

Brief Overview:

This program uses Principle Component Analysis (PCA) analysis and k-nearest neighbors' algorithm (KNN). PCA is used to project the data, in this case 3823x64, into any 3823xY matrix. With PCA the data can be represented by Y principle components that capture the maximum variance. An ordered eigen vector (Yx64) matrix of the covariance matrix of the training set is constructed such that the corresponding eigen values are $\lambda_1 \geq \dots \geq \lambda_3$. This projection matrix allows us to substantially reduce the size of the data without losing much information about the data. A KNN algorithm is used to classify the data based on based the training set with various values of K. The accuracy of the KNN algorithm is tested against various sizes of Y and K. As expected, As Y increases the PCA becomes more accurate at a decreasing rate until it approaches 64. At this point no dimensionality has been reduced and the projection matrix is a scalar of the covariance matrix. As shown in the video demonstration the projection matrix of 64 has the same information capture as the original data set. In a few short words, as Y decreases the accuracy of the learned KNN algorithm slightly decreases but the memory usage and speed of the algorithm substantially increases.