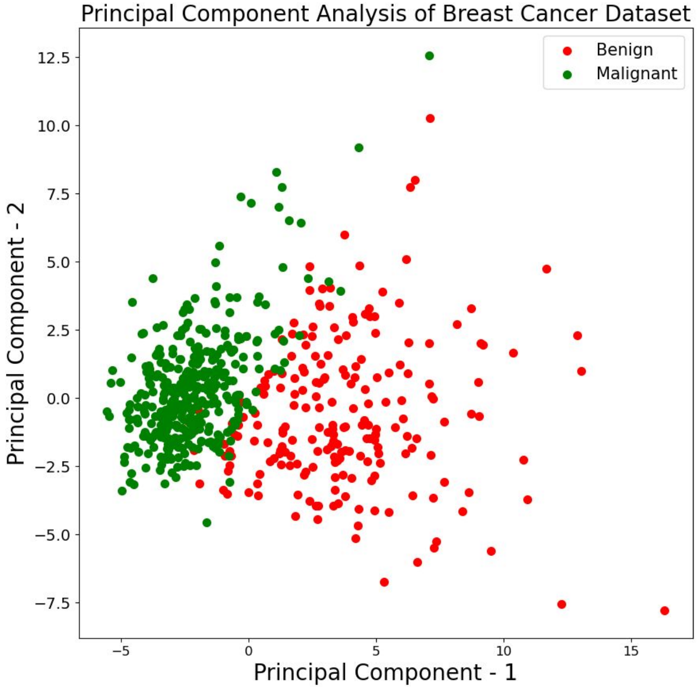
T1.2 Dimensionality reduction using PCA

Joshua Llano
joshua.llano@upc.edu

What reducing the dimensionality is useful for?

- Data Visualization:
  e.g. Breast Cancer dataset
  (569 samples x 30 features)



- Speeding Machine Learning Algorithm:

  more # features* = more training and testing time!

***Note**: Features, Dimensions, and Variables are all referring to the same thing.

Principal Component Analysis

”PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.”

Principal Component Analysis

”PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.”

“PCA seeks a small number of dimensions that are as interesting as possible, where the concept of interesting is measured by the amount that the observations vary along each dimension. Each of the dimensions found by PCA is a linear combination of the *p* features.”

Principal Component Analysis

”PCA is defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.”

“PCA seeks a small number of dimensions that are as interesting as possible, where the concept of interesting is measured by the amount that the observations vary along each dimension. Each of the dimensions found by PCA is a linear combination of the *p* features.”

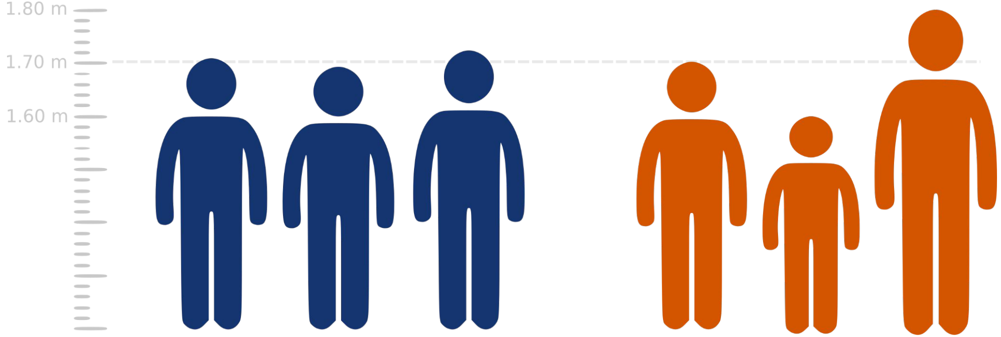
“PCA is an unsupervised approach, since it involves only a set of features X1, X2, . . . , Xp, and no associated response Y .”

