Unsupervised learning

Aprendizaje Automático para la Robótica
Máster Universitario en Ingeniería Industrial

Departamento de Automática

Objectives

1. TODO

Bibliography

- TODO Bishop, Christopher M. *Pattern Recognition and Machine Learning*. 2nd edition. Springer-Verlag. 2011
- TODO Müller, Andreas C., Guido, Sarah. *Introduction to Machine Learning with Python*. O'Reilly. 2016

Table of Contents

Clustering

K-means, agglomerative clustering, DBSCAN and GMM

Clustering

Applications

Set of unsupervised techniques that identify groups of data (named clusters)

- No universal definition of cluster: Centroid, medoid, dense regions, etc

Applications

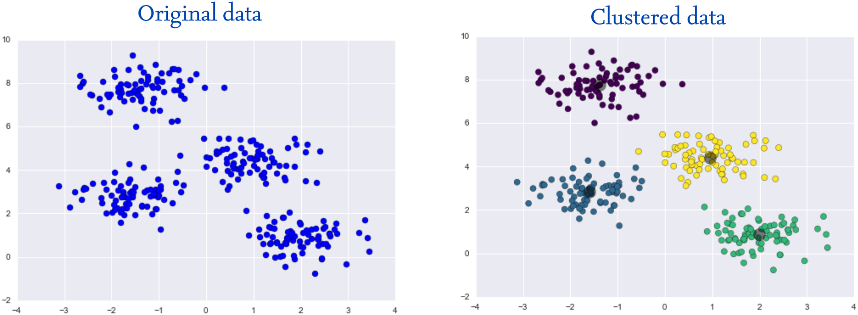
- Customer segmentation
- Data analysis
- Dimensionality reduction
- Anomaly detection
- Semi-supervised learning
- Search engines
- Image segmentation

Main algorithms

- K-means, DBScan, GMM, hierarchical clustering, EM, ...

K-means

Overview



(Source)

In k-means, clusters are identified by a centroid

K-means

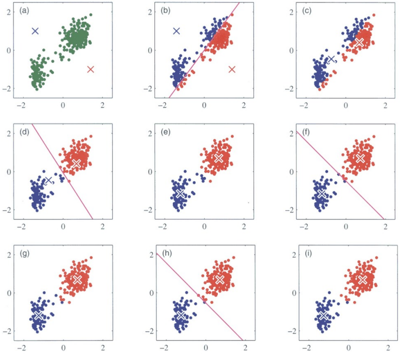
K-means algorithm (I)

K-means algorithm

1. Set k random centroids
2. Assign each data point to its closest centroid
3. Recompute centroids
4. Go to 2 until no point reassignment

k is an hyperparameter

- Number of clusters

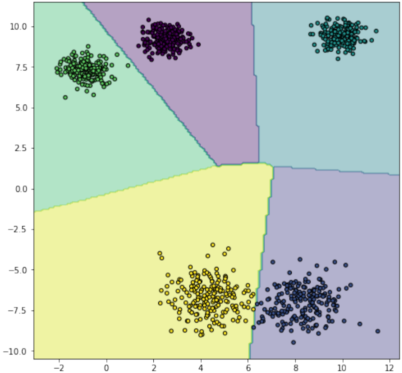


(Source)

K-means

K-means algorithm (II)

New data points are assigned to its closest centroid

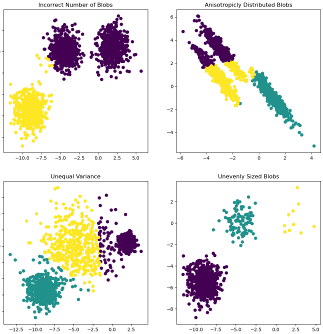


K-means

K-means limitations

K-means can fail in several conditions

- Incorrect number of clusters
- Different clusters variance
- Non-spheric clusters ⇒ normalization



