# Part II: Feature Extraction

Say we have a set of images of human faces denoted as matrices . Our goal is to use this data to derive some method for identifying moustaches within images not found in the original dataset. The matrices are first vectorized, i.e., their columns are stacked on top of one another to form vectors:

where . The resultant vectors are then stacked column-wise to form a matrix containing all the images:

We will identify the “average face” by taking the column-wise mean of :

This can be used to mean centre :

where is a column vector of ones. We will now consider the reduced singular value decomposition of :

where , and . This can be rewritten as a sum of rank one matrices:

where and and are the columns of and respectively. Since we have that . As such, we can see that the relative contribution of each matrix towards the reconstruction of is determined solely by the value of . Since , it can be deduced that said contribution monotonically decreases as increases. It turns out that in most instances this happens very quickly. As such, can often be very well approximated by

where . Returning to matrix format this can be expressed as

where , and . From this we can write an equation for the column of :

where is the row of . Let . Thus:

One can clearly see that is simply the coordinate vector of with respect to the basis . Note that is simply a projection of onto the lower rank basis . Now let’s say we have a new vectorized image and we want to find its closest representation with respect to the basis . To do this one would simply find the projection of onto with respect to . Since is orthonormal, said projection is given by:

This can then be expressed in terms of the standard basis as follow: