


(Contrastive) Principal Component Analysis

Independent Component Analysis too!

Sam Fuller¹

1. Thanks to Chris Hare for the slides from which this presentation is based 

PCA Review

Unsupervised Models

Unsupervised machine learning is, actually, relatively common in the social sciences but is known under a different label: **Dimensional Analysis**.

As a reminder, **Unsupervised** Machine Learning has *no* specified output only a set of explanatory variables (X_i). It attempts to find relationships between these variables, but is not attempting to predict any outcome variables (Y_i).

Really, unsupervised learning can be motivated by either the curse of dimensionality or by latent dimensions.

The Curse of Dimensionality (A Reminder)

- ▶ As p increases and the dimensionality \uparrow , the volume of the space explodes and the data points grow further and further apart.
- ▶ How many people in this room share the same birth month, as opposed to share the same birth month *and* height *and* party ID *and* ...
- ▶ Sparse data means we have less leverage to estimate the effects of X on Y (that is, f).
- ▶ Frankly, this is probably better appreciated in the machine learning world and too often ignored by social scientists. In machine learning, this motivates dimensional reduction/feature selection methods.

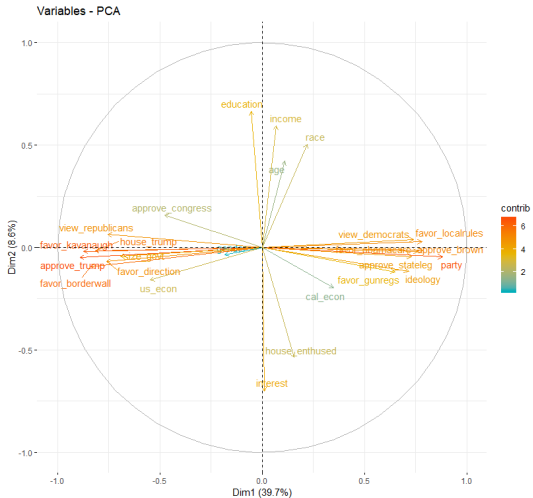
Principal Component Analysis

So let's take a look at **Principal Component Analysis (PCA)**.

Let's say we want to predict someone's vote-choice in a congressional election. However, we want to see if there is some underlying 'dimension' that exists among the survey questions that we ask (e.g. policy preferences, party ID, demographic variables etc.).

PCA generates dimensions using the data that, in order, contain the most variance. Looking at the plot of the first few dimensions that PCA finds can be very illustrative. And we can actually use these identified dimensions in regular statistical models (like logit or probit).

PCA Loadings Plot



PCA Code

Let's go over code for analyzing PCA results! You can follow along with either `PCA_cPCA.Rmd` or `PCA_cPCA.html`.

Also, we'll be using two datasets in this lecture:
`cal_voters.RDS` & `BTP_cPCA.RDS`.

Other Unsupervised Approaches

- ▶ Contrastive Principal Component Analysis (cPCA)
- ▶ Independent Component Analysis (ICA)
- ▶ Multiple Correspondence Analysis (MCA)
- ▶ Factor Analysis (FA)

Other Unsupervised Approaches

- ▶ **Contrastive Principal Component Analysis (cPCA)**
- ▶ **Independent Component Analysis (ICA)**
- ▶ Multiple Correspondence Analysis (MCA)
- ▶ Factor Analysis (FA)

Contrastive Principal Component Analysis (cPCA)

What is cPCA?

Often, components derived from PCA are uninformative or simply uninteresting. To deal with this, and to also further analyze subgroups, Abid et al. (2018) developed **Contrastive PCA (cPCA)**.

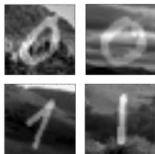
cPCA is conducted by:

1. Splitting the data into different groups, usually by predefined classes (e.g., Party ID).
2. Setting one group as the **target group** and the other(s) as the **background group**.
3. Estimating PCA such that the PCs vary **maximally on the target group** and **minimally on the background group**.

Contrasting

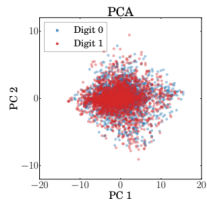
Unlike ordinary approaches that usually aim to capture the characteristics of an entire dataset, cPCA compares subsets of that dataset relative to one another (target versus background group).

cPCA Example

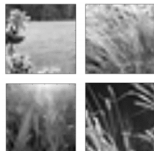


Target Data

(a)

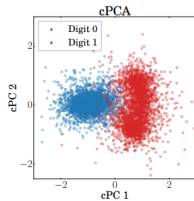


(b)



Background Data

(c)



(d)

cPCA Code

Back to the code! We'll still be using: `PCA_cPCA.Rmd` or `PCA_cPCA.html`.

Independent Component Analysis (ICA)

Using ICA

Independent Components are estimated using the following procedure (Attewell, Monaghan, & Kwong 2015):

1. Choose the number of components to estimate.
2. “Whiten” the dataset, most often by using PCA.
3. Find the decorrelation matrix that is **maximally non-gaussian**.

Thus, the only differences between running PCA and ICA is choosing the number of components to estimate and estimating the decorrelation matrix.

Using ICA cont.

Maximal nonnormality is achieved in one of two ways:

1. Find components that have kurtoses that vary the most from that of a normal distribution
2. Find components that have the largest negentropy (the difference between the component's entropy and that of a normal distribution).

ICA was developed primarily for “blind source separation,” which attempts to separate two independent, but mixed signals (for example mixed audio streams). It focuses on identifying nonnormality because source signals tend to have non-gaussian (normal) probability density functions, whereas the mixed signal's PDF tends to be gaussian.

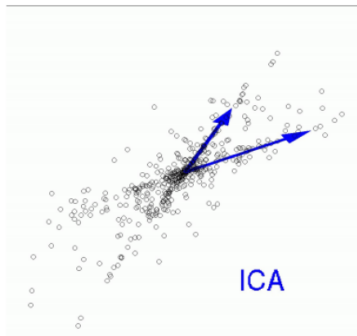
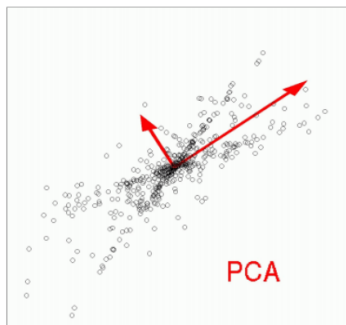
ICA vs. PCA

The primary difference between ICA and PCA is that PCA assumes a normal PDF for each component, ICA does the opposite. Consequently, ICA attempts to find components that are independent (but not orthogonal) from one another.

Other differences include:

- ▶ PCA removes correlations, **but not** higher order dependence.
 - ▶ ICA removes correlations **and** higher order dependence.
- ▶ Components have differing, descending importance in PCA.
 - ▶ All components are considered **equally important** in ICA.
- ▶ Components are orthogonal (so as to maximize variance explained) in PCA.
 - ▶ ICA vectors are not orthogonal.

ICA vs. PCA Cont.



ICA Code

While I've yet to find a social science application of ICA that performs well and differently than PCA, I did try the method on the BtP dataset.

So back to the code for a final time! We're still using:
`PCA_cPCA.Rmd` or `PCA_cPCA.html`.