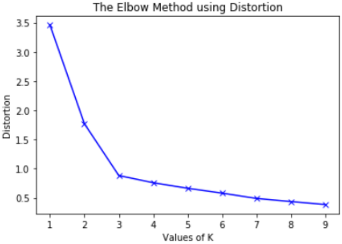
(Source)

K-means

Elbow's method

Election of k

- Not a problem when domain information is available
- ... that is rarely the case



Elbow's method

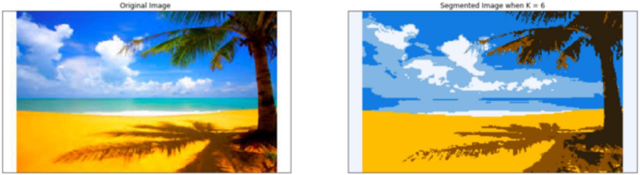
1. Select $K = 1, ..., n$
2. Visualize performance for each k
3. Choose K where metric stabilizes

Performance measures

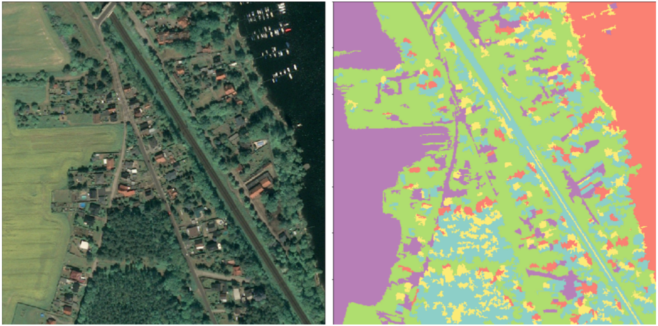
- Inertia: mean squared error between each instance and its closest centroid
- Silhouette: $(b - a)/max(a, b)$, where a mean intra-cluster distance, and b is the mean nearest-cluster distance

K-means

Application: Image segmentation



(Source)



(Source)

K-means

Application: Clustering for semi-supervised learning

Semi-supervised learning: Only a subset of the dataset is labeled

- Supervised and unsupervised learning
- Quite common in real-world applications (labels use to be expensive)

f_1	f_2	$\cdots$	f_n	y
$a_{1,1}$	$a_{2,1}$	$\cdots$	$a_{n,1}$	y_1
$a_{1,2}$	$a_{2,2}$	$\cdots$	$a_{n,2}$	
$a_{1,3}$	$a_{2,3}$	$\cdots$	$a_{n,3}$	
$a_{1,4}$	$a_{2,4}$	$\cdots$	$a_{n,4}$	y_4
$a_{1,5}$	$a_{2,5}$	$\cdots$	$a_{n,5}$	

Label propagation

1. Obtain k clusters
2. Get a representative instance of each cluster (medoid) measuring the distance to the centroid
3. Label the members of each cluster with its medoid's label

Clustering

K-means: Scikit-learn

TODO: SCikit-Learn

K-means

K-means: Summary

Hyperparameters	Advantages	Disadvantages
k	Fast	Simple shapes
	Few hyperparameters	Determine k
	Scalable	Random initialization

Other clustering algorithms

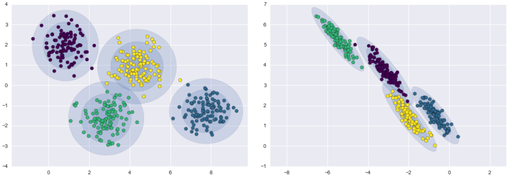
Gaussian Mixure Model (GMM) (I)

GMM is a generative clustering algorithm

- Assumes data coming from a set of multidimensional gaussian distributions

GMM fits a set $\{(\phi_i, \mu_i, \sigma_i)\}_{i=1,\ldots,k}$

- ϕ is a weight
- μ is a multidimensional mean
- σ is a covariance matrix
- k is the number of clusters (hyperparameter)



(Source)

Other clustering algorithms

Gaussian Mixure Model (GMM) (II)

