Unsupervised Learning

Survey 1

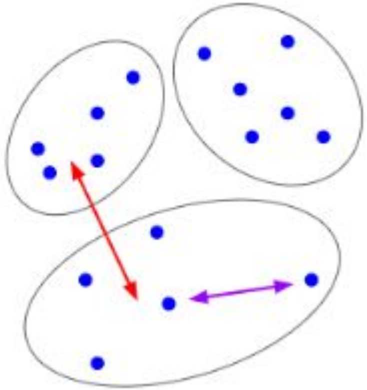
Template by: Laurent Lessard

Unsupervised learning

- **Unsupervised learning** is when our data consists of examples (rows) and features (columns). It is the broad task of describing how our data is organized. This includes:
    - Discovering groups of similar examples (e.g., clustering)
    - Reduce dimensionality for purpose of visualization (e.g., PCA)
    - Estimate the underlying distribution of data (density estimation)

Recap: K-Means clustering

- **Similar** items in the same cluster
- **Dissimilar** items in different clusters
- Each point falls in exactly one cluster (eg. KMeans)



Inter-cluster distances maximized

Intra-cluster distances minimized

Partitional Clustering

K-Means vs PCA

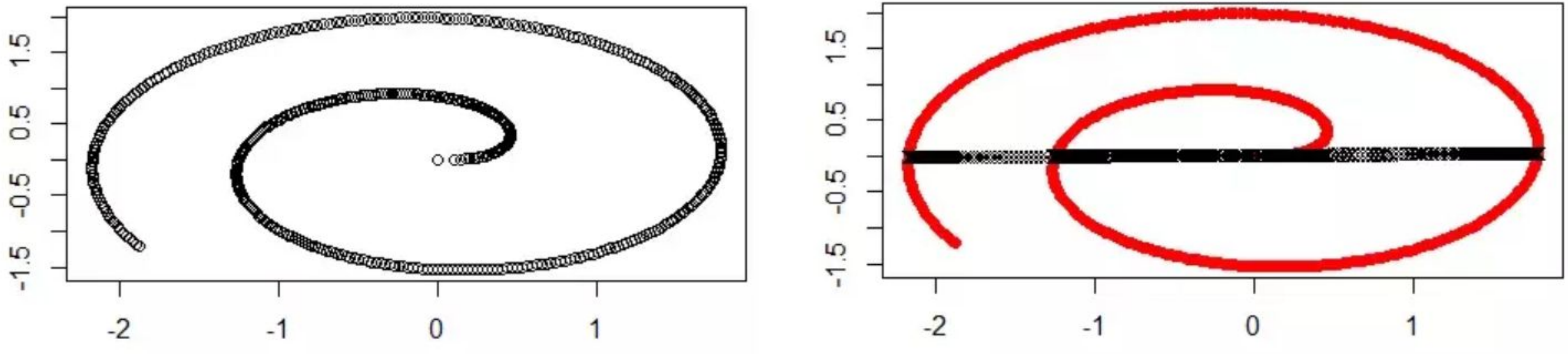
	K-Means	**PCA**
Goal	Find groups of similar points based on feature values.	Find n directions that contain as much variability in the data as possible.
Structure	The dataset is approximated by k centroids. Each point is associated with its nearest centroid.	The dataset is approximated by k orthogonal directions. Each point is associated with its projection.
Method	Find centroid locations such that we minimize the sum of the squares of the distances between each data point and its nearest centroid.	Find a direction that minimizes the sum of the squares of the projected distances of all points from their mean along this direction.

Going further

- Discuss/recap limitations of K-Means and PCA
- What else can we do?

When can PCA fail?

- PCA is works well when data is distributed along a line, flat plane, or higher dimensional equivalents. What happens when data is non distributed like this?

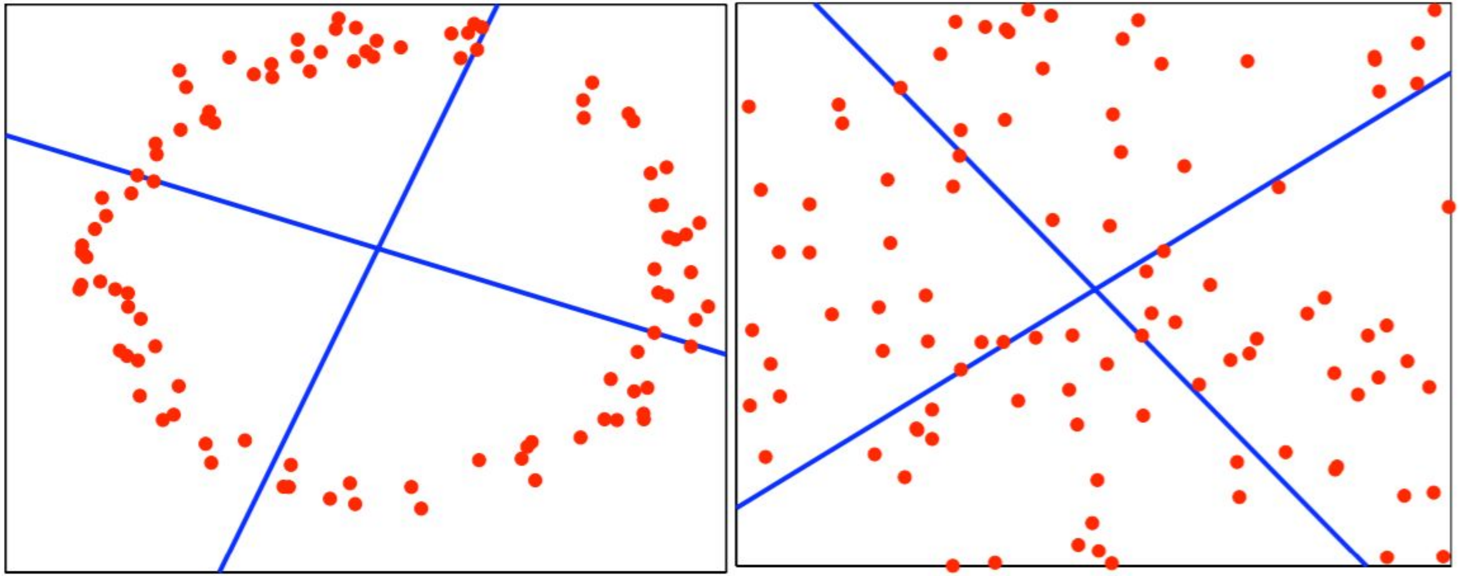


Here, which direction captures the maximum variability ?

Source

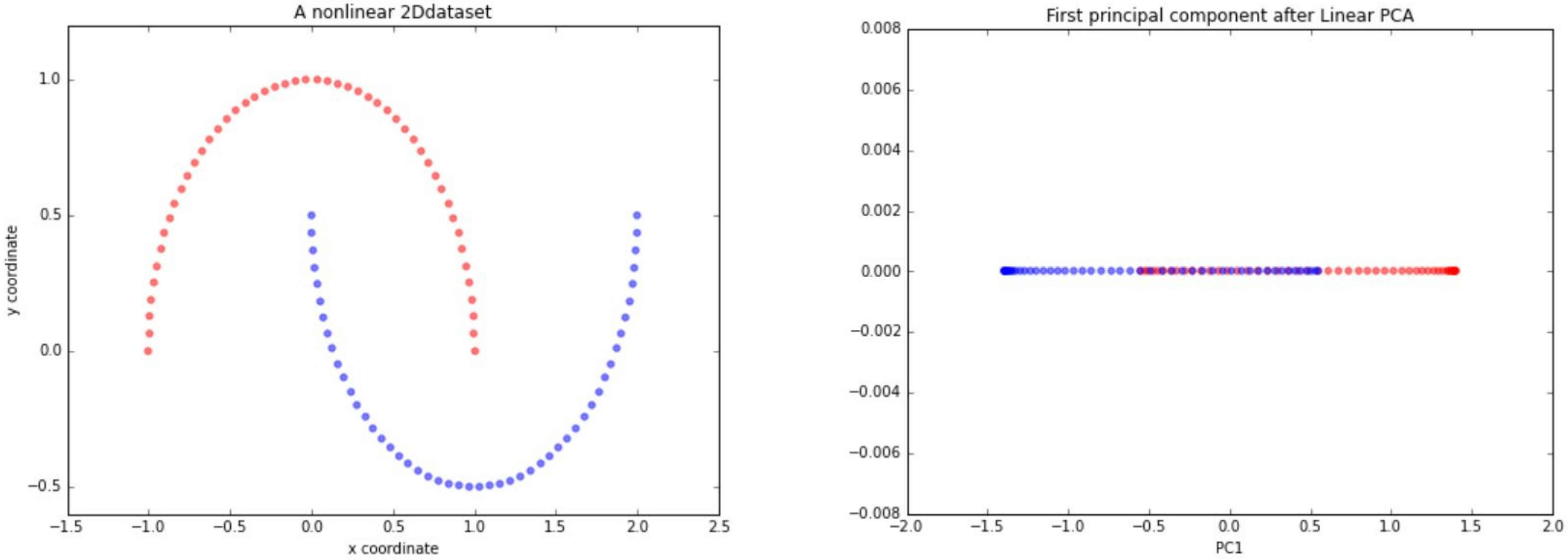
When can PCA fail?

