Isabelle Pardew

06/14/17

PCA Notes and Observations

**PCA** creates a linear combination of the many variables we are given about an object (say, a photograph of a face), and we can create a scatter plot with the first principal component on the x-axis, and the second on the y-axis.

**The first component shows the greatest variance of all the variables, and the second is the second most variance.**

PCA looks for properties that show as much variance across faces as possible.

(You look for properties that would allow you to reconstruct the original image of the face)

* Each dot in a scatter plot would be an image of a face, and we would see how well they can be distinguished
* Eigenvectors are just the linear combinations of the original variables (in the simple or rotated factor space); they described how variables "contribute" to each factor axis

Ex.: "Let's say we have high dimension data - say 30 measurements made on an insect. The bugs have different genotypes and slightly different physical features in some of these dimensions, but with such high dimension data it's hard to tell which insects belong to which group.”

PCA is a technique to reduce dimension by:

1. Taking linear combinations of the original variables.

2. Each linear combination explains the most variance in the data it can.

3. Each linear combination is uncorrelated with the others" (orthogonal)

SVD of a data matrix,after mean-centering data for each component (see Lay’s Linear Algebra book)

**Identification** - eigenfqces

**Recognition**- matching eigenfaces

**Categorization** - grouping

An NxM image is a point in R^NM

Eigenvectors of the covariance matrix of the set of face images are called the eigenfaces.

PCA makes it easier to do these calculations - dimensionality reductions - calculate eigenvectors from a covariance of reduced dimensionality

* The eigenfaces form a basis for all faces
* Reduce dimensionality by preserving as much variance as we can (the most important distinguishing factors)
* 2 new axes- eigenvector 1&2
* All the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data.