EPITECH.
PARIS GRADUATE SCHOOL OF DIGITAL INNOVATION

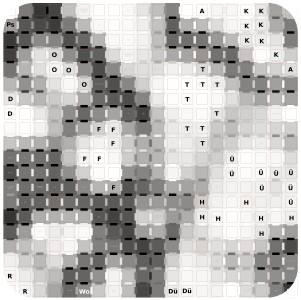
Recommender

Kick-off

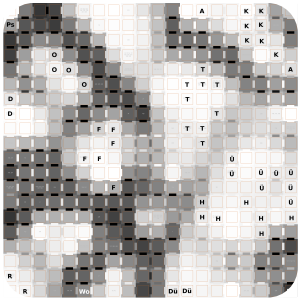
T9 - Big Data

T-DAT-901

Manipulating big data

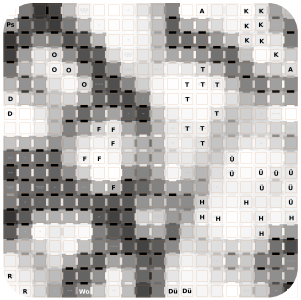


Manipulating big data



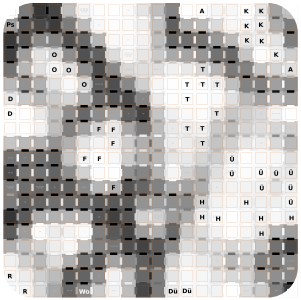
- Millions of lines

Manipulating big data



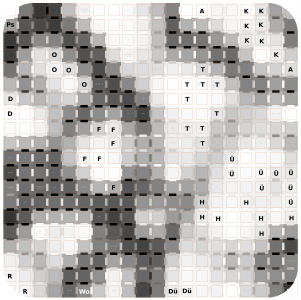
- Millions of lines
- High dimensional space

Manipulating big data



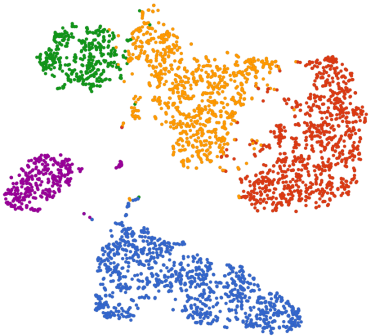
- Millions of lines
- High dimensional space
- Unclean and diverse sources of data

Manipulating big data

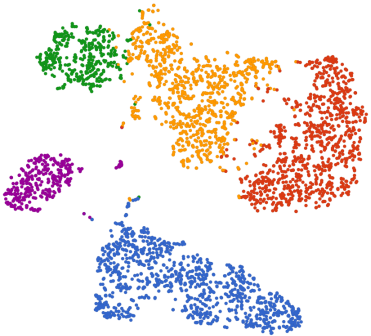


- Millions of lines
- High dimensional space
- Unclean and diverse sources of data
- Short time for response

Clustering

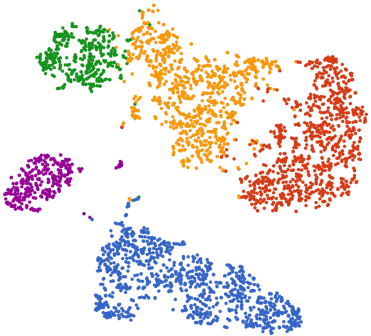


Clustering



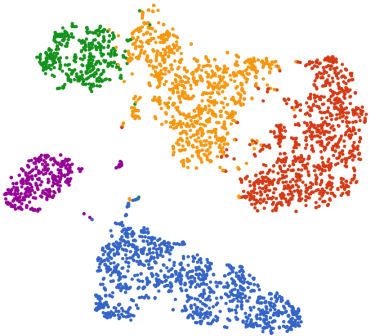
- Unsupervised learning

Clustering



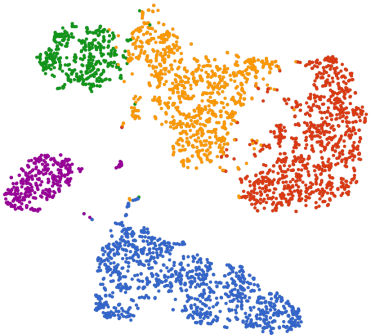
- Unsupervised learning
- Separate data into clusters

Clustering



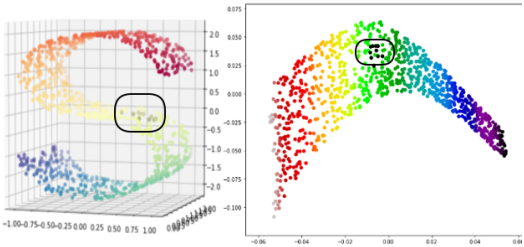
- Unsupervised learning
- Separate data into clusters
- Minimize dissimilarity inside a cluster

Clustering

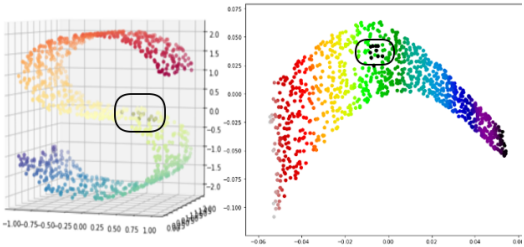


- Unsupervised learning
- Separate data into clusters
- Minimize dissimilarity inside a cluster
- Optimize shape and number of clusters