- More examples where PCA is likely to fail



Source

When can PCA fail?

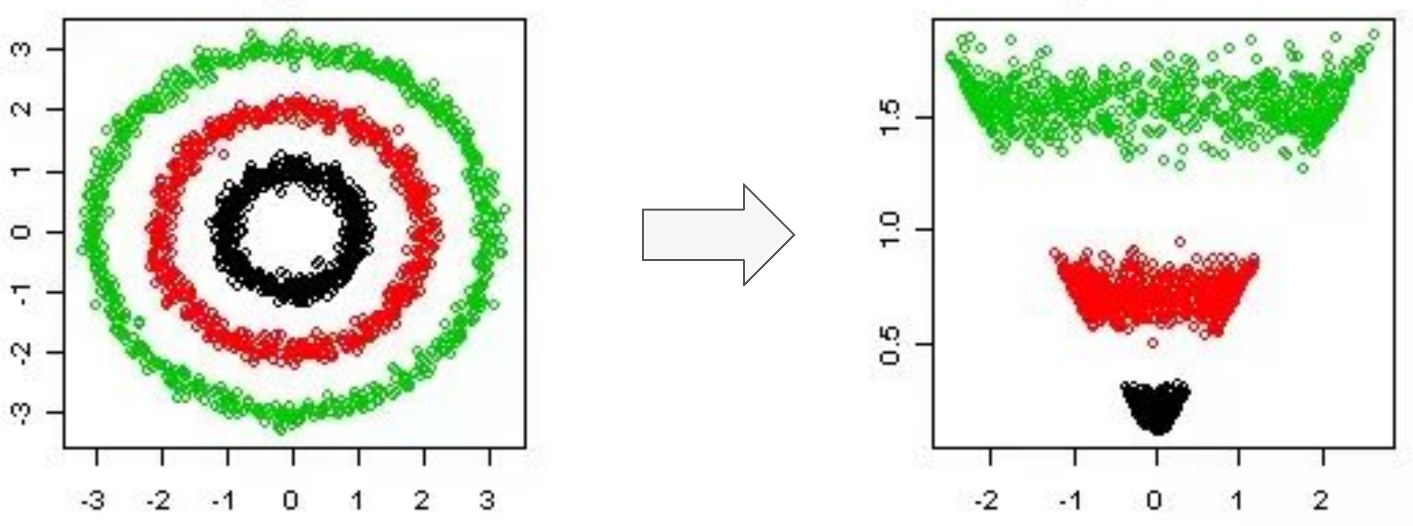


Reducing dimensions from 2D to 1D using Linear PCA

Source

Kernel PCA (one of many methods…)

