

Introduction to Machine Learning Applications

Spring 2023

Dimensionality reduction

Minor Gordon

gordom6@rpi.edu



Rensselaer

Curse of dimensionality

When dimensionality increases, data becomes increasingly sparse in the space that it occupies

Dimensionality reduction

- Purpose:
 - Avoid curse of dimensionality
 - Reduce amount of time and memory required by algorithms
 - Allow data to be more easily visualized
 - May help to eliminate irrelevant features or reduce noise
- Techniques
 - Principal Component Analysis (PCA)
 - Singular Value Decomposition
 - Others: supervised and non-linear techniques

Feature selection

- Redundant features
 - Duplicate much or all of the information contained in one or more other attributes
 - Example: purchase price of a product and the amount of sales tax paid
- Irrelevant features
 - Contain no information that is useful for the task at hand
 - Example: students' ID is often irrelevant to the task of predicting students' GPA
- Many techniques developed, especially for classification

Feature creation

- Create new attributes that can capture the important information in a data set much more efficiently than the original attributes
- Three general methodologies:
 - Feature extraction
 - Example: extracting edges from images
 - Feature construction
 - Example: dividing mass by volume to get density
 - Mapping data to new space
 - Example: Fourier and wavelet analysis

Principal Component Analysis



Step-1: Standardization



Step-2: Compute covariance matrix



Step-3: Compute the eigenvalues and eigenvectors of the covariance matrix

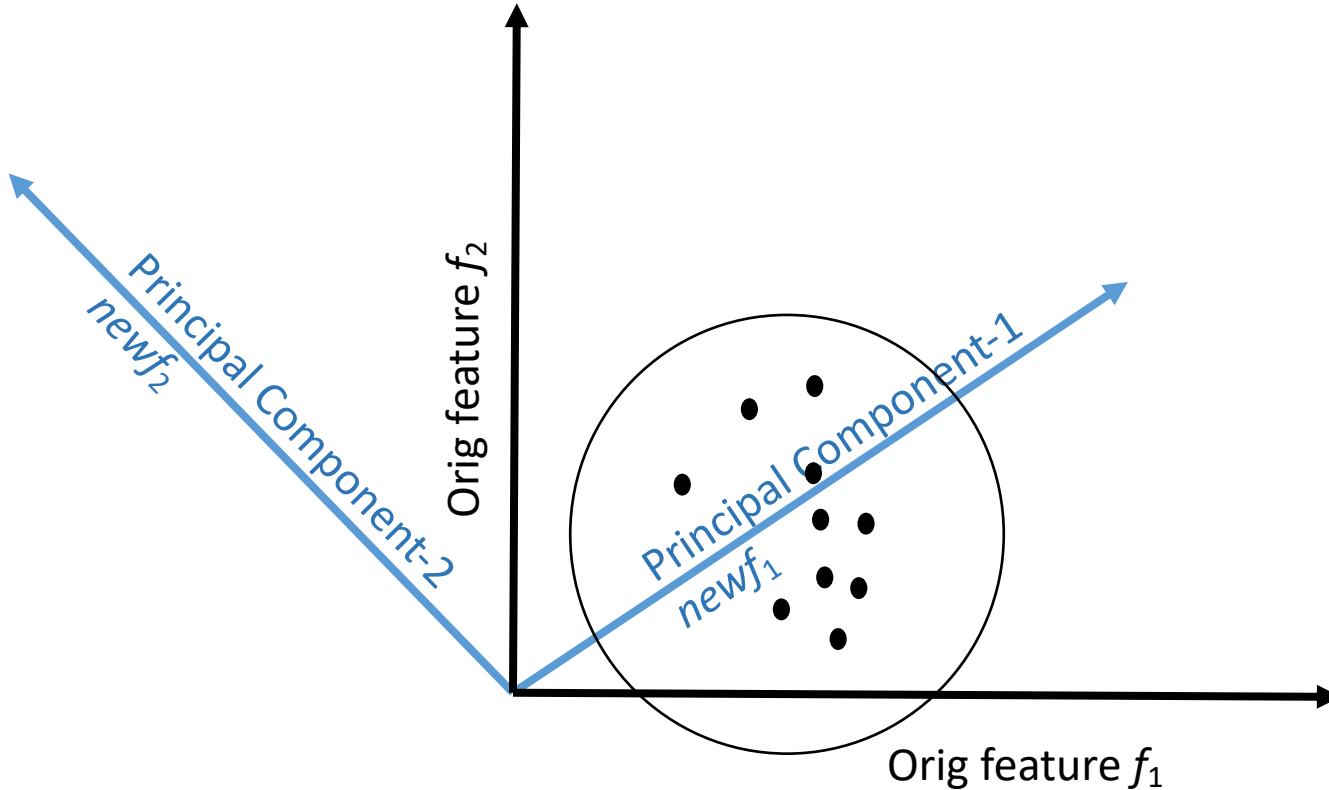


Step-4: Sort the eigenvalues in a decreasing order



Step-5: Choose the top-k eigenvectors which are the principal components – these will be the transformed feature vectors (reorient the data is the common approach)

Dimensionality Reduction PCA Example



Learning PCA

- [Making sense of principal component analysis, eigenvectors & eigenvalues](#)
- [StatQuest: Principal Component Analysis \(PCA\), Step-by-Step](#)