Machine Learning - Session 2

Unsupervised Learning

Content

1. Introduction
2. Clustering
3. Association Rules
4. Hands - On

Introduction

- Unsupervised learning includes all kinds of machine learning where there is no known output, no teacher to instruct the learning algorithm.

Introduction

- Unsupervised learning includes all kinds of machine learning where there is no known output, no teacher to instruct the learning algorithm.

- In unsupervised learning, the learning algorithm is just shown the input data and asked to extract knowledge from this data.

Introduction

- Unsupervised learning includes all kinds of machine learning where there is no known output, no teacher to instruct the learning algorithm.

- In unsupervised learning, the learning algorithm is just shown the input data and asked to extract knowledge from this data.

- In contrast with supervised learning, in unsupervised learning, there is no direct measure of success.

Introduction

Some tasks include

- Clustering
- Anomaly Detection
- Density estimation
- Dimensionality Reduction
- Association Rules

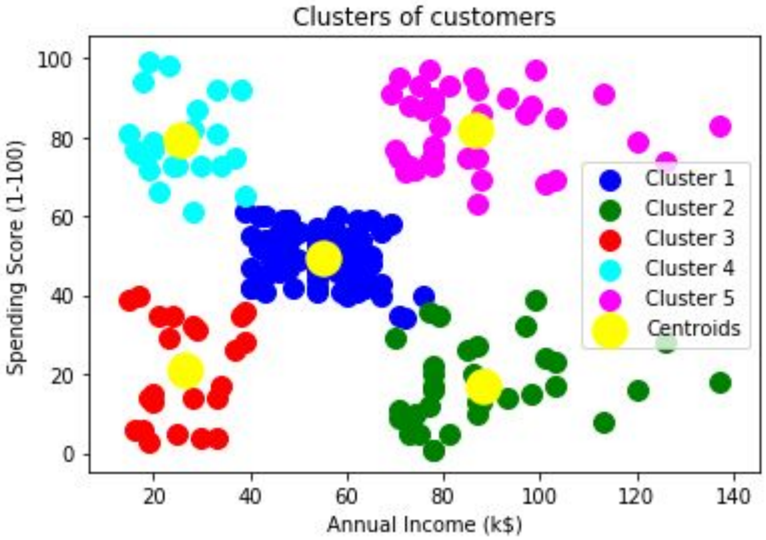
Introduction

Clustering:

- The goal is to group similar instances together into clusters.
- Clustering is a great tool for data analysis, customer segmentation, recommender systems, search engines, image segmentation, semi-supervised learning, dimensionality reduction, and more.

Introduction

Clustering



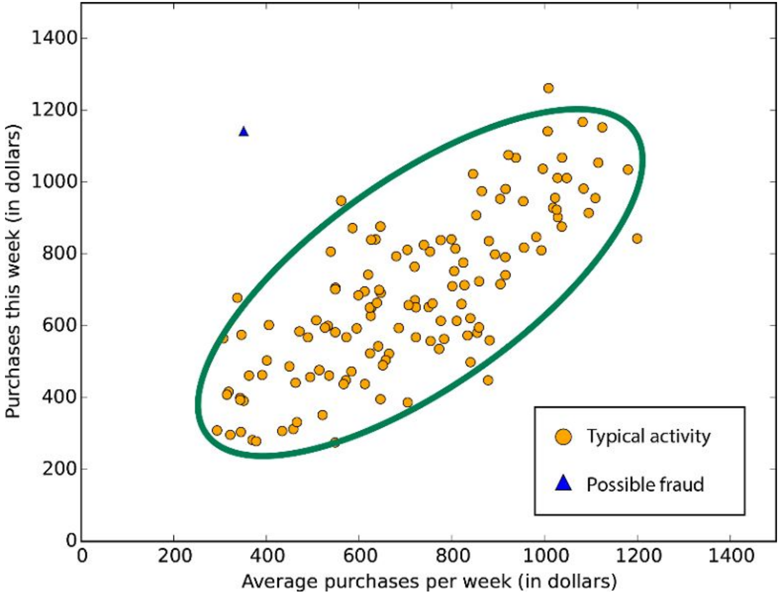
Introduction

Anomaly Detection:

- The objective is to learn what “normal” data looks like, and then use that to detect abnormal instances, such as defective items on a production line or a new trend in a time series.