Gaussian parameters are fit with the Expectation-Maximization (E-M) algorithm

- E-M is a generalization of K-means

Expectation-Maximization algorithm

1. Init parameters randomly
2. Expectation step: Assign each instance to a cluster
   - Assignment is probabilistic
3. Maximization step: Update cluster parameters
   - Each cluster is updated using *all* the data
   - Instances contribution to a cluster parameters is weighted by the probability that it belongs to it
4. Go to 2

GMM can be seen as a fuzzy clustering algorithm

Other clustering algorithms

Gaussian Mixure Model (GMM) (III)

Gaussian parameters are fit with the Expectation-Maximization (E-M) algorithm

- E-M is a generalization of K-means

Expectation-Maximization algorithm

1. Init parameters randomly
2. Expectation step: Assign each instance to a cluster
   - Assignment is probabilistic
3. Maximization step: Update cluster parameters
   - Each cluster is updated using *all* the data
   - Instances contribution to a cluster parameters is weighted by the probability that it belongs to it
4. Go to 2

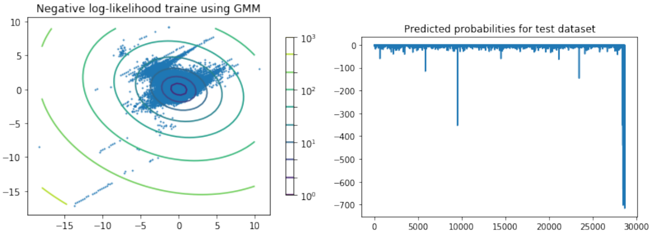
GMM can be seen as a fuzzy clustering algorithm

Other clustering algorithms

Gaussian Mixure Model (GMM) (IV)

GMM provides a probability of an instance to belong to a cluster

- This can be used to detect anomalies
- Just assign a probability threshold



(Source)

Other clustering algorithms

DBSCAN: Scikit-learn

TODO: Scikit-Learn

Other clustering algorithms

GMM: Summary

Hyperparameters	Advantages	Disadvantages
Number of clusters	Probabilistic clustering	Number of clusters
Covariance matrix type	Generative model	Gaussian data
	Anomaly detection	Sensitive to outliers

Other clustering algorithms

DBSCAN (I)

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

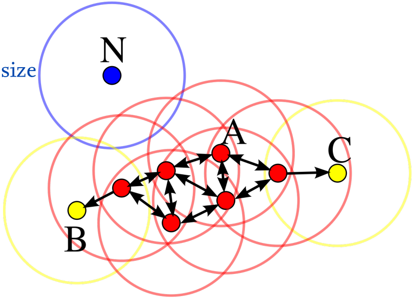
- Identifies high density regions (dense regions) in feature space
- Asumtion: Clusters form dense regions separated by empty areas

Hyperparameters

- ϵ: Radius of a neighborhood
- *min_samples*: Minumun cluster size

Type of points

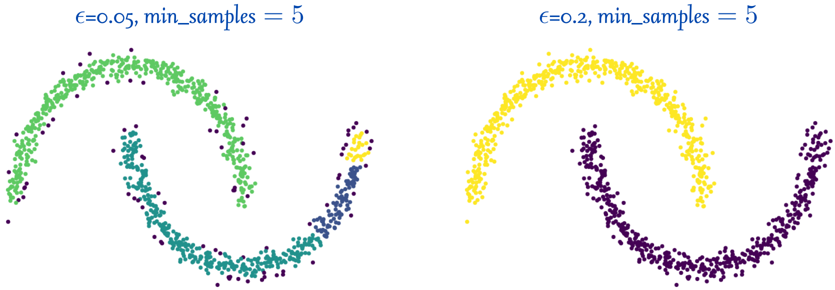
- Core instance
- Outliers



(Source)

Other clustering algorithms

DBSCAN (II)



(Source)

