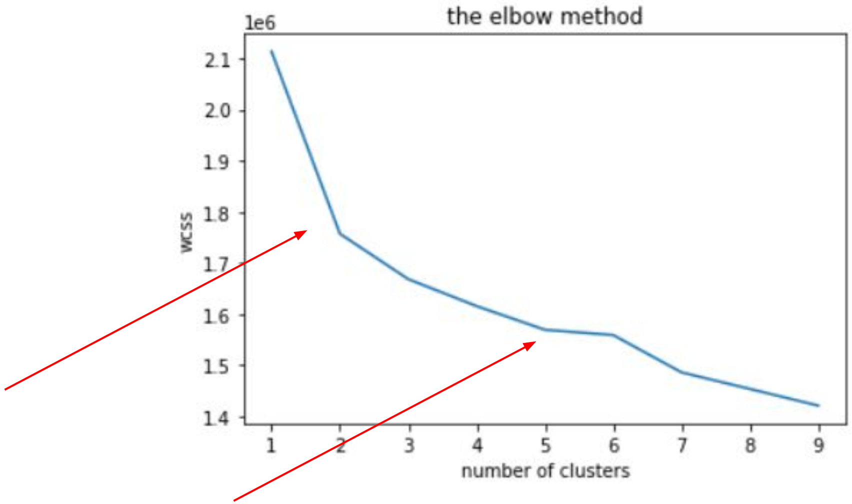
A mini-batch k-means was performed on each dataframe. The average ARI and Silhouette score (of 3 runs) was taken for a mini-batch K-means using 2 through 9 clusters.

The best performing dataframe was chosen to pursue more in depth. In this case the dataframe where we removed the thumb markers had the best ARI and Silhouette scores at 5 clusters. This is good because we already know there are 5 known clusters we should expect.

```
1 loop, best of 3: 679 ms per loop
1 loop, best of 3: 593 ms per loop
1 loop, best of 3: 593 ms per loop
1 loop, best of 3: 569 ms per loop
1 loop, best of 3: 883 ms per loop
1 loop, best of 3: 585 ms per loop
1 loop, best of 3: 584 ms per loop
1 loop, best of 3: 1.04 s per loop
                ARI Scores  Silhouette Scores
Cluster Number
2                 0.268724           0.167603
3                 0.326467           0.099651
4                 0.264176           0.072089
5                 0.313081           0.082799
6                 0.308886           0.062979
7                 0.277690           0.073847
8                 0.249641           0.087942
9                 0.229363           0.071599
ARI and Silhouette Scores for df_nothumb
```

# The Elbow Method

To choose the appropriate number of clusters we can utilize the 'Elbow Method'.  Here we plot WCSS (Within-Cluster-Sum-of-Squares) and the point at which we start to see diminishing returns (the elbow) is the number of clusters to use.



# Choosing a Clustering Method

Now that we know to use 5 clusters we compare the results of the following clustering methods:

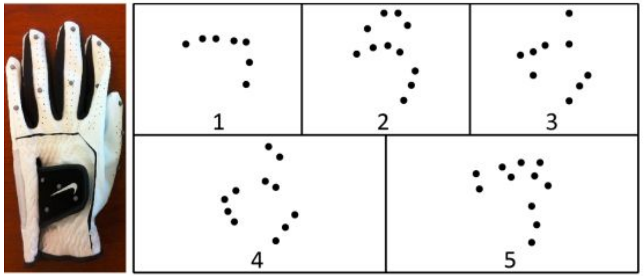
- K-Means, Mini-Batch K-Means, GMM, DBSCAN, Hierarchical Clustering (Agglomerative)

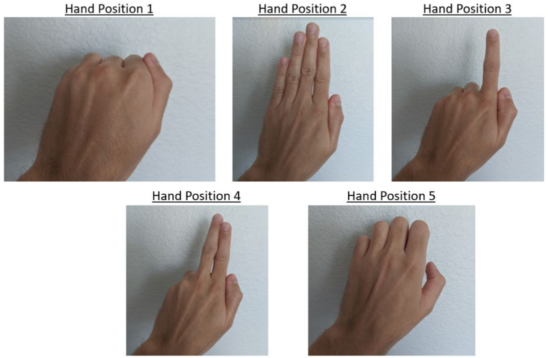
Of the above methods the following proved to be the best performing with an ARI score of ~0.45

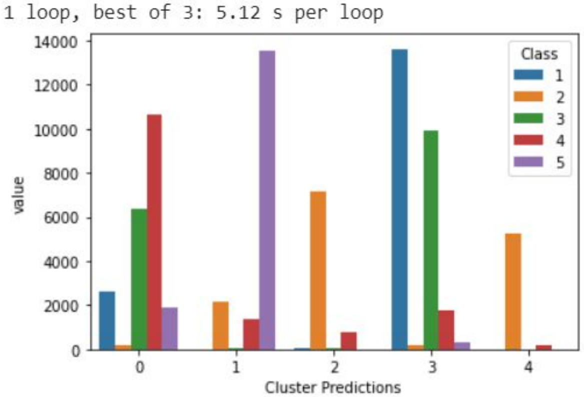
- GMM, n_components=5, covariance_type='full', max_iter=100

