

PCA on MNIST Digit Data

In this practical, we experiment with PCA on the MNIST Digit Dataset

The MNIST dataset consists of 1024-dimensional inputs corresponding to pixels of a 32×32 image showing a handwritten digit from 0 to 9. For this practical, we use only 5000 images, 500 for each digit; the first 500 are zeros, the next five hundred ones, and so on.

The file `p6_process_data.lua` performs normalisation and singular value decomposition. It stores this decomposition so that you don't have to repeatedly run the time-consuming svd code during your remaining experiments. After running `p6_process_data.lua` once, you can modify and re-run `practical6.lua` as you wish.

The files for this practical can be found here: <https://github.com/oxford-cs-ml-2016/practical6>

Read `README.md` for setup instructions for the lab machine.

Task 1: Visualization

Plot the data on the first 2 and 3 principal components using gnuplot. Notice that images corresponding to the same digits are clustered quite well together even with just three PCs.

Try plotting different pairs of digits and investigate which ones are not well separated using the first two or three PCs. Look at the PCs below in part 2 and see if your observations make sense.

If you run `dofile('practical6.lua')` you will be able to rotate the 3d image made by gnuplot.

Task 2: Principal Components

The principal components themselves can be visualised as 32x32 images.



These are the first, second, third, twenty-fifth, hundredth, five hundredth, and thousandth principal component respectively. The images are created using a scaled version of the PCs to avoid excessively dark images. Notice that the last one only has a single white pixel in the bottom row. Can you explain why this might be happening?

Task 3: Reconstruction (code not provided)

Try reconstructing the images, one for each digit using projections on to the first few principal components. Notice how the number of components required for a reasonable reconstruction changes with the digit.

For this part, we have not provided code. Fill in the part that says TODO in `practical6.lua`.



Machine Learning
Hilary Term 2016
Week 8

Handin

There is no handin for this practical