Intro to Principal Component Analysis (PCA) and Independent Component Analysis (ICA)

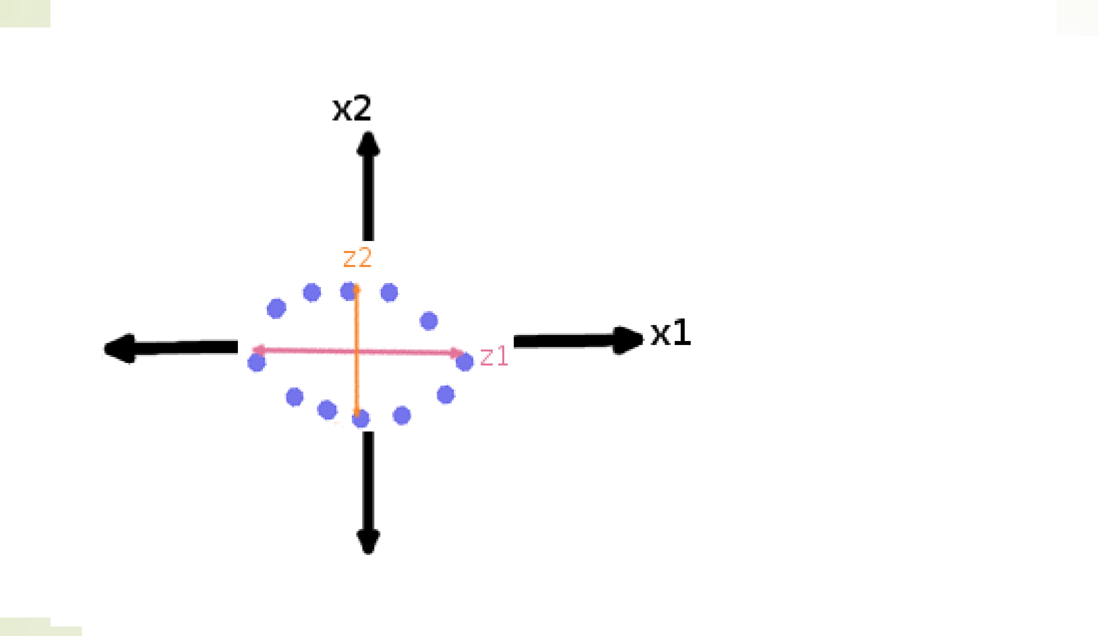
Similarities and Differences

- Both are statistical transformations
  - PCA: information from second order statistics
  - ICA: information that goes up to high order statistics
- Both used in various fields:
  - Blind source separation, feature extraction, neuroscience!

PCA

- Often used prior to running machine learning algorithm
- Finds principal components of the dataset
  - Each succeeding step finds direction that explains most variance
  - Transforms data into new subspace
    - First axis corresponds to first principal component (explains greatest amount of variance in the data)

Visualizing PCA



- 2-dimensional data (x_1 and x_2)
- 2 (or less) orthogonal principal components
- Reframed data

PCA

- Principal Components = eigenvectors of covariance matrix of original dataset
  - Eigenvectors are orthogonal (covariance matrix is symmetric)
  - Principal components correspond to **direction** (in original space) with greatest variance in data
- Each eigenvector has an associated eigenvalue
  - Eigenvalue is a scalar that indicates **how much** variance there is in the data along that principal component
  - If PCA is used for dimensionality reduction, generally discard principal components with zero or near-zero eigenvalues

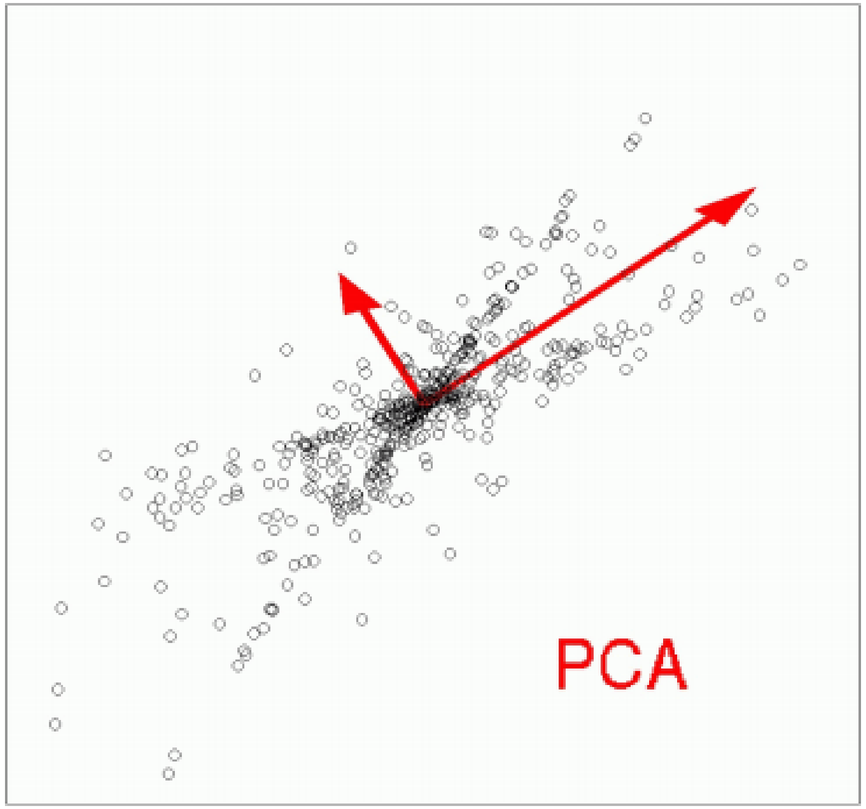
Algebraic Definition of Principal Components

