

Homework 2: Applied Machine Learning

Tim Delisle and Sam Raudabaugh

09/29/2015

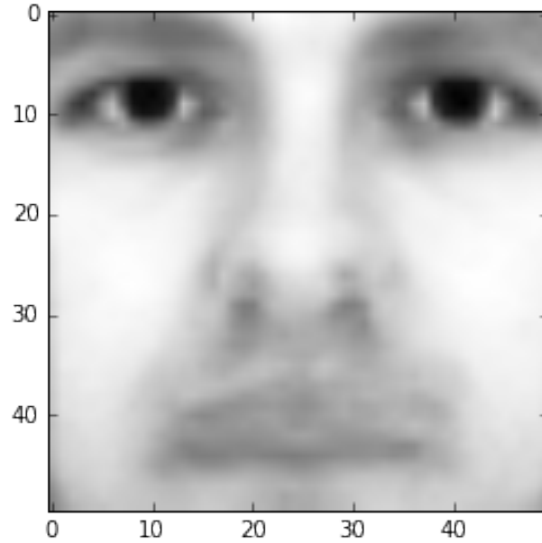


Figure 1: Mean face image

1. Face Recognition

For this assignment, we investigated a subset of the Yale Face Database. Our goal was to train a logistic regression classifier that uses eigenfaces to predict which subject is portrayed in the image.

Once again, we use the `LogisticRegression` model contained in `scikit-learn` for this classification. To train the classifier, however, we must first obtain the mean face μ from computing the average grayscale intensity of each pixel across the training set. The mean face is shown in Figure 1.

We then subtract this from every image and perform Singular Value Decomposition (SVD) on the dataset to compute the eigenfaces, an example of which is shown in Figure 2. Using the SVD factors U , Σ , and V^T , where V^T is the matrix of eigenfaces, we can construct low-rank approximations by truncating the number of elements in each factor.

In Figure 3, a plot of approximation error versus r , the number of eigenfaces (first r rows of V^T) used in the approximation, is presented. To better understand how our classifier will perform, we examine the relationship between information loss and the size of the subset of eigenfaces, and see that the error slope is fairly asymptotic for $r > 100$.

To use a logistic regression classifier with this data, we build feature matrices out of the training and test images by projecting them on the subset of

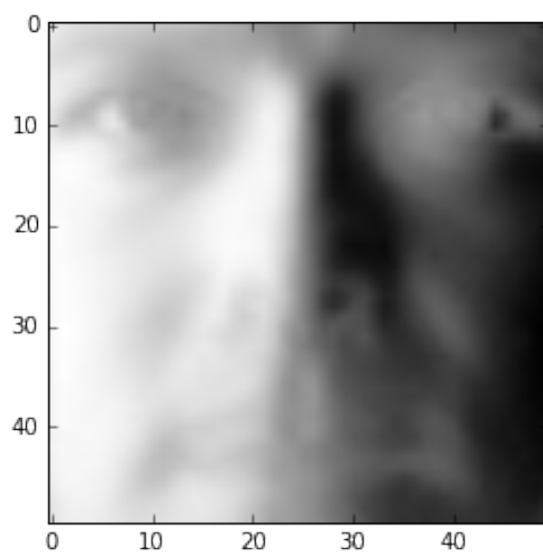


Figure 2: Example eigenface

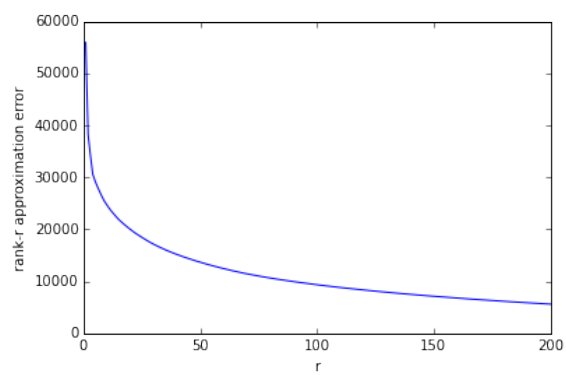


Figure 3: Plot of approximation error

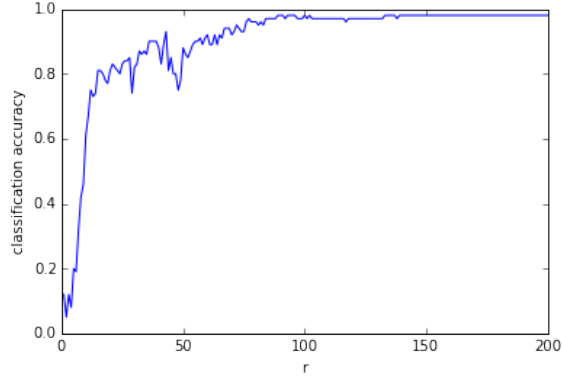


Figure 4: Plot of classifier accuracy

eigenfaces (i.e. multiplying $XV^T[:, r, :]$). At $r = 10$, the logit model performs with classification accuracy 0.61. Based on the previous plot of approximation error in Figure 3, we would expect our classifier to see little improvement in accuracy beyond $r = 100$, and in fact, the plotted accuracy in Figure 4 shows that this is the case.