Homework 1: Eigendigits

Shun Zhang

1 Introduction

In this report, I applied principal components analysis on digits classification problem. 60000 training samples and 10000 testing samples are provided.

Eigenvectors on first 2000 training samples.

Reconstructed digits using first 100 and 200 eigenvectors correspondingly. First 2000 training samples are used.

2 Experiments

On different size of data.

Using different number of eigenvectors.

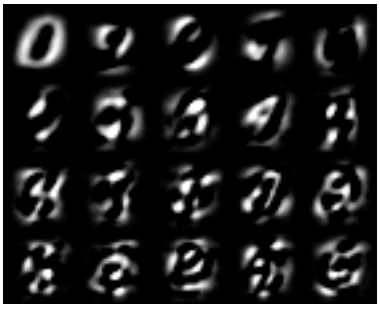

Figure 1: First 20 eigen-vectors, aligned from left to right and top to down. Each value is multiplied by 10, i.e., each vector has norm of 100, instead of 1. They become dark after normalization.

Figure 2: Reconstruction of digits. From left to right on each line, there are original digits, digits constructed by first 100 eigenvectors, digits constructed by first 200 eigenvectors, digits constructed by first 600 eigenvectors.

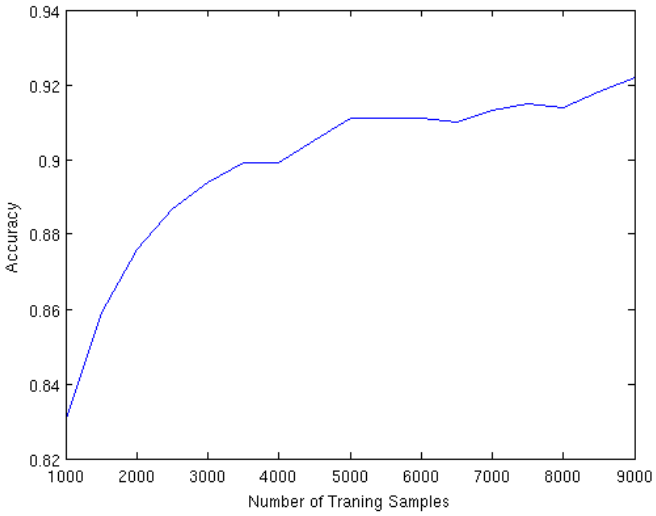

Figure 3: Comparison on using different number of training samples. Testing on first 1000 testing set. Using first 200 eigen-vectors. K for K-nearest-neighbors is 1.

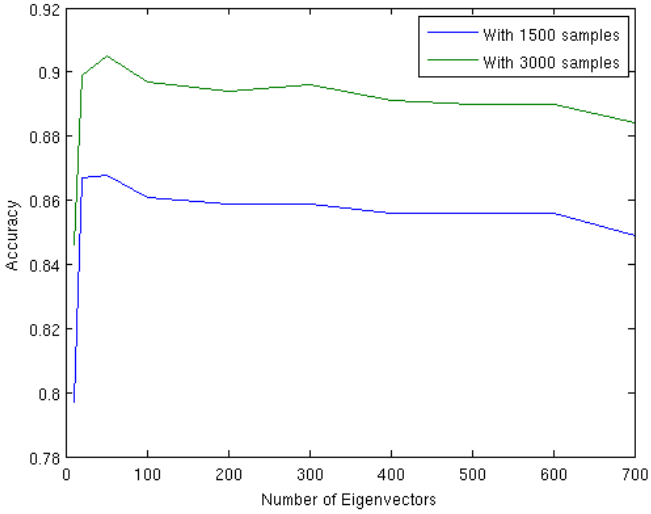

Figure 4: Comparison on using different number of eigenvectors. Testing on first 1000 testing set. Using first 1500 and 3000 training samples for each line. K for K-nearest-neighbors is 1.