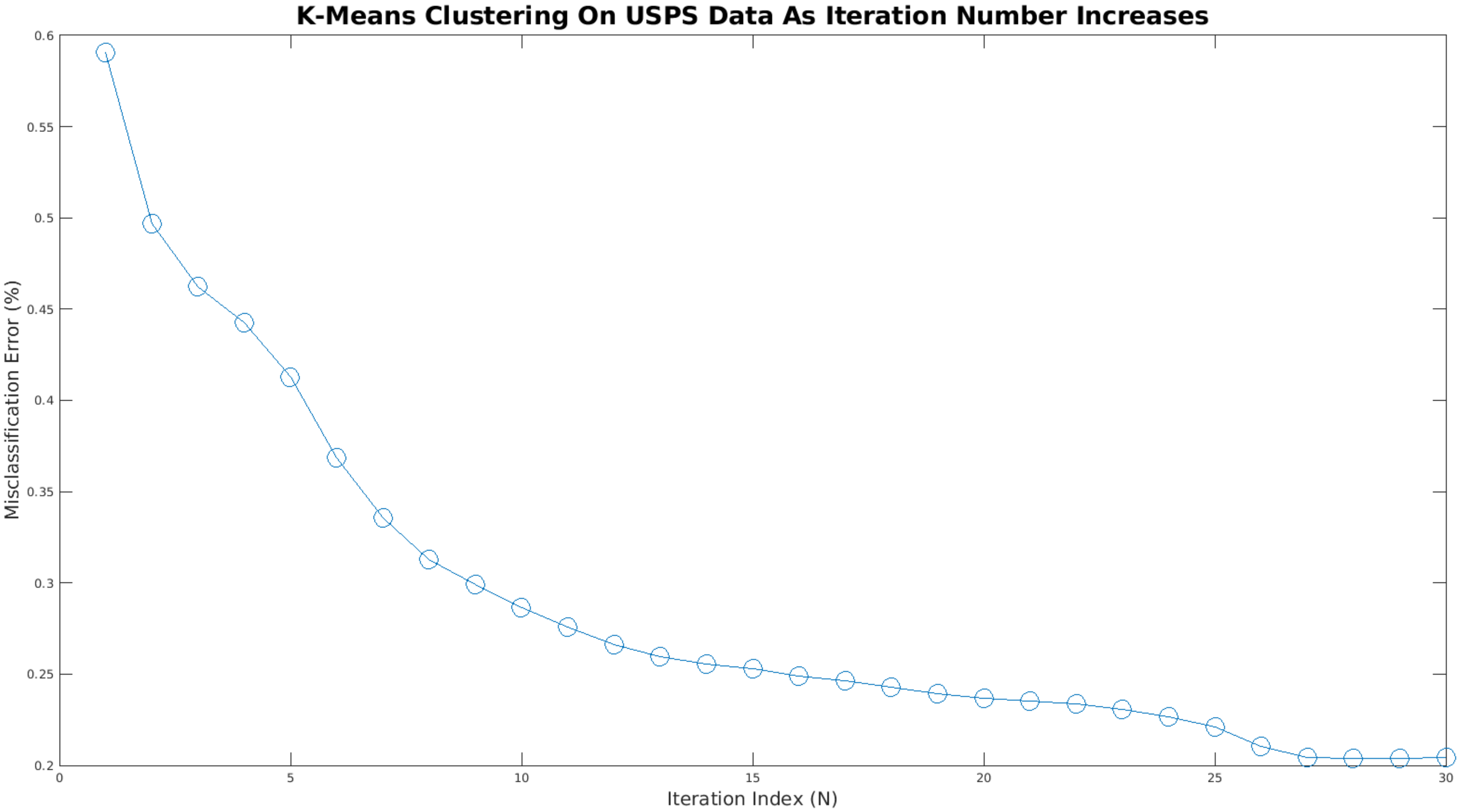
PCA + K-Means Clustering - HW#8

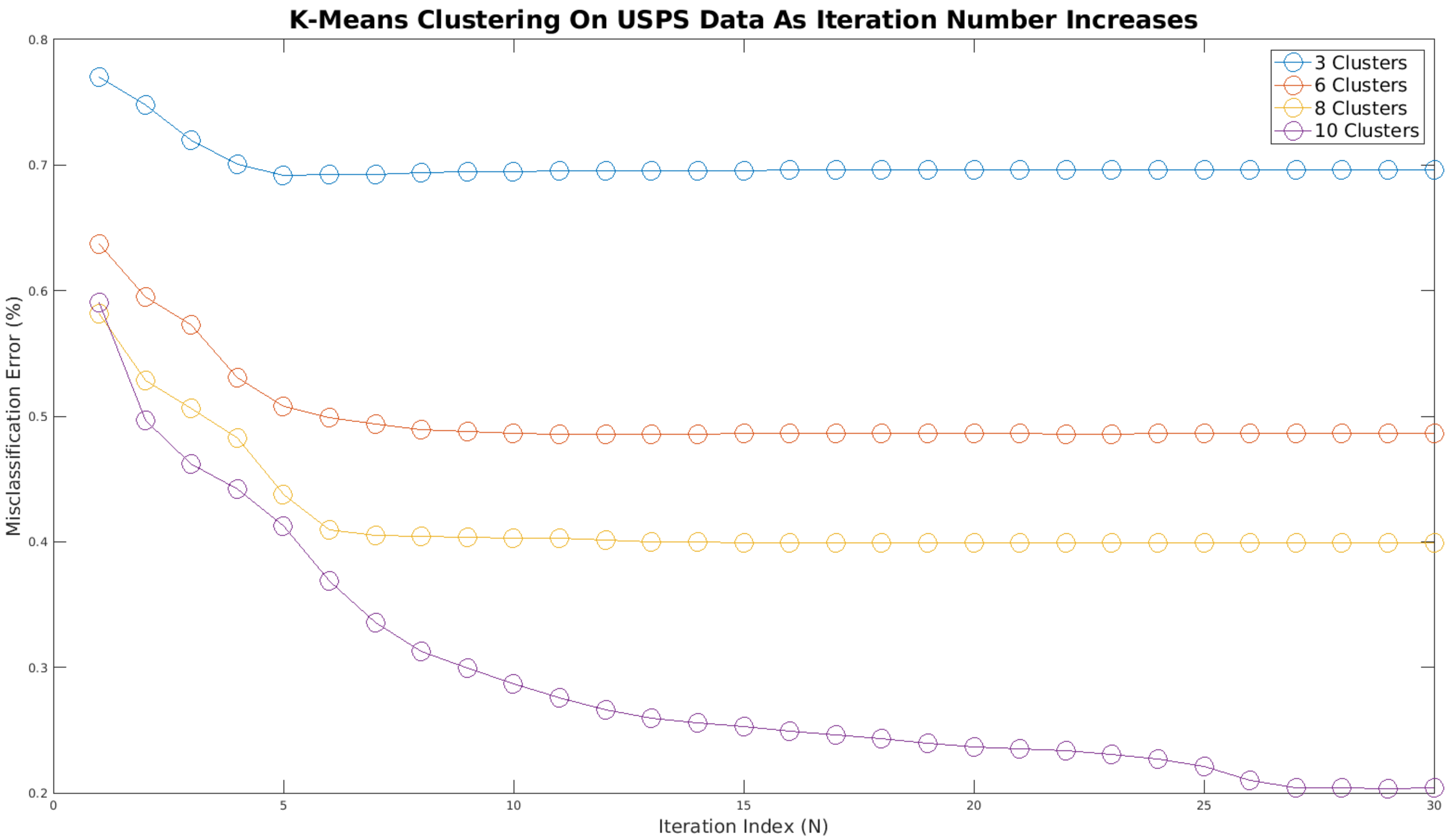
MACHINE LEARNING EE5354

Javier Salazar
1001144647

Clustering (Clusters = 10)



