Principal Component Analysis: The Algorithm & Important Properties

Ira A. Fulton Schools of Engineering
Arizona State University

Objective

Objective

Implement the PCA algorithm

Objective

Discuss some important properties of PCA

Principal Components

- We found $\mathbf{e}_1$, which gives the direction of the largest variance after projection

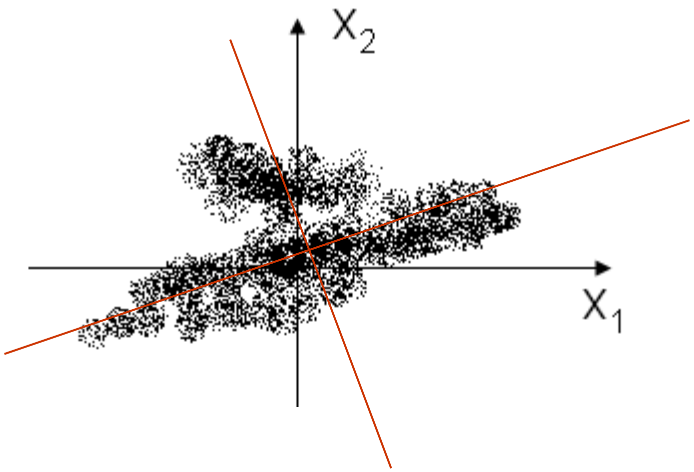
  ➔ The first **principal component.**

- The process can be continued in the subspace orthogonal to $\mathbf{e}_1$, and so on and so forth.

  ➔ Obtaining other principal components: $\mathbf{e}_2$, $\mathbf{e}_3$, etc., corresponding to other eigenvectors of C, ordered by the corresponding eigenvalues λ_i

Principal Components (cont'd)

- The principal components are orthogonal to each other → $\{\mathbf{e}_i\}$ forms an orthonormal basis in the d-dimensional space.



- The total variance is given by the sum of the variances of the projections.

$$\sigma^2 = \sum_{j=1}^{d} \lambda_j$$

How Many Principal Components to Keep?

- To reduce dimensions, we will need to keep only *d'* << *d* projections.

- We can measure how much of the total variance a *d'*-dimensional subspace captures, by the ratio

$$\sum_{j=1}^{d'} \lambda_j \Big/ \sum_{j=1}^{d} \lambda_j$$

- Variance may be related to the "energy" of a signal: how accurately we want to represent the data.

➔ The ratio can be used to guide in choosing a proper *d'* for desired accuracy.

The PCA Algorithm

1. Compute the dxd sample covariance matrix C

2. Find the eigenvalues and corresponding eigenvectors of C

3. Project the original data onto the space spanned by the eigenvectors
   - The projection may be done onto a d'-dimensional subspace spanned by the first d' eigenvectors (ordered by the eigenvalue in descending order)
     - d' is determined by the desired accuracy

Important Properties of PCA

- PCA represents the data in a new space, in which the components of the data is ordered by their "significance".

  ➔ Dimension reduction can be done by simply discarding less significant dimensions.

- Linearity assumption ➔ extensions exist

- "**Variance ≈ Importance**" is meaningful only under large *signal-to-noise ratio*

PCA as Feature Mapping

- When we use only *d*' dimensions from PCA (with original dimension *d* > *d*'), this may look like feature section.
  - But in general they are different approaches.

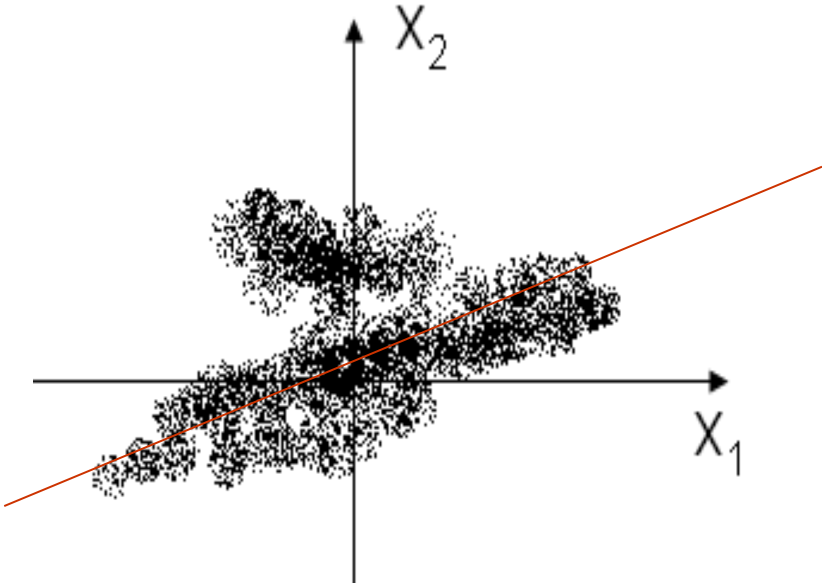
- PCA
  - Unsupervised (in general)
  - Generates new features (linear combination of original ones)

- Feature Selection
  - Supervised (in general)
  - Selects a few original features (e.g., for better classification)

Can PCA help classification?

- Can we do better classification in a lower-dimensional space from *d'* principal components given by PCA?

    ➔ Not necessarily.



- LDA may be better posed for such a task.