Problem 3 of Take Home Exam #2

Using the dataset of Problem #1, explore the clustering algorithms hard K-Means and fuzzy c-means for classification of the data into 10 digits. To reduce the computational cost, apply the principal component analysis algorithm to reduce the dimension from 256 to 22, achieving over 90% accuracy of the approximation.

Assume first that you do not know the labels for the training set. Join together the feature training and testing sets, x = [ x1 x2 ]; x is a 256x9298 array. Apply principal component analysis and retain the strongest 22 features, call it xr, array of 22x9298.

Apply the clustering algorithms for K=10, i.e. to obtain 10 clusters. Verify the accuracy of the clustering for the digits in the training set.

Based on the clustering results, label the digits of the testing set, save the result in array y2.