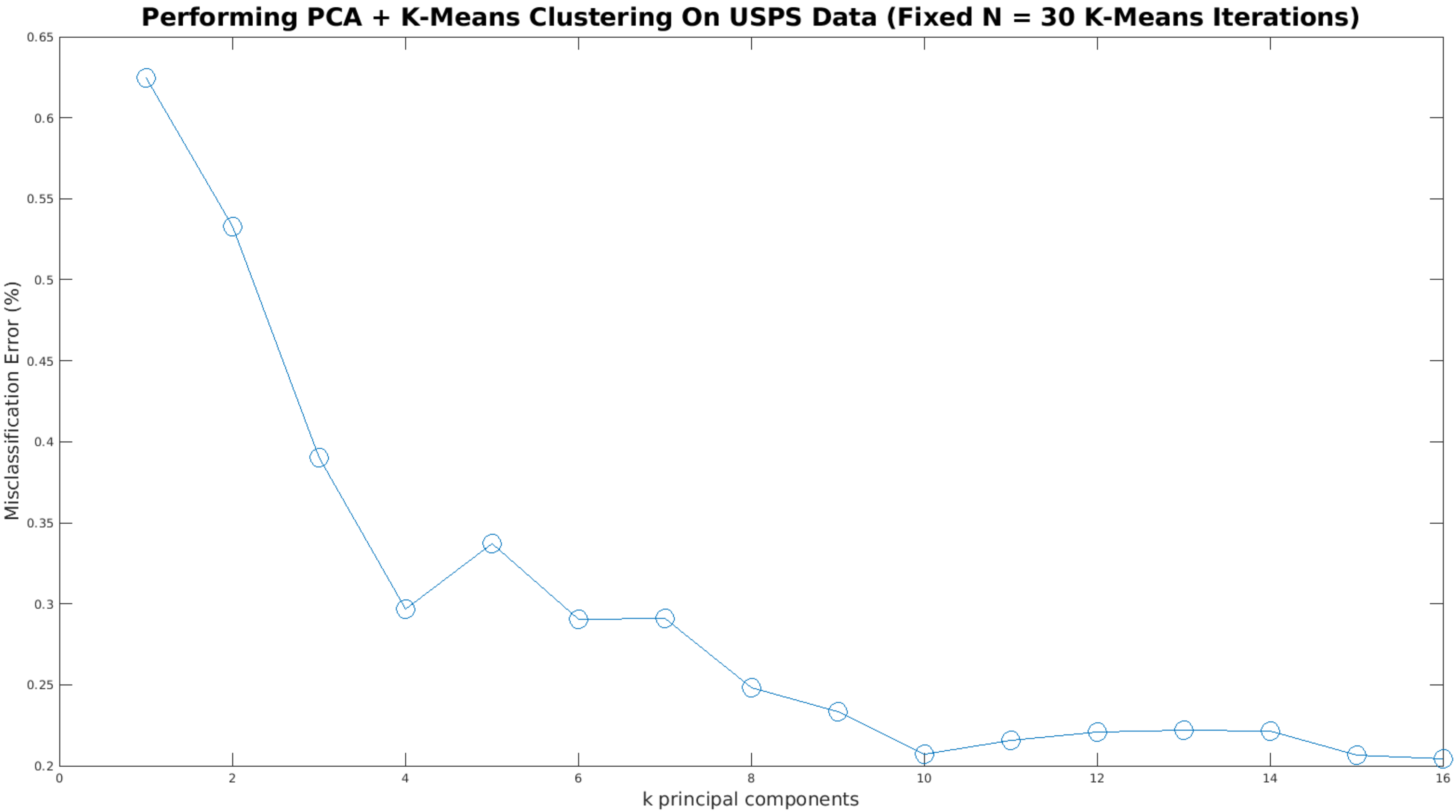
Clustering (Multi Cluster)