Introduction

Anomaly Detection



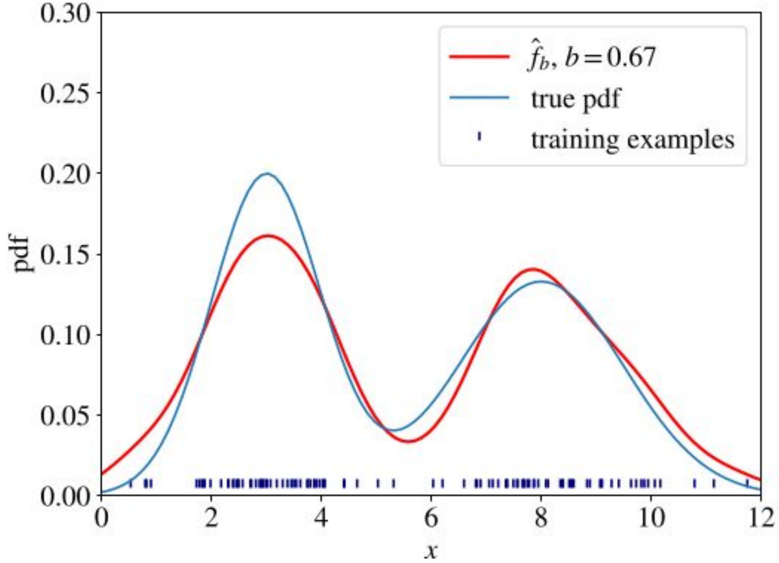
Introduction

Density estimation:

- This is the task of estimating the probability density function (PDF) of the random process that generated the dataset.
- Density estimation is commonly used for anomaly detection: instances located in very low-density regions are likely to be anomalies.
- It is also useful for data analysis and visualization.

Introduction

Density estimation



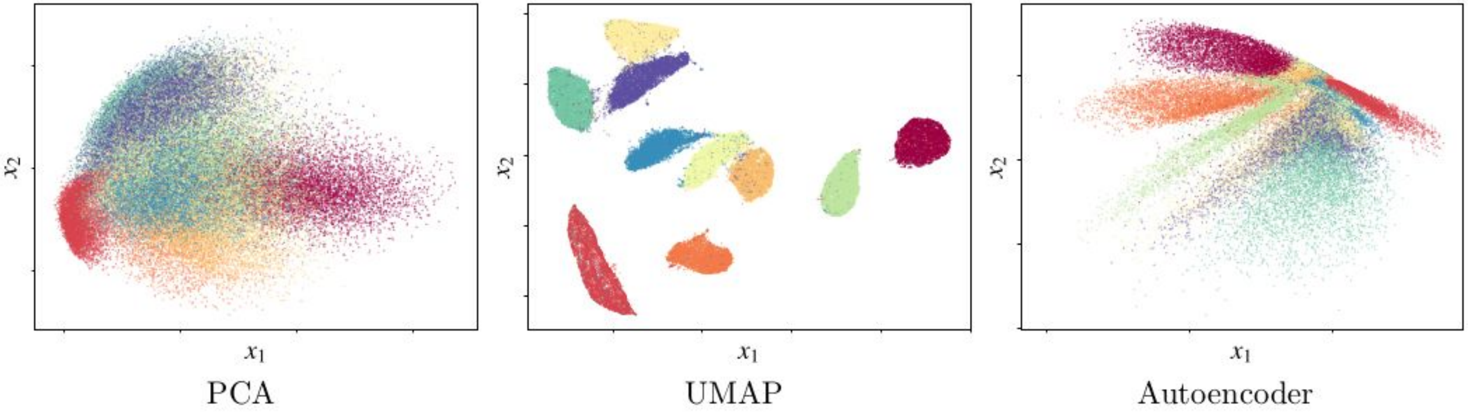
Introduction

Dimensionality Reduction:

- Considers Latent Variables that explain observed variables.
- Decomposition techniques.
- Linear, nonlinear, parametric and non parametric.