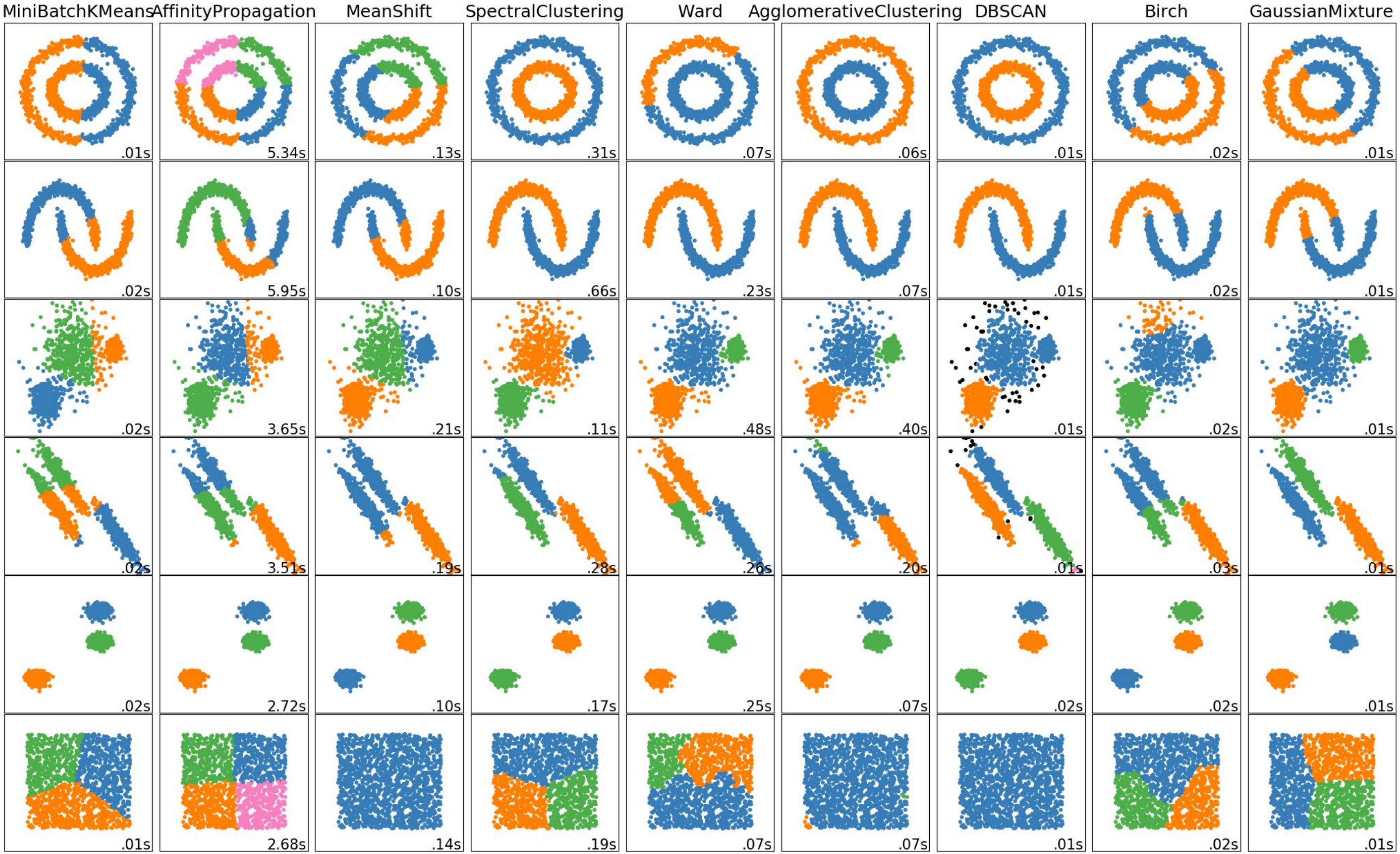
- **Idea:** augment dataset with new columns:
  [x, y] → [x, y, x^2, y^2, xy, …]



When can K-Means fail?

- K-Means works best when:
  - clusters are well-separated
  - clusters are of comparable size
  - clusters contain similar number of points
  - clusters are roughly round / spherical

What other clustering methods can we use?


MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
Birch
GaussianMixture
.01s
5.34s
.13s
.31s
.07s
.06s
.01s
.02s
.01s
.02s
5.95s
.10s
.66s
.23s
.07s
.01s
.02s
.01s
.02s
3.65s
.21s
.11s
.48s
.40s
.01s
.02s
.01s
.02s
3.51s
.19s
.28s
.26s
.20s
.01s
.03s
.01s
.02s
2.72s
.10s
.17s
.25s
.07s
.02s
.02s
.01s
.01s
2.68s
.14s
.19s
.07s
.07s
.01s
.02s
.01s

Source

Unsupervised learning

- **Unsupervised learning** is when our data consists of examples (rows) and features (columns). It is the broad task of describing how our data is organized. This includes:
    - Discovering groups of similar examples (e.g., clustering)
    - Reduce dimensionality for purpose of visualization (e.g., PCA)
    - Estimate the underlying distribution of data (density estimation)

Supervised learning

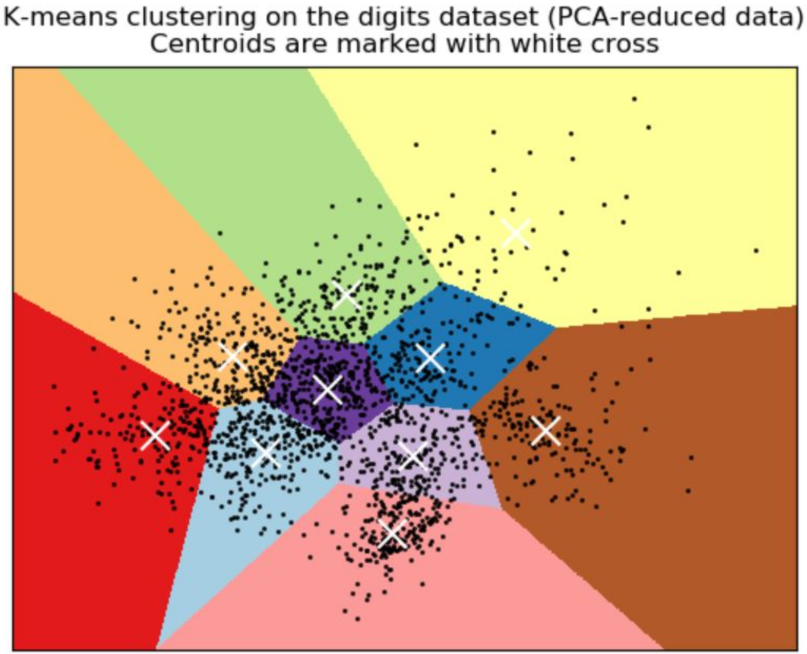
- **Supervised learning** is when our data consists of examples and features, as well as outcomes (labels) for each example.
    - Main application: Predict the labels of new unlabeled examples.
    - **Note:** The Iris and MNIST datasets had labels, but we never used them while doing clustering or PCA!
- We will see many examples of supervised learning in the second half of this course!

Applications of unsupervised learning

- Clustering
- Image segmentation
- Compression
- Denoising
- Anomaly detection
- Data generation

