

DS-630 HW5

Assigned April 10, 2018

Due April 17, 2018

In this assignment you will perform a principal components analysis of the dataset *mnist.csv* using the `princomp` or `prcomp` function in R. Note that the data set consists of 60,000 rows, each of which contains the grey scale pixel representation of a given hand written digit which appears in the first column of the row. The remaining $784 = 28 \times 28$ columns of the row contain the grey scale values for each pixel. Recall the grey scale values range from 0 - 255, with 0 encoding white and 255 encoding black.

1. Employ either the `princomp` or `prcomp` function to perform PCA on the 784 columns of the data set which contain the *grey scale pixel* information. Do not include the first column as this column merely indicates the digit 0-9 that a given row encodes.
2. Report the number of principal components needed to account for 98% of the variance of the original data set.
3. For each of the principal components reported in 2. supply the percentage of the variance that each component contributes to the overall variance.
4. Are there pixels in the original data set that consistently appear with significant loadings in the principal components reported in 2.?

Bonus: Graph the first 10 principal components in a 28×28 pixel grid.