Introduction

Dimensionality Reduction



PCA

UMAP

Autoencoder

Figure 7: Dimensionality reduction of the MNIST dataset using three different techniques.

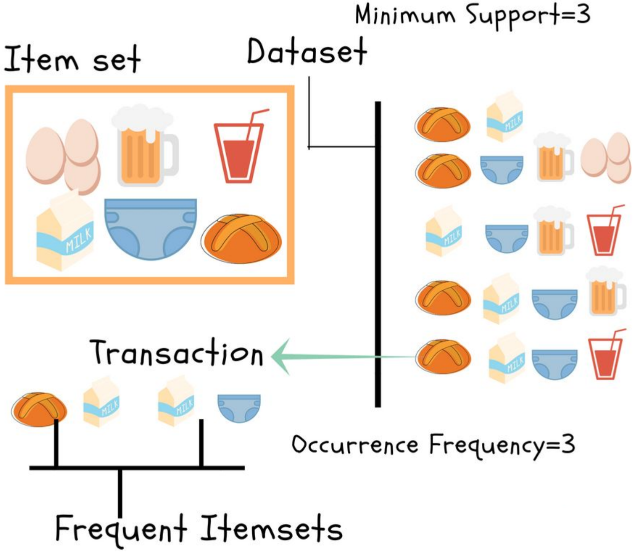
Introduction

Association Rules:

- Association rule analysis has emerged as a popular tool for mining commercial databases.
- It is most often applied to binary-valued data where it is referred to as "market basket" analysis.

Introduction

Association Rules:



Clustering

K-means

Defines an objective function J, sometimes called a distortion measure, that computes the sum of the squares of the distances of each data point to assigned vector of centroids μ.

The goal is to find values for the clusters assignations r and the μ so as to minimize the objective function.

Clustering

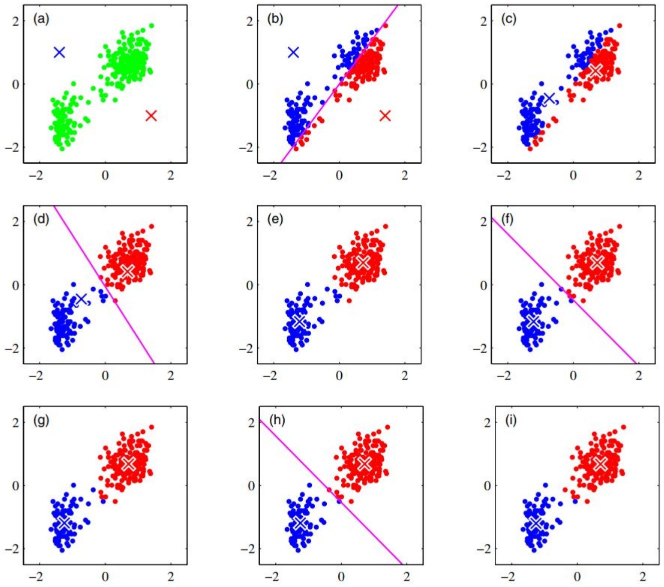
K-means

Steps:

- First we choose some initial values for the μ.
- Then in the first phase we minimize J with respect to the r, keeping the μ fixed.
- In the second phase we minimize J with respect to the μ , keeping r fixed.
- This two-stage optimization is then repeated until convergence.

