Amath 482 Homework 4 Report
Classifying Digits

Colin Kong

March 10th, 2021

Abstract: This assignment started with an SVD analysis of the given images, and then we will perform the digits of the analysis.

Section I: Introduction and Overview

Introduction: The experiments are an attempt to illustrate various aspects of the PCA, its practical usefulness, and the effects of noise on the PCA algorithms. We are showing the experiments in four different tests of 'Ideal Case', 'Noise Case', 'Horizontal Displacement' and 'Horizontal Displacement and Rotation'.

Overview:

I started with loading the videos from the three cameras, and then, in order to narrow down the whole image of the video down to the targeted part of the mass and the spring, I applied a Shannon filter. After filtering, I transformed the colored image to the grayscale image. I calculated for the region of the position of the mass and after that, I applied the PCA algorithm to do the analysis for the videos.

Section II: Theoretical Background

SVD (Singular Value Decomposition):

The reduced SVD is not the standard definition of the SVD used in the literature. What is typically done is to construct a matrix U from $\widehat{U}$ by adding an addition m−n column that are orthonormal to the already existing set $\widehat{U}$. In this way, the matrixUbecomes a squarem×mmatrix, and in order to make the decomposition work, and additional m−n row of zeros is also added to the matrix $\widehat{\sum}$, resulting in the matrix Σ. This leads to the singular value decomposition of the matrix A: A = UΣV^*,where U ∈ $\mathbb{R}^{m*n}$ and V ∈ $\mathbb{R}^{m*n}$ are unitary matrices, and Σ ∈ $\mathbb{R}^{m*n}$ is diagonal. Geometrically, the SVD is doing the following. Multiplying by V^* on the left rotates the hypersphere to align the v_n vectors with the axes. Then we multiply on the left by a diagonal matrix which stretches in each direction. Then we multiply on the left by U to rotate the hyper ellipse to the proper orientation.

To compute SVD, we have:

$$\begin{aligned} A*A &= (U\Sigma V^*)*(U\Sigma V^*) \\ &= V\Sigma U*U\Sigma V^* \\ &= V\Sigma^2 V^*. \end{aligned}$$

Multiplying on the right by V gives

$$A*AV = V\Sigma^2,$$

so that the columns of V are eigenvectors of A∗A with eigenvalues given by the square of the singular values. Similarly,

$$AA^{*} = U\Sigma^{2}U^{*}.$$

and so, multiplying on the right by U gives the eigenvalue problem

$$AA^{*} * U = U\Sigma^{2}$$

which can be used to solve for U.

PCA (Principle Component Analysis):

Let us now imagine we have multiple vectors. For example, imagine there were 10 assignments and each vector contains grades from the same 20 randomly chosen students. We can put each of those row vectors into a matrix:

$$X = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

We can compute all the variances and covariances between the rows of X with one matrix multiplication:

$$C_X = \frac{1}{n-1}XX^T = \begin{bmatrix} \sigma_a^2 & \sigma_{ab}^2 & \sigma_{ac}^2 & \sigma_{ad}^2 \\ \sigma_{ba}^2 & \sigma_b^2 & \sigma_{bc}^2 & \sigma_{bd}^2 \\ \sigma_{ca}^2 & \sigma_{cb}^2 & \sigma_c^2 & \sigma_{cd}^2 \\ \sigma_{da}^2 & \sigma_{db}^2 & \sigma_{dc}^2 & \sigma_d^2 \end{bmatrix}$$

You should notice that C_X is a square symmetric matrix. Unsurprisingly, it is called the covariance matrix. The goal of principal component analysis is to find a new set of coordinates (a change of basis) so that the variables are now uncorrelated. That will mean that each variable contains completely new information, i.e. no redundancies. It would also be nice to know which variables have the largest variance because these contain the most important information about our data. Therefore, we want to diagonalize this matrix so that all off-diagonal elements (covariances) are zero:

$$C_X = V\Lambda V^{-1}.$$

The basis of eigenvectors contained in V are called the principal components. They are uncorrelated since they are orthogonal. Why? Since C_X is a symmetric matrix, its eigenvalues are real, and the corresponding eigenvectors are orthogonal. The diagonal entries of Λ, the eigenvalues of C_X, are the variances of these new variables. The connection to the SVD for PCA starts with the SVD of a matrix A is connected to the eigenvalue decomposition of AA^T. To account for the $\frac{1}{n-1}$ factor, consider

$$A = \frac{1}{\sqrt{n-1}}X$$

Then,

$$C_X = \frac{1}{n-1} X X^T = \mathrm{A} A^T.$$

Then, from the following of SVD, we have that $C_X = \mathrm{A}A^T = \mathrm{U}\Sigma^2 U^T$, where U is the (orthogonal) matrix of left-singular vectors and Σ is the diagonal matrix of singular values. So, the eigenvalues of the covariance matrix are the squares of the (scaled) singular values. Recall that this factor of $\sqrt{n-1}$ is exactly what we scaled the singular values by to get the right units. Recall that we used a change of basis to work in the basis of the principal components. To do this you multiply by $U^{-1} = U^T$. The data in the new coordinates is

$$Y = U^T X.$$

The covariance of Y is

$$C_Y = \frac{1}{n-1} Y Y^T = \frac{1}{n-1} U^T X X^T U = U^T \mathrm{A} A^T U = U^T \mathrm{U} \Sigma^2 U^T \mathrm{U} = \Sigma^2$$

Since the off-diagonal elements of Σ are zero, it follows that the variables in Y are uncorrelated.

LDA (Linear Discriminant Analysis):

The goal for LDA is to find a suitable projection that maximizes the distance between the inter-class data while minimizing the intra-class data. First, we calculate the means for each of our groups for each feature. As above, we will call these μ1 and μ2. Note that these μ are column vectors since they are means across each row (see the code below if you're confused). We can then define the between-class scatter matrix

$$S_B = (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T$$

This is a measure of the variance between the groups (between the means). Then we can define the within-class scatter matrix

$$S_\omega = \sum_{j=1}^{2} \sum_{x} (x - \mu_j)(x - \mu_j)^T$$

This is a measure of the variance within each group. The goal is then to find a vector w such that

$$w = argmax \frac{w^T S_B w}{w^T S_\omega w}$$

This is a tough problem to solve, but luckily for us it turns out (we won't prove this) that the vector w that maximizes the above quotient is the eigenvector corresponding to the largest eigenvalue of the generalized eigenvalue problem

$$S_B w = \lambda S_\omega w$$