Other clustering algorithms

DBSCAN: Scikit-learn

TODO: Scikit-Learn

Other clustering algorithms

DBSCAN: Summary

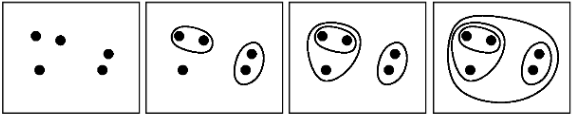
Hyperparameters	Advantages	Disadvantages
ϵ	No explicit number of clusters	Slower than K-means
$min\_samples$	Scales relatively well	Clusters with different densities
	Almost deterministic	
	Robust to outliers	
	Anomaly detection	

Other clustering algorithms

Agglomerative clustering (I)

Agglomerative clustering

1. Initially, each instance forms a cluster
2. Merge the two most similar clusters according to a metric
3. Repeat 2 until a stop criterion is satisfied



We need a similarity measure between two clusters

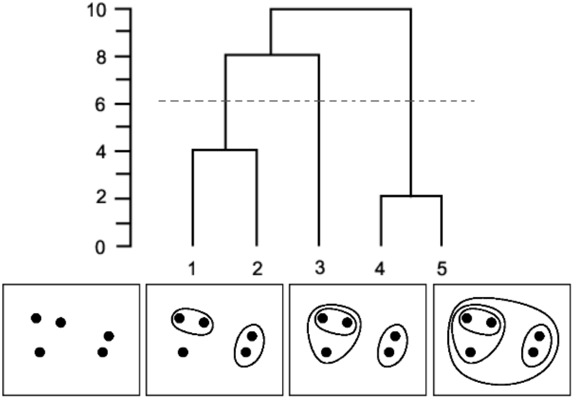
- *Ward*: Minimizes variance within merged clusters. Leads to equally sized clusters
- *Average*: Minimizes average distances between their points
- *Complete*: Minimizes maximun distance between their points

Other clustering algorithms

Agglomerative clustering (II)

Agglomerative clustering is a special case of hierarchical clustering

Dendrogram



(Source)

Other clustering algorithms

Agglomerative clustering: Scikit-Learn

TODO: SCikit-Learn

Other clustering algorithms

Agglomerative clustering: Summary

Hyperparameters	Advantages	Disadvantages
	Complex shapes	
	Hierarchical clustering	

Anomaly detection

TODO

Dimensionality reduction

PCA and manifold learning

Dimensionality reduction

Main approaches for dimensionality reduction

Dimensionality reduction transforms data into more convenient representations

- Reduce data dimensionality
- Visualize multidimensional data

Projection

- TODO

Manyfold learning

- TODO

Dimensionality reduction

Principal Components Analysis (I)

Dimensionality reduction transforms data into more convenient representations

- Reduce data dimensionality
- Visualize multidimensional data

Main algorithms

- Isomap
- T-distributed Stochastic Neighbor Embedding (t-SNE)
- Principal Components Analysis (PCA)

Dimensionality reduction

Principal Components Analysis (I)

Dimensionality reduction transforms data into more convenient representations

- Reduce data dimensionality
- Visualize multidimensional data

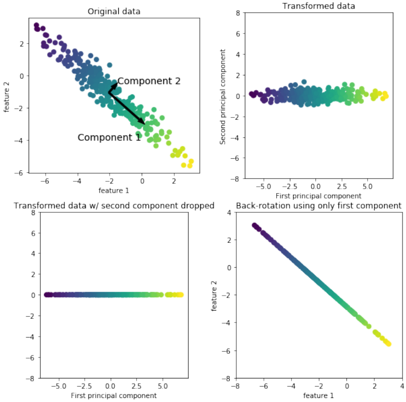
Main algorithms

- Isomap
- T-distributed Stochastic Neighbor Embedding (t-SNE)
- Principal Components Analysis (PCA)

Dimensionality reduction

Principal Components Analysis (II)

PCA maximizes data variance



(Source)

Dimensionality reduction

Principal Components Analysis (III)

Example: Hand-written digits recognition

- Images of hand-written digits
- 8x8 images (64 dimensions)
- 10 digits
- Classification problem