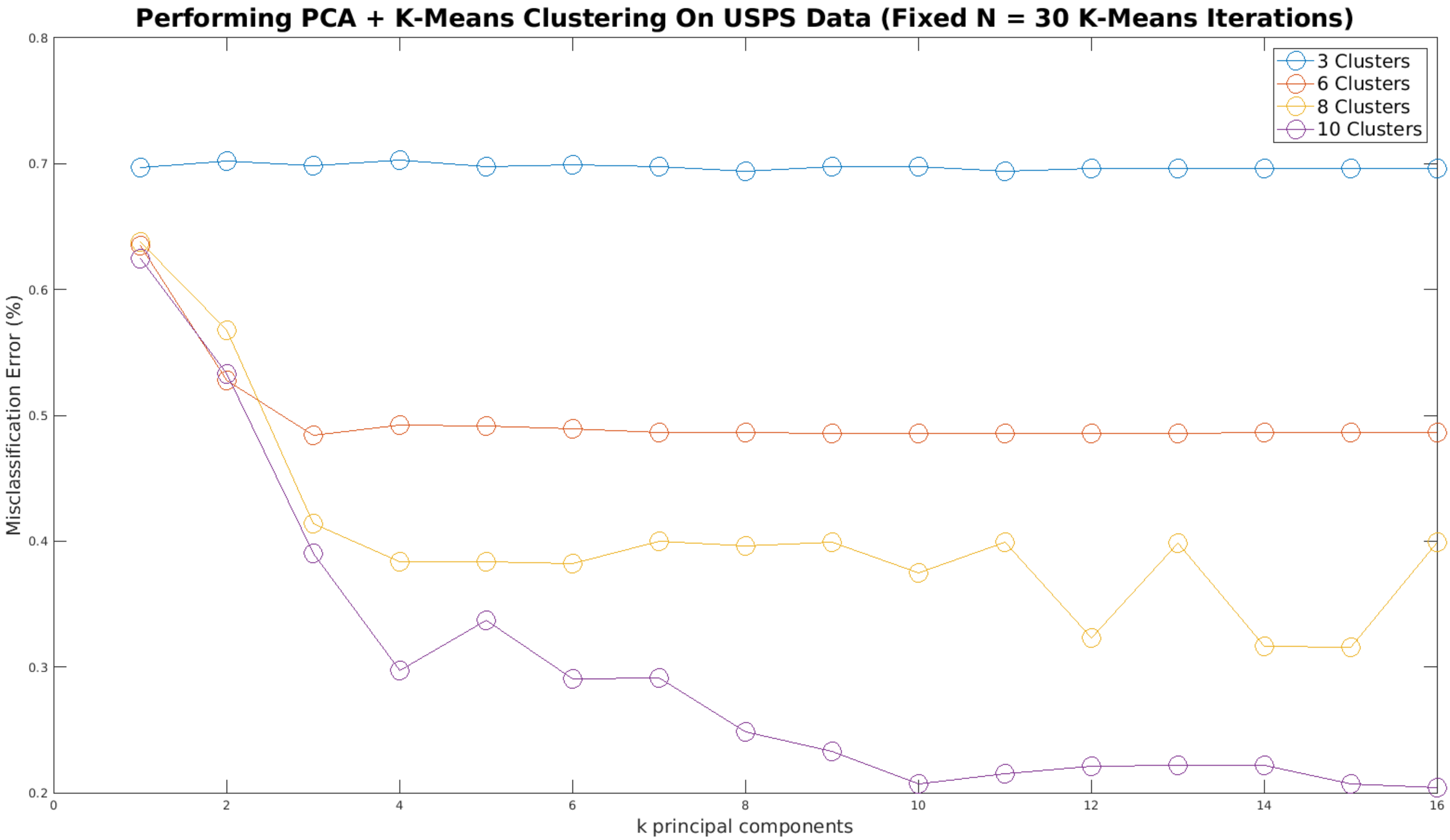
Clustering Analysis

- Only one iteration → ~60% error (Clusters = 10)
- 30 iterations → ~20% error (Clusters = 10)
- Randomly assigning labels would result in 90% misclassification error so results are pretty good after 30 iterations for 10 clusters
- Results seem to stagnate after 30 iterations
- L2 norm utilized for distance metric

PCA + Clustering (Clusters = 10)



PCA + Clustering (Multi-Class)



PCA + Clustering Analysis

- Clustering iterations fixed to N = 30 iterations since that lead to good results in clustering problem
- 1 comp. → ~63% misclassification error (Clusters = 10)
- 10 comp. → ~20% misclassification error (Clusters = 10)
- 16 comp. → ~20% misclassification error (Clusters = 10)
- Looks like we can keep the first 10 component vectors without really losing any important information in this example
- L2 norm used for distance metric during clustering

File Information

- **The file ‘cluster.m’ is to generate the plot of how the misclassification error changes as we increase K-Means iterations**
- **The file ‘script_clusters.m’ is to generate the PCA plot of misclassification error changes as we increase number of component vectors**
- **Both files have fixed random seed (e.g. rng(7777)) to directly observe effects of increasing iterations or components on misclassification error otherwise results could be misleading**