Clustering

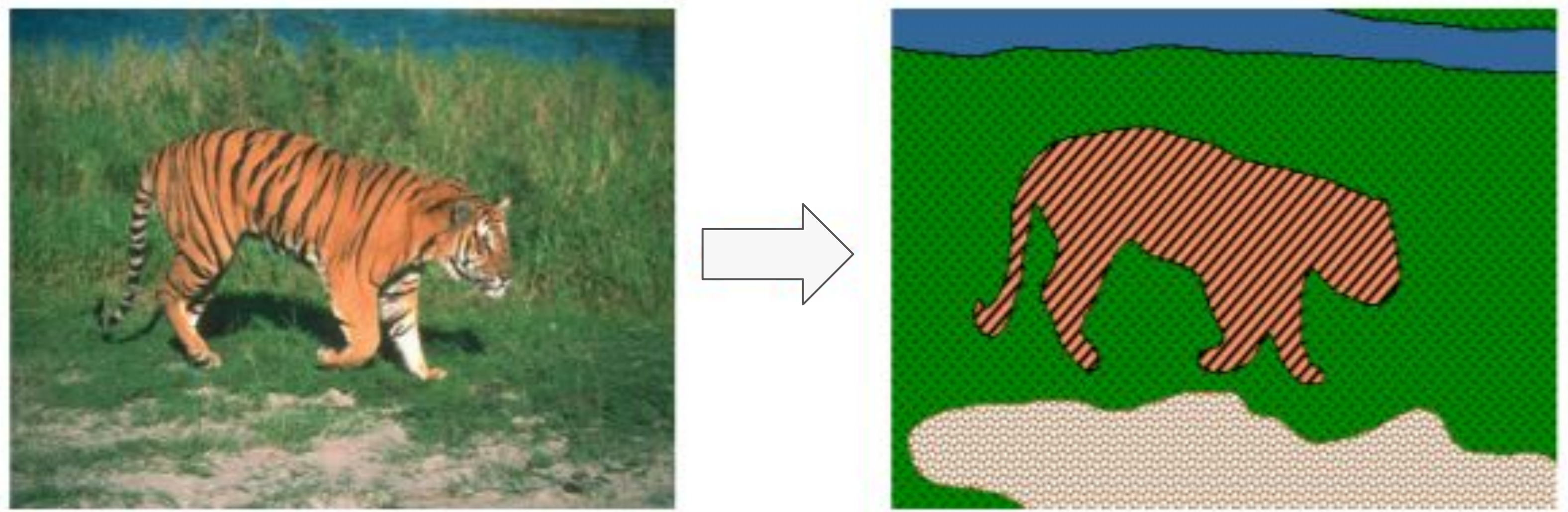
- Optical character recognition for converting written text to machine text.
- Segmenting customer base for targeted marketing, or to identify spending patterns.



Source

Image segmentation

Divide an image into its constituent components.



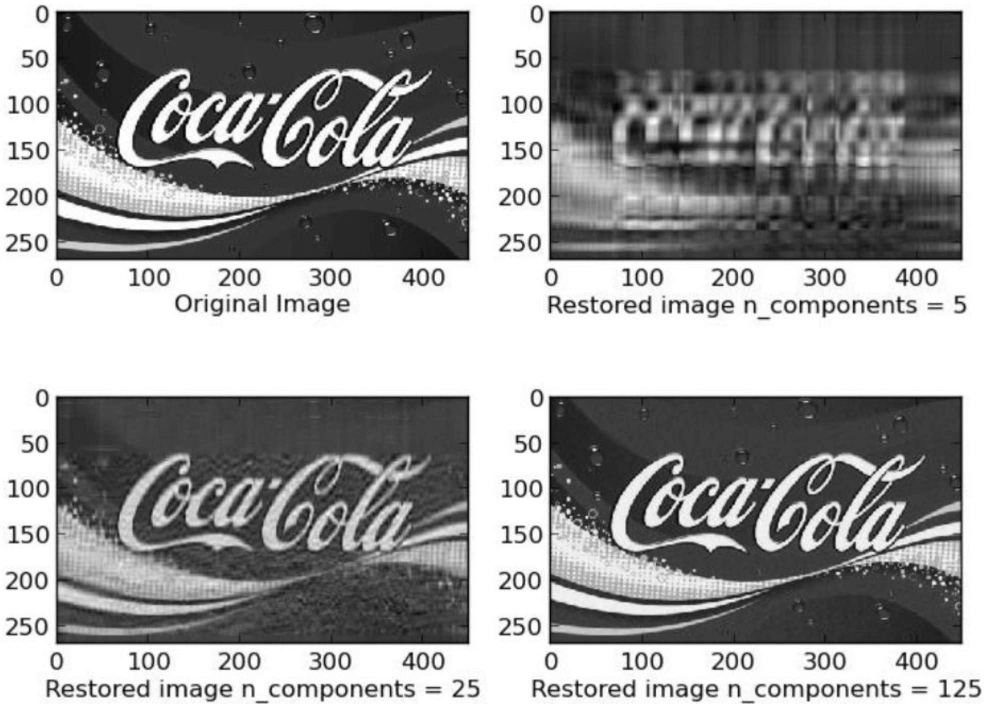
Example of **Image Segmentation**. Often used as a pre-processing step for detecting objects in an image.

