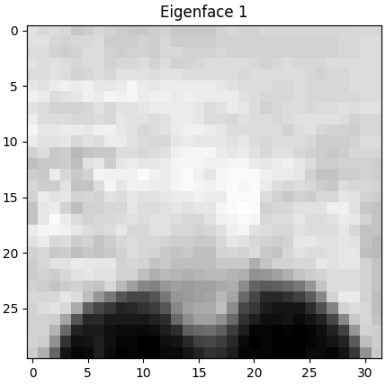
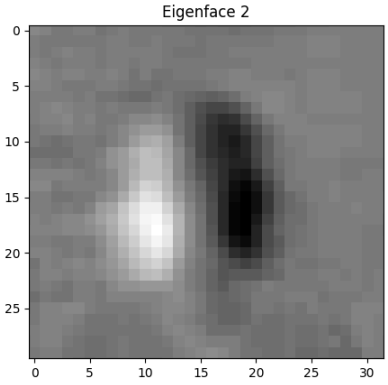
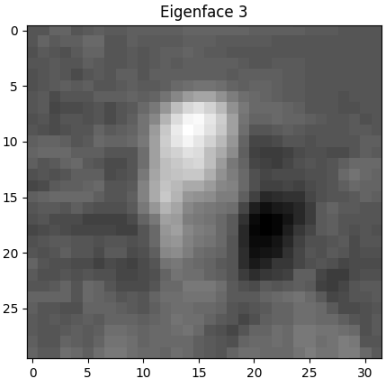
Bryce Rothschadl
Dr. Hien Nguyen
COMPSCI 767-01: Big Data and Data Mining
2025-05-14