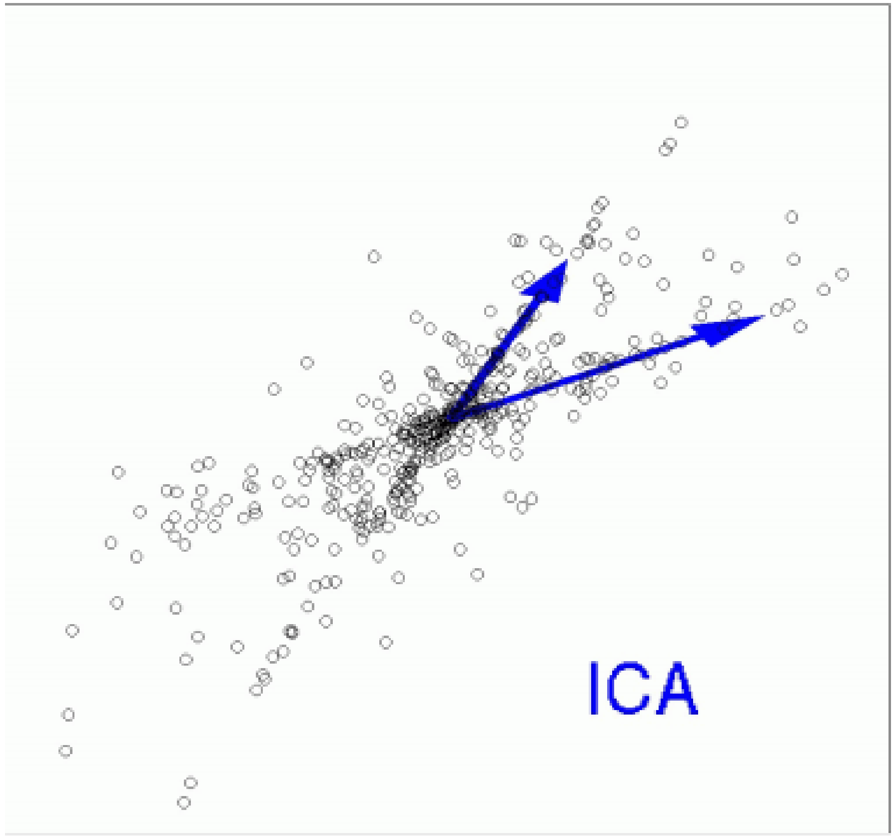
- Sample of *n* observations, each with *p* variables:
  - $x = (x_1, x_2, \dots, x_p)$
- First principal component: $z_1 \equiv a_1^T x = \sum_{i=1}^{p} a_{i1} x_i$
  - Where vector $a_1 = (a_{11}, a_{21}, \dots, a_{p1})$ st. $var[z_1]$ is a maximum
- k[th] principal component: $z_k \equiv a_k^T x = \sum_{i=1}^{p} a_{i1} x_i$
  - Where vector $a_k = (a_{1k}, a_{2k}, \dots, a_{pk})$ st. $var[z_k]$ is a maximum
  - Subject to: $cov[z_k, z_l] = 0 \; for \; k > l \geq 1$
  - And to: $a_k^T a_k = 1$

Differences between ICA and PCA

- PCA removes correlations, **but not** higher order dependence
  - ICA removes correlations **and** higher order dependence
- PCA: some components are more important than others (recall eigenvalues)
  - ICA: all components are **equally important**
- PCA: vectors are orthogonal (recall eigenvectors of covariance matrix)
  - ICA vectors are **not orthogonal**

PCA vs ICA





Algebra Behind ICA

- Assume there exist independent signals: $S = [s_1(t), s_2(t), \dots, s_N(t)]$
- Linear combinations of signals: $Y(t) = A\,S(t)$
  - Both A and S are unknown
  - A is called the mixing matrix
- Goal of ICA: recover original signals, S(t) from Y(t)
  - Ex. find a linear transformation, L, ideally A^{-1} st. LY(t) = S(t)

ICA

- Generally, preprocess data before applying ICA to remove correlation (“whitening”)
  - PCA is one way to whiten signals
- Address higher order dependence