Clustering

K-means iteration illustration



Clustering

Gaussian Mixtures

- A Gaussian mixture model (GMM) is a probabilistic model that assumes that the instances were generated from a mixture of several Gaussian distributions whose parameters are unknown.
- Each cluster can have a different ellipsoidal shape, size, density, and orientation

ŏzaru

Clustering

Gaussian Mixtures

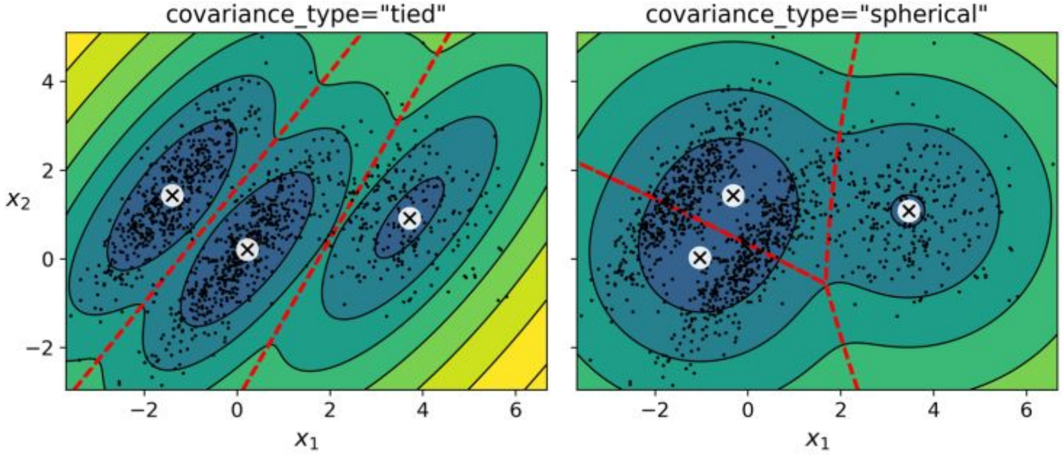


Figure 9-18. Gaussian mixtures for tied clusters (left) and spherical clusters (right)

Clustering

DBSCAN (density-based spatial clustering of applications with noise)

The idea behind DBSCAN is that clusters form dense regions of data, separated by regions that are relatively empty. Also:

- It does not require the user to set the number of clusters a priori
- It can capture clusters of complex shapes.
- It can identify points that are not part of any cluster.
- Works by identifying points that are in “crowded” regions of the feature space, where many data points are close together.

Clustering

Others:

- Spectral Clustering
- HDBSCAN
- Agglomerative Clustering
- BIRCH

Clustering

Evaluation

- Because the dataset is completely unlabeled, deciding on whether the learned model is optimal is much more complicated than in supervised learning.

Clustering

Evaluation

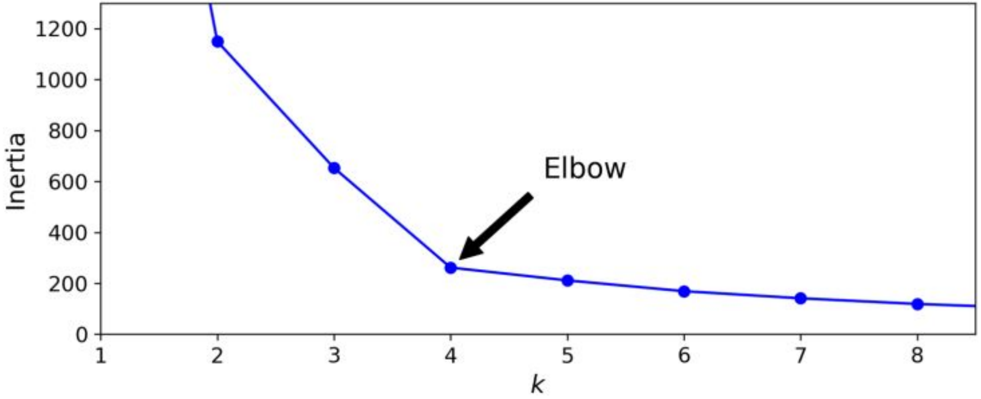
K-means:

- For evaluating the quality of clustering, we need to use intrinsic metrics—such as the within-cluster SSE (distortion)—to compare the performance of different k-means clusterings.