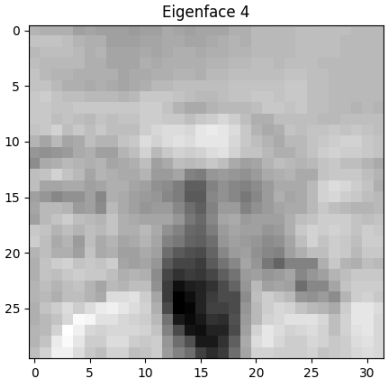
Assignment 6: PCA Implementation

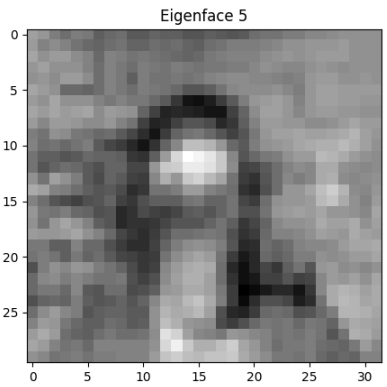
B. Eigenfaces



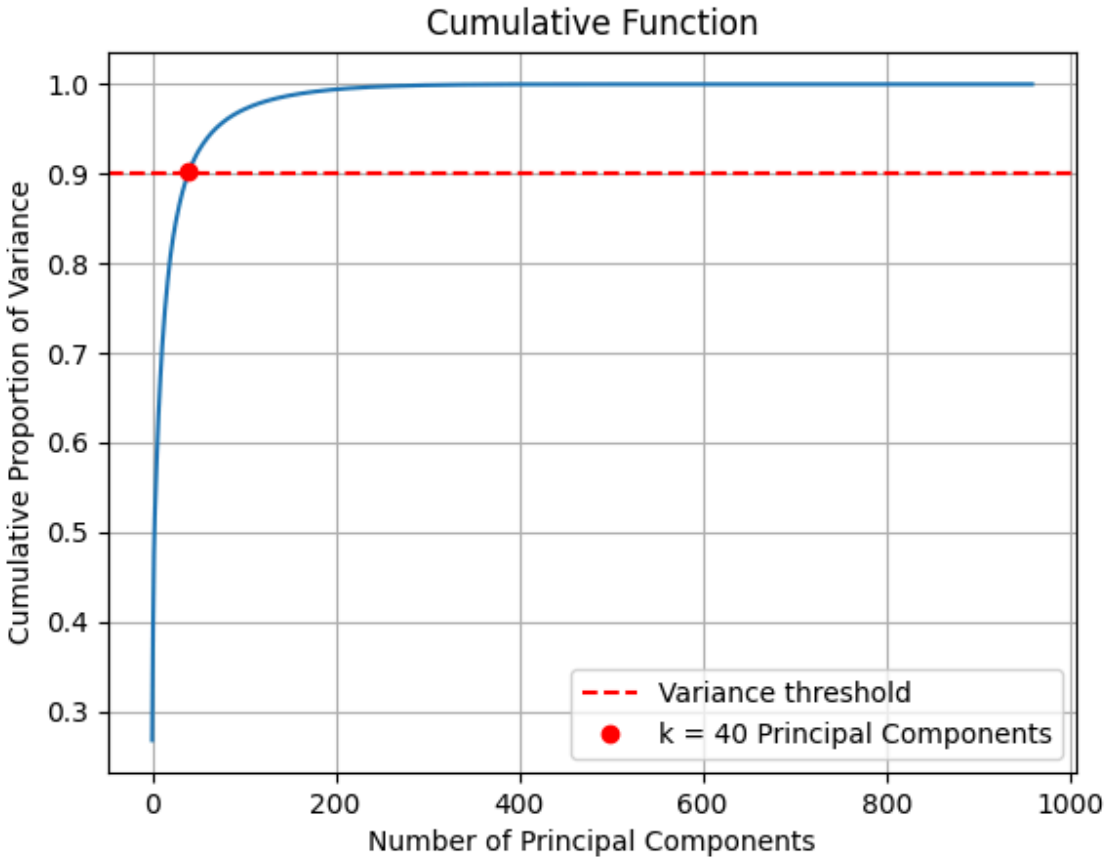








C. Proportion of Variance

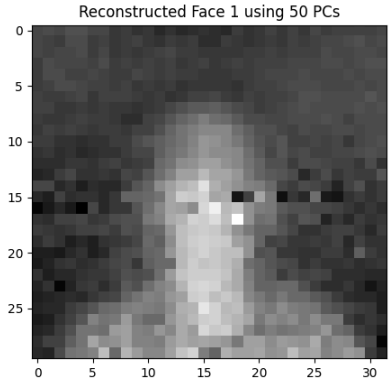


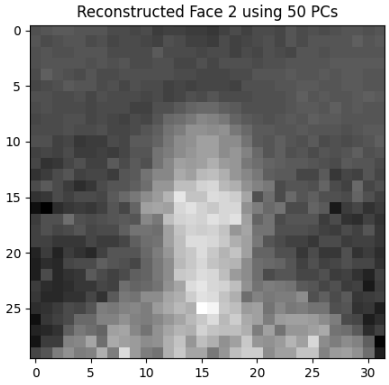
D. KNN Prediction Accuracy

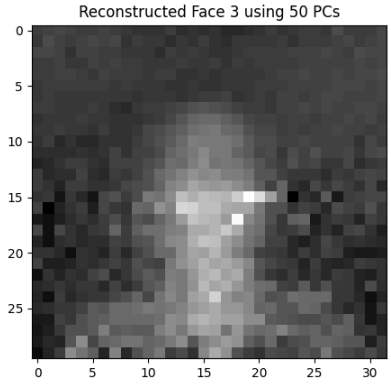
Number of Principal Components (K)	Accuracy (for KNN=3)
1	48.78%
3	60.16%
5	66.67%
7	71.54%