Source

Compression

- PCA is one way to do this

Original image requires storing 450x270 values. →



Original Image

Restored image n_components = 5

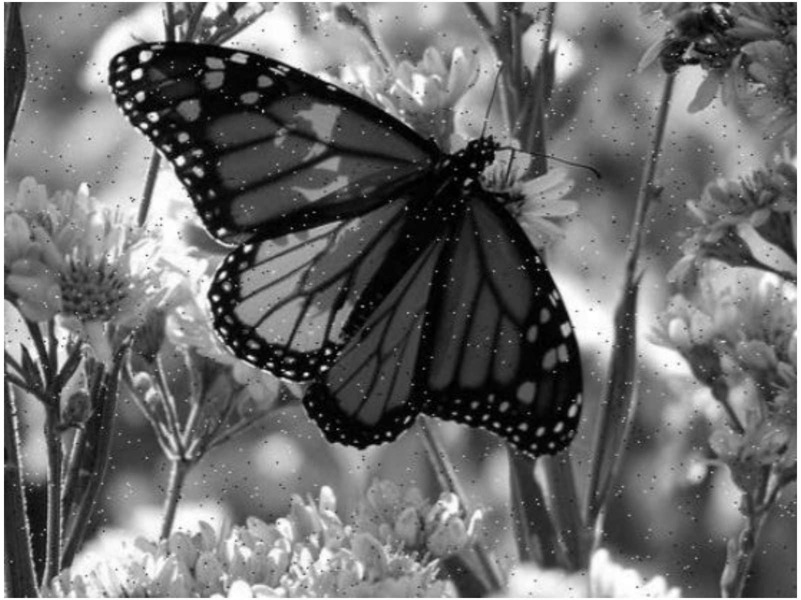
Restored image n_components = 25

Restored image n_components = 125

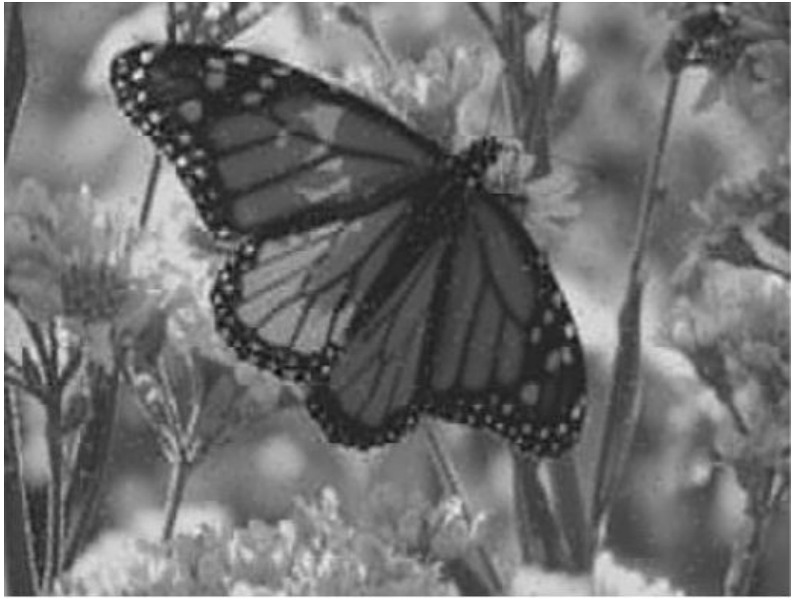
Source

Denoising

- PCA can also be used for this application!



