Isabelle Pardew

24 July 2017

Notes on PCA / Coordinate Indices

Important factors

**Performing PCA:** If we are given **p** number of photos, import them as matrices into MATLAB with dimensions **m** X **n**, we make each matrix one large column vector. (by successively placing each column under the first)

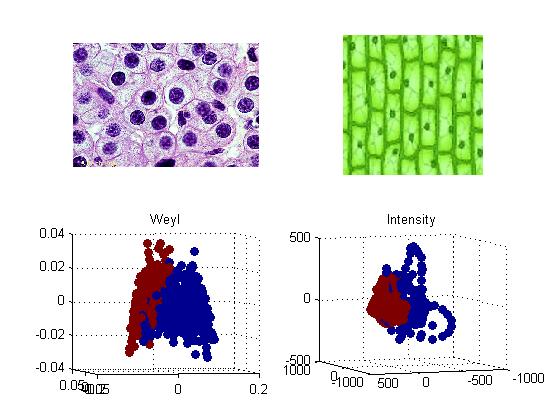
Essentially ending up with a **matrix of observations** of dimension **(m x n)** X **p**.

This matrix is centered around the mean of the data, then we find the **covariance matrix**, which will account for the variance in the data.

The eigenvectors (unit) of the covariance matrix are bases for the new coordinate indices.

**Principal component 1** – largest eigenvector

**Principal component 2** – next largest eigenvector

**PCA and Image Detection:** PCA is a useful tool for recognizing images. If there are thousands of pictures of faces run through the PCA process, there should be about the same variance across all the images. If we have our new basis after performing PCA with the “training set”, or the “base set”, then have replotted our data on the new 2-dimensional axes (with principal component 1 and principal component 2 on the x and y axes, respectively), the cluster of points should be relatively uniform.

Now, if we input a bunch of images that are *not* faces, and replot them according to the new basis/axes we have, they should be separated from the cluster of points with the faces, essentially producing a visual presentation of the image detection. (see figure 5 on the poster – we can see what *is* a plant cell, and what *is not* a plant cell, or rather a human cell).

**Coordinate Indices (Weyl, etc…):** The Weyl representation builds on PCA, in that it is representing data differently through the use of Kronecker products (no need to get into). A sequence of binary numbers can easily be shortened into a smaller, concise product of binary base vectors.