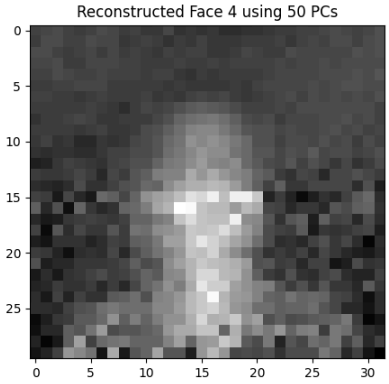
E. Reconstructed Faces

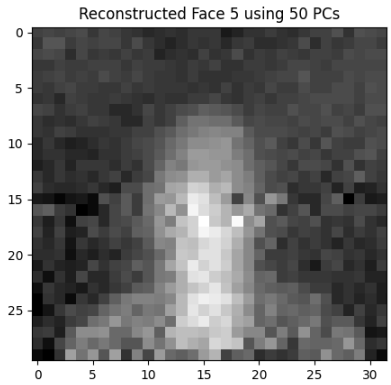
50 Principal Components





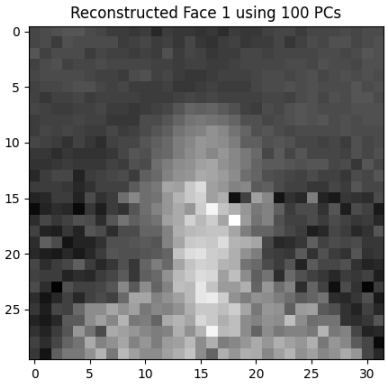


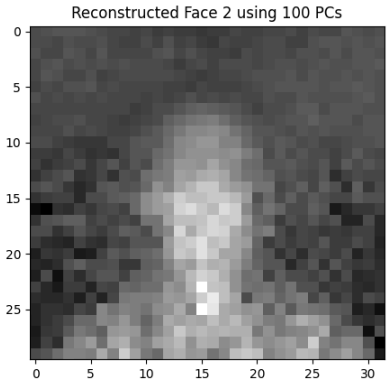


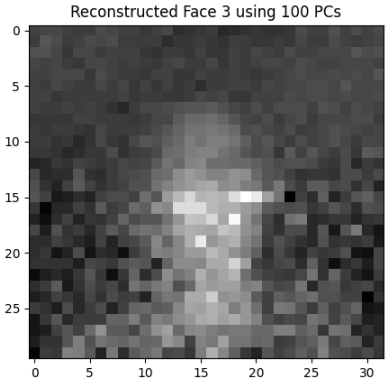


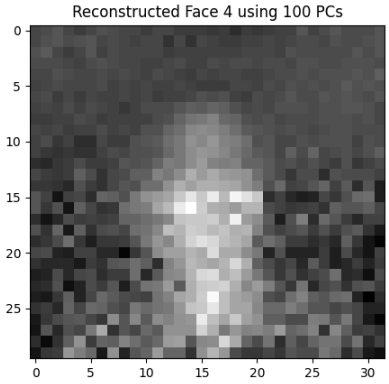
I noticed that these reconstructed images look significantly more similar to one another than their original counterparts. This makes sense, as many distinctive features are inevitably lost after dimensionality reduction. If I were given the two sets of images and was tasked with determining which two from each set corresponded to one another I don't think I would be able to do it with better accuracy than random chance.

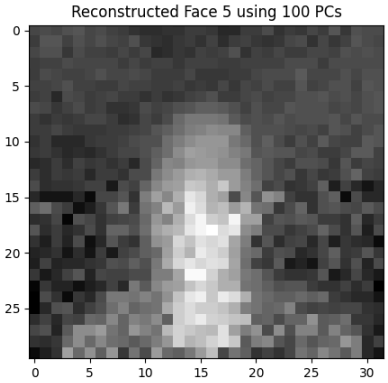
100 Principal Components











Increasing the number of principal components to 100 does not cause the images to change much, they still look nearly identical to the images reconstructed using 50 PCs. This makes sense because we found that 40 PCs already explain >90% of the variance, so moving from 50 to 100 PCs yields diminishing returns. According to my plot of the cumulative function, the additional 50 PCs explain <5% of the total variance.