Principal Component Analysis (PCA) Examples

Statistics 185

9/8/2019

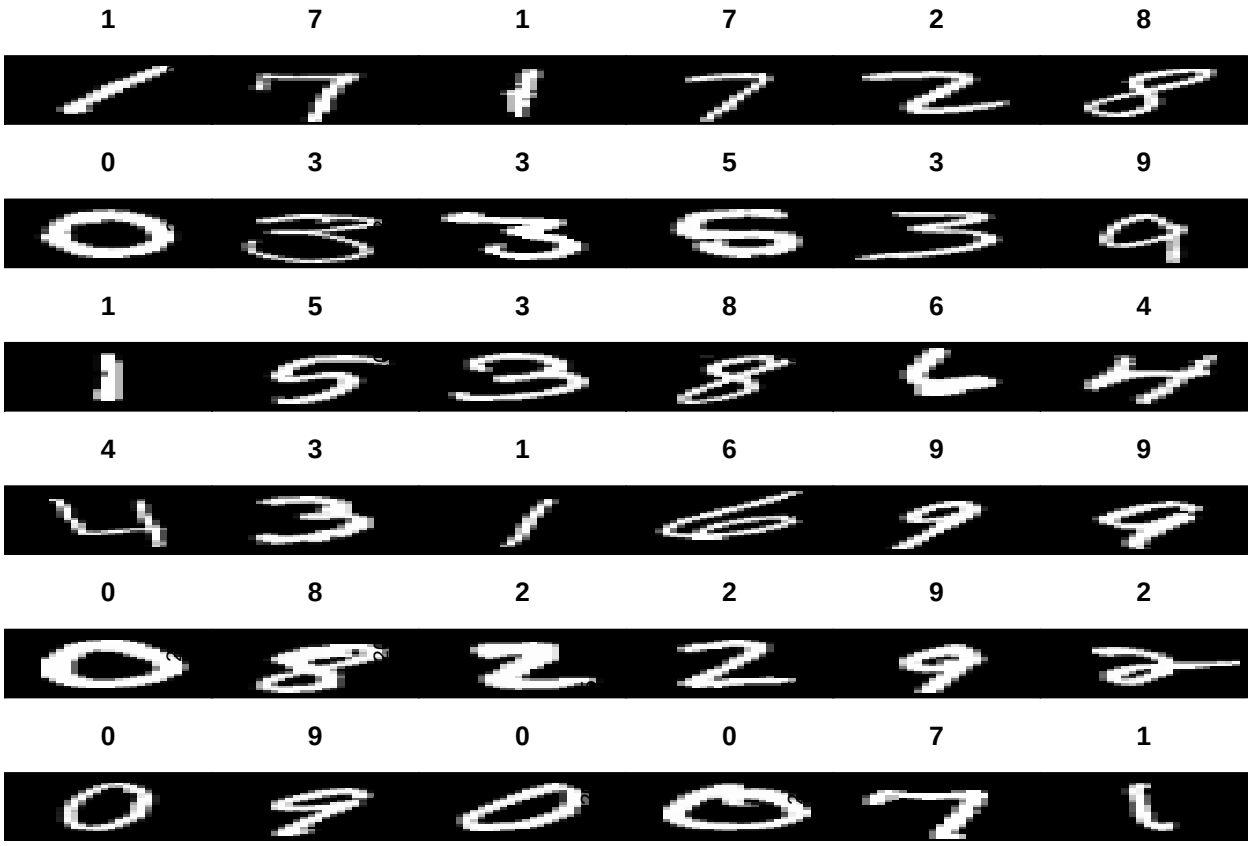
MNIST Dataset

MNIST a database of images of handwritten digits. Each image in MNIST is a 28 x 28 pixel images of a single handwritten digit. For each pixel, we have a number between 0 (black) and 255 (white) indicating the color of the pixel. Thus, each image can be thought of as a 28 by 28 matrix. We convert this matrix into a $28^2 = 784$ dimensional vector by stacking its columns, $\vec{a}_1, \ldots, \vec{a}_{28}$, each a vector in $\mathbb{R}^{28}$.

$$[\vec{a}_1, \ldots, \vec{a}_{28}] \longmapsto \begin{bmatrix} \vec{a}_1 \\ \vdots \\ \vec{a}_{28} \end{bmatrix}$$

This file contains a list *digits* with 42,000 digits.

```
# visualize the digits
par(mfcol=c(6,6))
par(mar=c(0, 0, 3, 0), xaxs='i', yaxs='i')
for (idx in 1:36) {
    im <- matrix(digits$pixels[idx, ], ncol = 28, nrow = 28)
    image(1:28, 1:28, im, col=gray((0:255)/255),
          xaxt='n', main=paste(digits$labels[idx]))
}
```



Flow Cytometry Data