

Machine Learning Project 4

- b03902089 資工三 林良翰

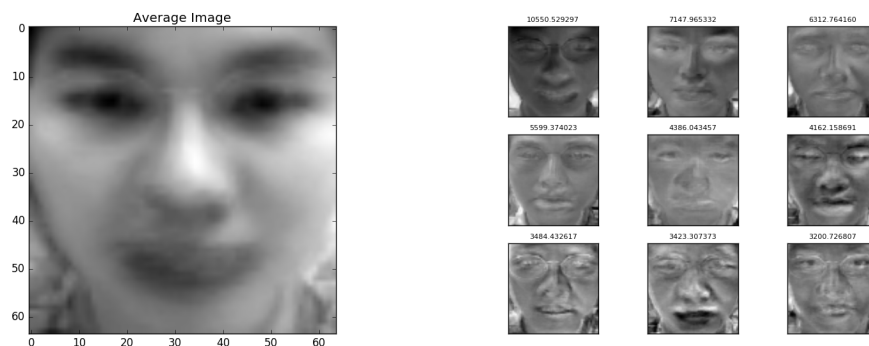
Eigenfaces with PCA

Data Sets

- Carnegie Mellon University - Electrical & Computer Engineering
- Project: Eigenflow Based Face Authentication
(<http://chenlab.ece.cornell.edu/projects/FaceAuthentication/Default.html>)
- Download: link (<http://chenlab.ece.cornell.edu/projects/FaceAuthentication/download.html>)

Questions

1. Perform PCA using the first 10 faces of the first 10 subjects to obtain the eigenfaces. Plot the average face. Also plot the top 9 eigenfaces in a figure.



2. Project the 100 faces onto the top 5 eigenfaces, and then reconstruct the original images. Plot the 100 original faces and the recovered faces.



3. In 2., we can choose top k eigenfaces and check the reconstruction error (RMSE). Find the smallest k such that the error is less than 1%.

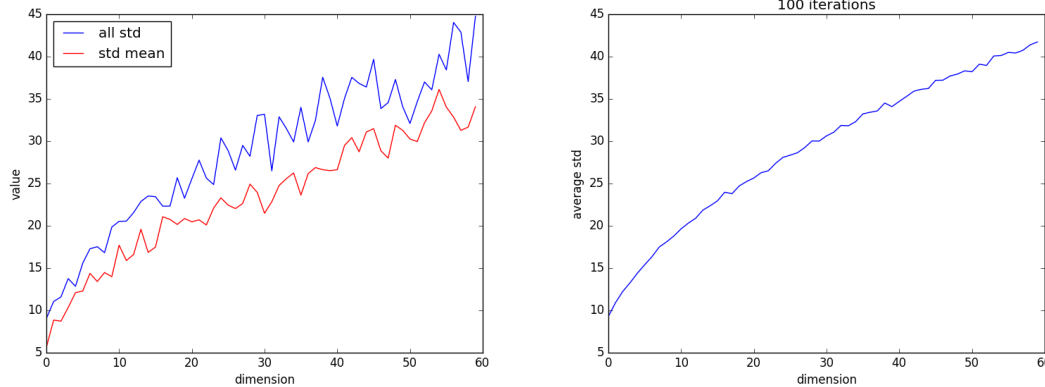
- Smallest $k = 56$, $RMSE = 0.98\%$

matrix W , vector b are sampled from $N(0, 0.5)$

Questions

1. Please elaborate your method and why you used that method. Discuss the results in detail.

- I modified `gen.py` to run on $[1, 60]$ dimensions, with random sample size $N \in [10^4, 10^5]$, $h_i \in [60, 79]$, and found that the dimension of input d_i and the standard deviation of output σ_i are highly positive related.



- Method 1: K-Means Clustering
 - According to the positive correlation between input d_i and output σ_i , use K-means clustering to find 60 clusters $[k_1, \dots, k_{60}]$ and label the dimension of each cluster with respect to the mean of standard deviation $[\sigma_{k_1}, \dots, \sigma_{k_{60}}]$.
 - Error on Kaggle public test: 0.15632
- Method 2: K-Means Clustering with Initial Centers
 - Besides K-means clustering, I generate 60 averaged centers of output σ_i from input $d_i \in [1, 60]$ for 100 iterations, and let them be the initial centers of k-means clustering.
 - Error on Kaggle public test: 0.13157
- Method 3: Initial Centers ONLY!
 - Simply trust the centers generated by myself, and find the closest center for each data set for labeling dimensions.
 - Error on Kaggle public test: 0.11435

2. Download the hand rotation sequence dataset, try to estimate the intrinsic dimension of this dataset and discuss your result.

- Download: link (<http://vasc.ri.cmu.edu/idb/html/motion/hand/index.html>)