Dimension reduction

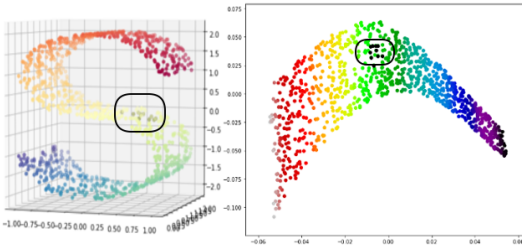


Dimension reduction



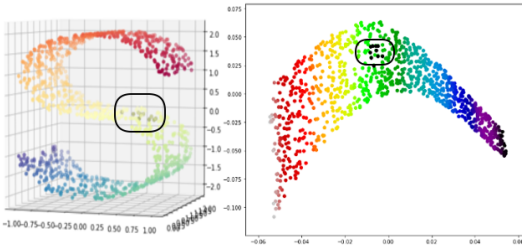
- From many dimensions to 2D

Dimension reduction



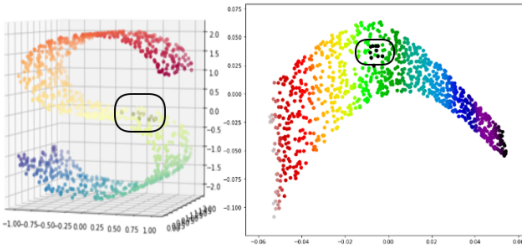
- From many dimensions to 2D
- Allow visualizations

Dimension reduction



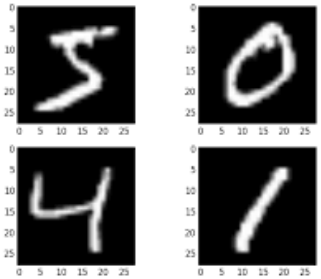
- From many dimensions to 2D
- Allow visualizations
- Fasten algorithms

Dimension reduction

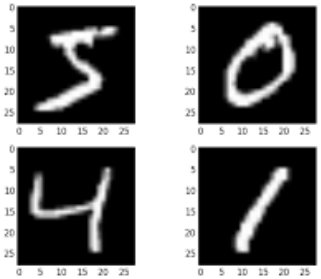


- From many dimensions to 2D
- Allow visualizations
- Fasten algorithms
- Control loss of information

Supervised learning

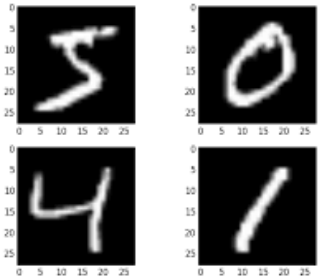


Supervised learning



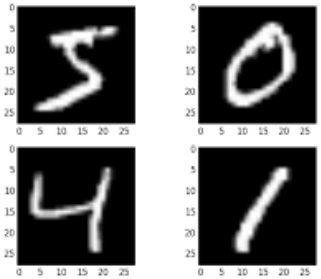
- Associate groups with outcomes

Supervised learning



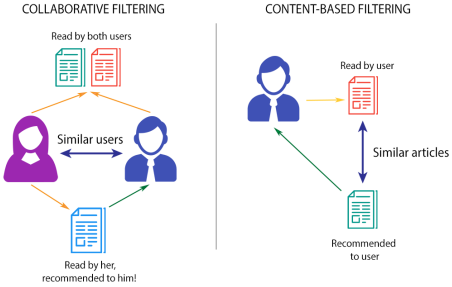
- Associate groups with outcomes
- Learn on data and predict

Supervised learning



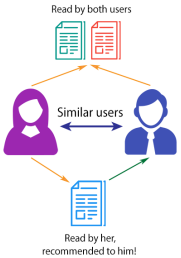
- Associate groups with outcomes
- Learn on data and predict
- Beware of complexity and robustness

Back to the project

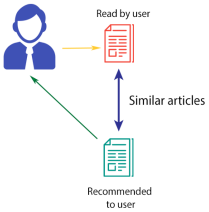


Back to the project

COLLABORATIVE FILTERING

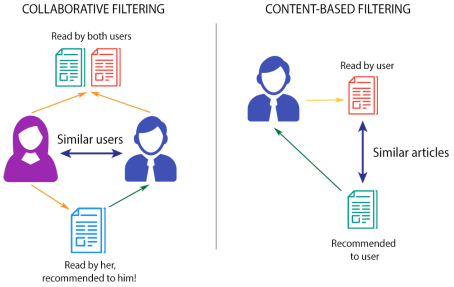


CONTENT-BASED FILTERING



Given a bunch of user behaviors:

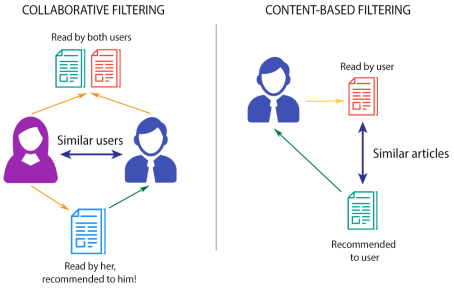
Back to the project



Given a bunch of user behaviors:

- Cluster users and/or products

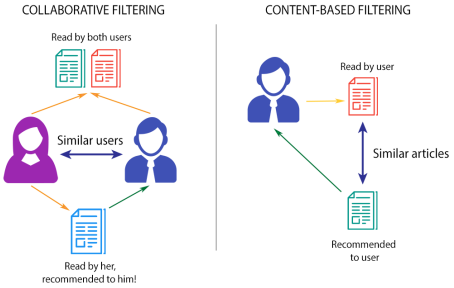
Back to the project



Given a bunch of user behaviors:

- Cluster users and/or products
- Identify user preferences

Back to the project



Given a bunch of user behaviors:

- Cluster users and/or products
- Identify user preferences
- Issue recommendations

Any questions

?