# Cluster Predictions vs Actual Labels

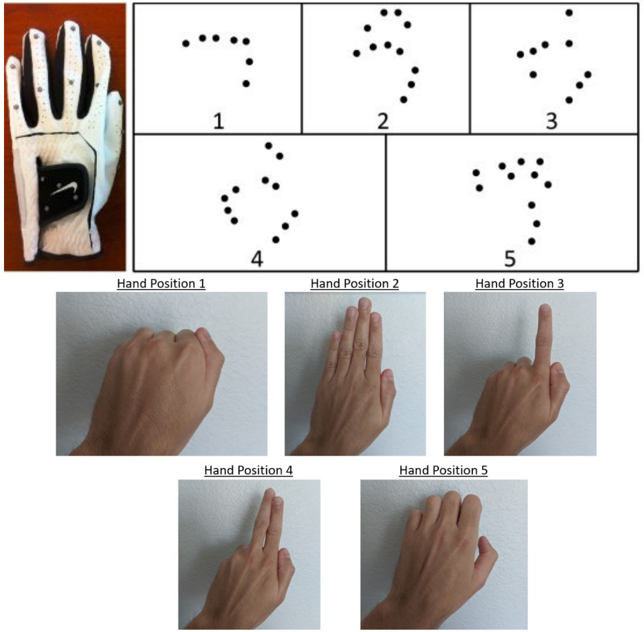
Next a comparison was done between cluster predictions and actual class labels for each data point