Variance

The variance is a measures of the average distance between each point and their mean:

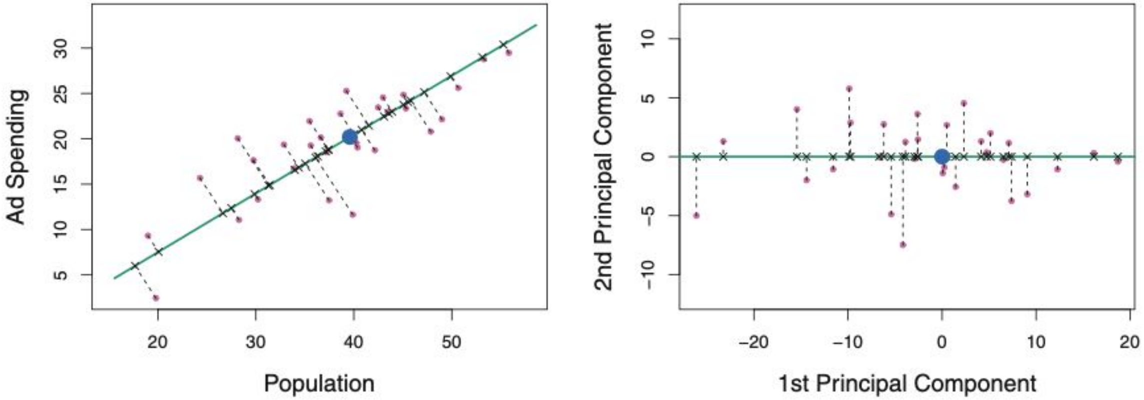
$$\sigma^2 = \frac{\Sigma(xi - \bar{x})^2}{N}$$

The greater variance, the more the information!



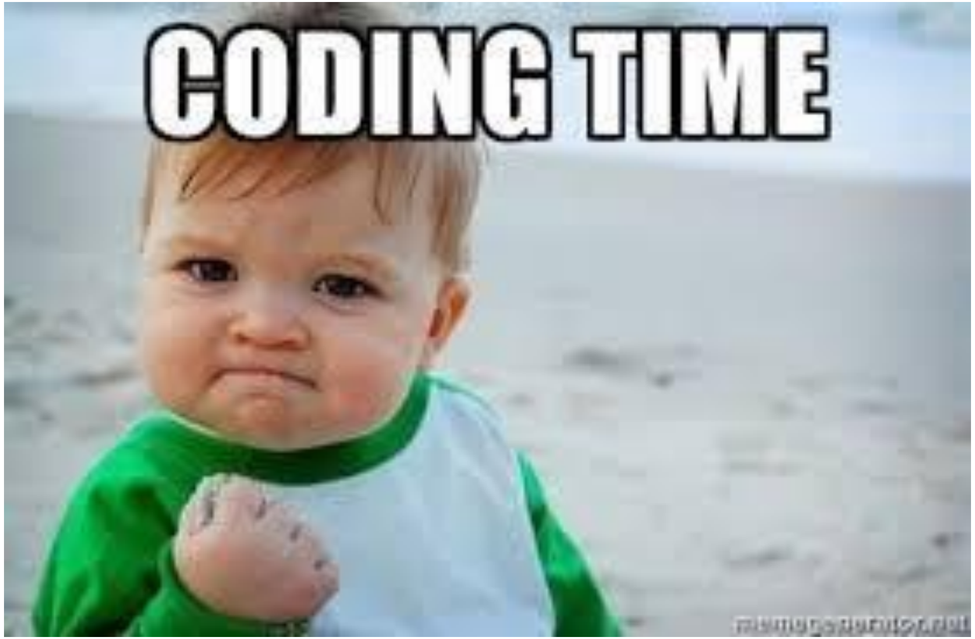
How PCA works?

Main idea (with no math): To create PC1, a line is anchored at the center of the points and rotate in all directions while points are projected in this line. This rotation continues until the total distance among projected points is maximum. The rotating line now describes the direction with the most variation among the data.



Key Steps of PCA in Practice

1. Mean subtraction (It is not strictly necessary but reduces the risk of numerical problems)
2. Standardization (Divide the data points by the standard deviation of the dataset for every dimension)
3. Eigendecomposition of covariance matrix (Compute the data covariance matrix and its eigenvalues and corresponding eigenvectors)
4. Projection (We can project any data point onto the principal subspace)


CODING TIME