

Machine Learning Exercise 5

Shun-Lung Chang, Anna Ruby

With 50% variance (5 PC's):



The above images show the image vectors of the first 5 images in our original training sample, reconstructed by our PCA algorithm such that 50% of the variance information of the original image vector was preserved. We found that, in order to preserve 50% of the variance information, our algorithm needed 5 PC's. As we have lost half of the original variance information, our images appear notably more blurry than the original images, and have taken on more true 0 shapes.

With 80% variance (24 PC's):



To increase the amount of variance information preserved, the number of PC's needed for compression/reconstruction has increased. We now have preserved 80% of the variance information, using 24 PC's. Here we notice that the images have become somewhat sharper, but have begun to lose the appearance of a true 0.

With 95% variance (68 PC's):



The images above show the reconstructed image vectors, having maintain 95% of the variance information. In order to preserve such a high level of the original variance in our image vectors our algorithm needed 68 PC's. It is worth noting that we have preserved almost the entirety of the original variance information, and we have reduced the necessary dimensions notably. Again we notice an increase in the sharpness of our images, but a less obvious 0 shape.

With 99% variance (118 PC's):



The images above show our reconstructed image vectors, having preserved 99% of the original variance using 118 PC's. Again we see an increase in the sharpness of our images, but, as in the original training set, we have lost the true 0 shape in the imageness. Again, we have greatly decreased the necessary dimensions for representing our data set, while maintain a high level of the original variance information.

With 100% variance (240 PC's):



Here we have the image vectors, reconstructed with 100% of the original variance information. In order to maintain all of the variance information in the reconstruction, we need 240 PC's: one PC for every image vector. Only by fully spanning the set can we perfectly reconstruct the original data set.