Noisy Image

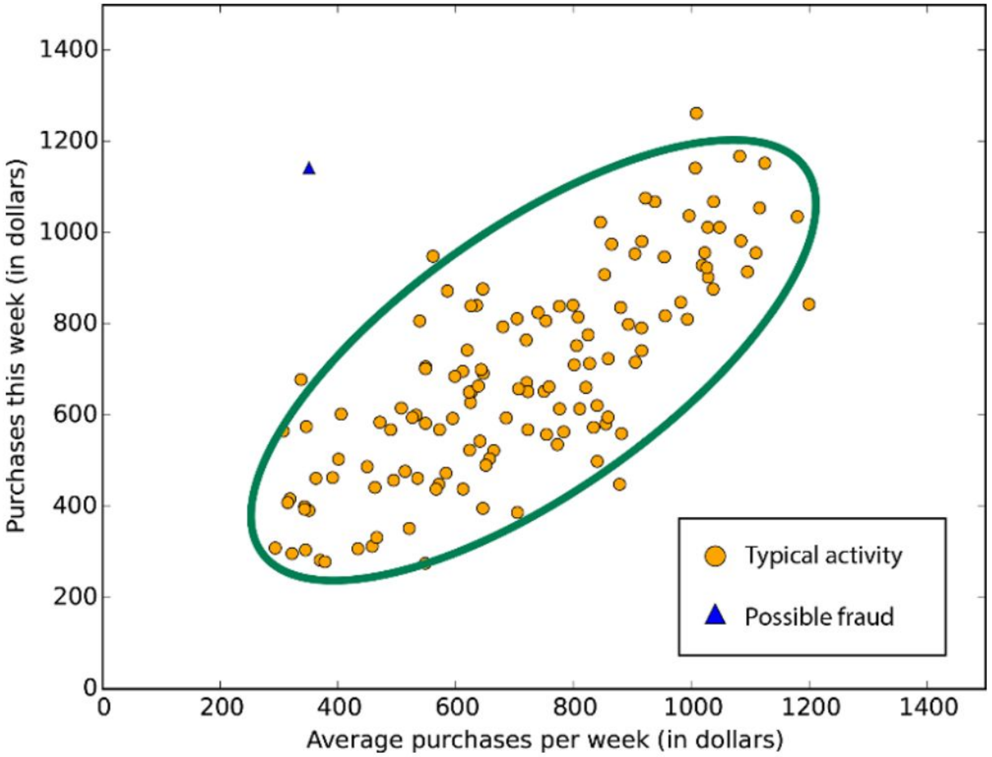


Denoised image using 15 PCA components

Source

Anomaly detection

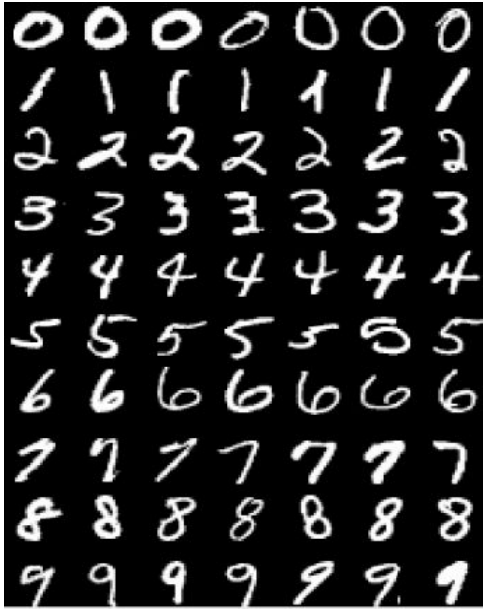
- Most points lie in a certain distribution. Points that deviate dramatically are flagged for review.



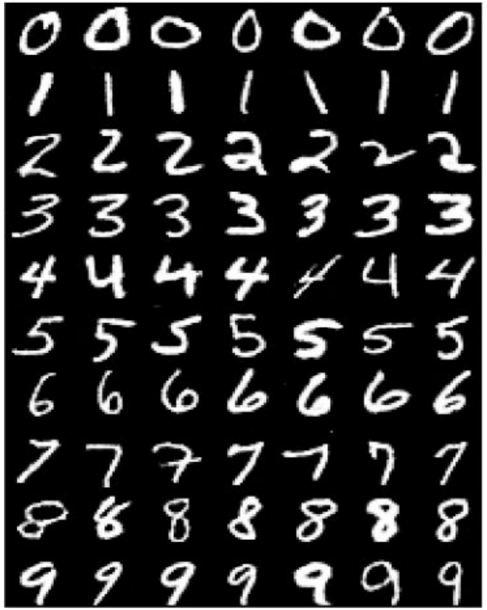
Source

Data generation

- **Generate** more data similar to given training data.



Ground Truth Handwritten Digits



Generated Data

Source