

Dimensionality Reduction

Andrés Ponce

April 21, 2020

Some problems inherently can be modeled using less dimensions than might appear at first sight. Taking the well-known digit recognition problem, even though a sample image might consist of 100x100 pixels, any digit on the image can only be transformed by rotation, translation, and scaling. We can study this pattern further.

Principal Component Analysis

Principal Component Analysis (PCA) is a popular way of reducing dimensions in a given problem by attempting to find a lower dimensional space on which to map the data points.

However, we can also think of PCA as the projection of the points onto a plane such that their *variance* is maximized.¹ Similar to a linear discriminant, we want to maximize the variance when we project it onto a smaller subspace. This will make it so that the points overlap less when brought down to the same line. Then we have the **data mean** and the **projected data mean**

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad \text{and} \quad u_1^T \bar{x} \quad (1)$$

Where u refers to the mean once the data is projected onto the line.

Finally, the **covariance matrix** refers to ²

$$S = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^T$$

The covariance then tells us how the values of the projected data and the regular data differ.

Lagrangian Multipliers

To remind, **Lagrangian Multipliers** allow us to solve **constrained optimization**. Basically, we want to optimize an equation subject to a set of constraints. In this case, we have the unit direction vectors in M dimensional space u , so $u_1^T u_1 = 1$. This equation acts as our constraint in this case, ³. And the equation we are then trying to optimize becomes $u_1^T S u_1$.

To conclude discussion on PCA, we have to do

1. Find the mean \bar{x} and covariance matrix S

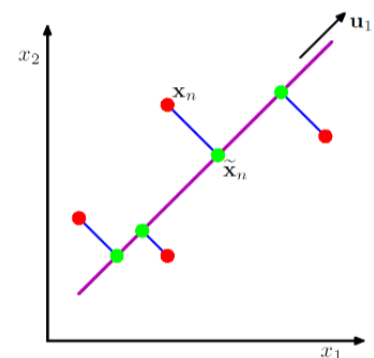


Figure 1: For original dimensions x_1 and x_2 , the **line** is the new space we want to map onto. We minimize the sum of the projections of the **data point** onto the line, which is at the **dot**

¹ This seems similar to a **linear discriminant**?

² So, basically the amount the projected data and the regular data differ squared, and then take the mean?

³ is this because the covariance can be at most 1?

2. Find the M eigenvectors of S corresponding to the M largest eigenvalues

The eigenvector decomposition takes $O(D^3)$ and finding the M largest eigenvectors and eigenvalues takes $O(MD^2)$

Minimum Error Formulation

We are still trying to bring down the original dimensionality to a lower level, and we introduce another method called the **Minimum Error Formulation**. This approach again reduces a space to another M dimensional space.

Originally, we can have D **orthonormal**⁴ vectors. We then express any position in D dimensions as a linear combination of the D orthonormal vectors. Our goal is to then express those same points with a linear combination of M vectors instead, which would be equivalent to projecting the vectors onto a smaller subspace.

Then, for this approach to work, if we are to use only M eigenvectors in subspace u , we would still need to perform the decomposition since we would need to find the largest M eigenvectors. Then the approximate point \tilde{x}_n can be written as

$$\tilde{x}_n = \bar{x} + \sum_{i=1}^M (x_n^T u_i - \bar{x}^T u_i) u_i$$

⁴ Two vectors are **orthonormal** if they are both orthogonal (i.e. perpendicular) and unit vectors.