CS7641 – Machine Learning

Assignment 3 – Unsupervised Learning and Dimensionality Reduction

Mason J. Kelchner  
mason.kelchner@gatech.edu

# 

# INTRODUCTION

In this analysis, several techniques used in unsupervised learning are explored. Namely clustering and dimensionality reduction are used to determine likeness among data set samples to create more meaningful insight into the relationship between data features ultimately producing better prediction outcomes. Several algorithms are implemented and applied to two data sets and their performance, likeness, advantages, and disadvantages are discussed on both data sets. Two clustering algorithms are implement: Expectation Maximization and K-Means. Four dimensionality reduction algorithms are implement, three of which are linear models including Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Randomized Projections (RP). The fourth dimensionality reduction algorithm is a nonlinear manifold learning algorithm which was selected as t-Distributed Stochastic Neighbor Embedding (t-SNE). Several experiments were run for each algorithm with the goal of disseminating how the clustering and dimensionality reduction techniques affect each data set.

# Clustering

## Expectation Maximization

## K-Means

# Dimensionality Reduction

## Principal Component Analysis

## Independent Component

## Randomized Projections

## t-Distributed Stochastic Neighbor Embedding

# Clustering Applied to Dimensionality Reduction

# Neural Network using Dimensionality Reduction

# Neural Network using Clustering

# Summary

# References

|  |  |
| --- | --- |
| [1] | F. Pedregosa, Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825-2830, 2011. |