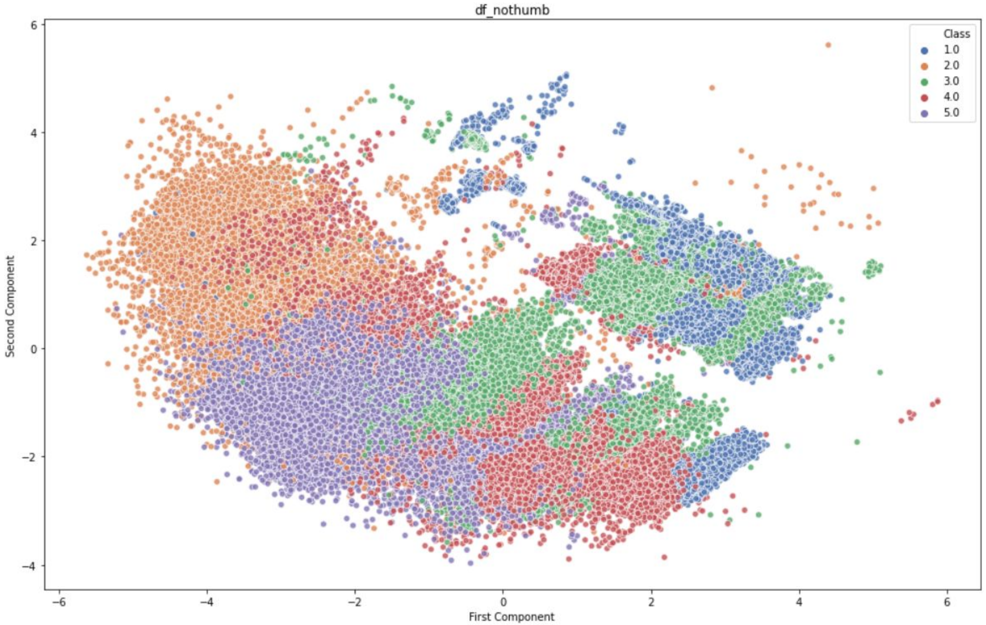


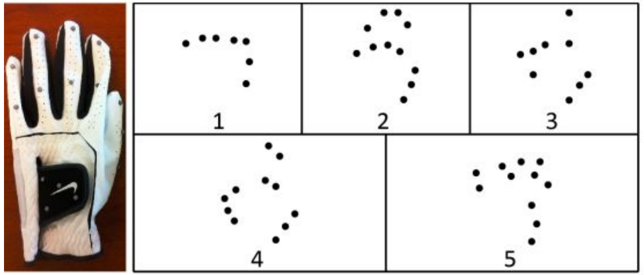



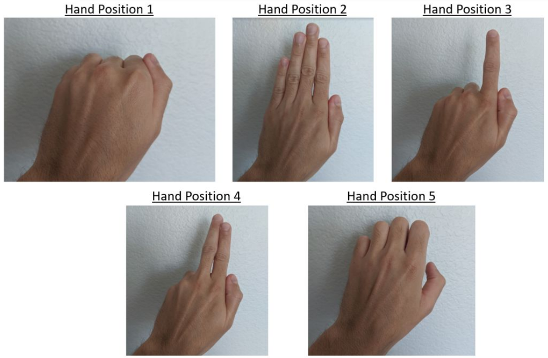
# PCA Visualization

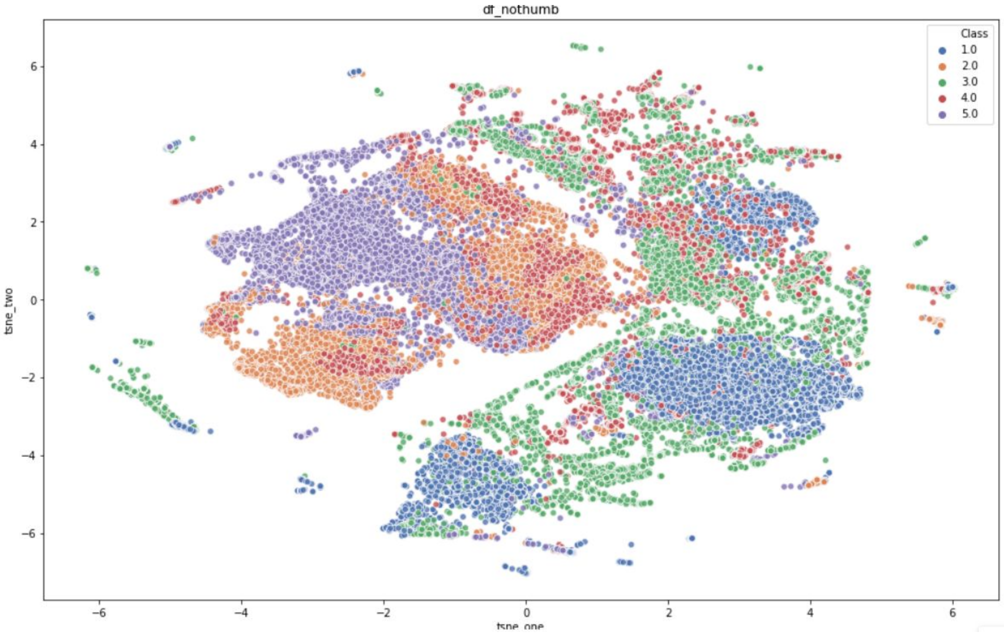


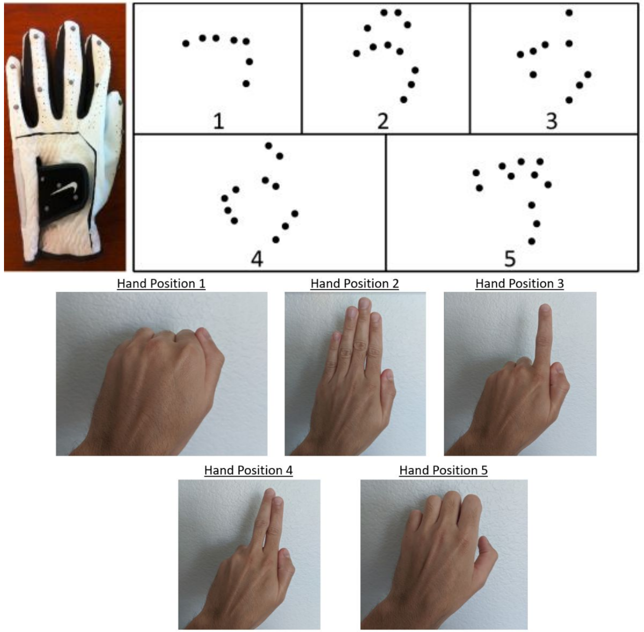
# t-SNE Visualization



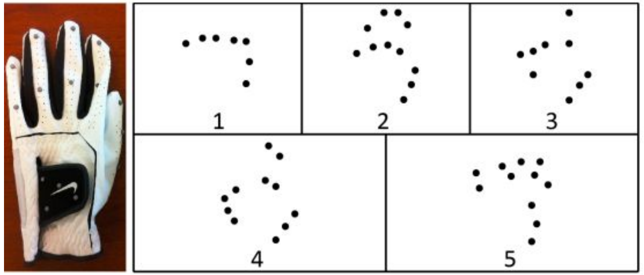
# UMAP Visualization

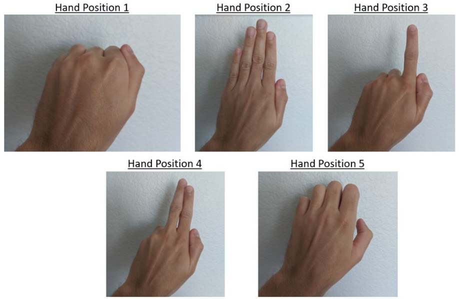



# Conclusions

What we see from the dimensionality efforts is that although there is a lot of data points and noise, the models did a good job of identifying regions where data can be clustered together.

- Hand postures 2 (stop) and 5 (grab) were the most distinct in the visualizations and are more distinguished from the other postures.
- Hand postures 3 (point 1-finger) and 4 (point 2-finger) had the most noise / overlap. They are also the most similar hand postures to each other.



# Thank you!