Clustering

Evaluation

- The so-called elbow estimates the optimal number of clusters, k, for a given task.
- The idea behind the elbow method is to identify the value of k where the distortion begins to increase most rapidly



Clustering

Evaluation

Gaussian Mixtures:

- In practice, the number of clusters is obtained with criterions from the information theory: AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion)
- Both the BIC and the AIC penalize models that have more parameters to learn (e.g., more clusters) and reward models that fit the data well. They often end up selecting the same model.

Clustering

Evaluation

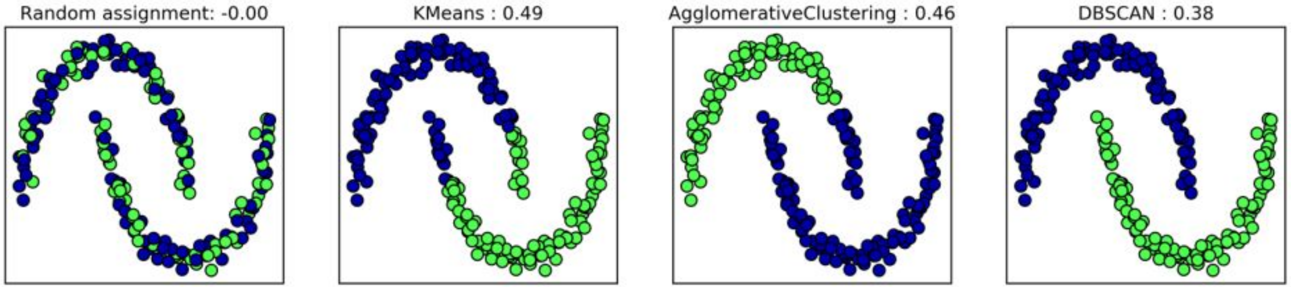
Others:

- For k-means and other methods, the silhouette score is frequently used.
- It computes the compactness of a cluster, where higher is better, with a perfect score of 1.
- However, this often don’t work well in practice. While compact clusters are good, compactness doesn’t allow for complex shapes.

Clustering

Evaluation

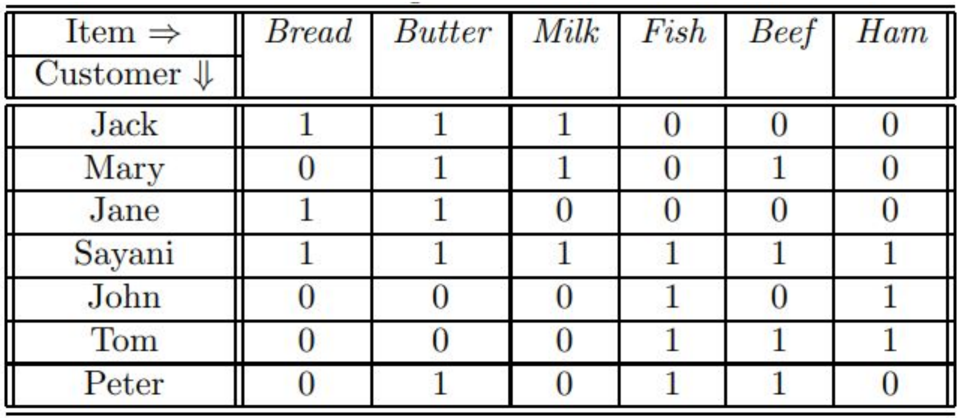
Silhouette scores:



Association Rules

Association rules attempt to construct simple rules that describe regions of high density in the special case of very high dimensional binary-valued data.

For example, suppose a data set of users and purchased items:



Item ⇒ Customer ⇓	*Bread*	*Butter*	*Milk*	*Fish*	*Beef*	*Ham*
Jack	1	1	1	0	0	0
Mary	0	1	1	0	1	0
Jane	1	1	0	0	0	0
Sayani	1	1	1	1	1	1
John	0	0	0	1	0	1
Tom	0	0	0	1	1	1
Peter	0	1	0	1	1	0

Association Rules

- The goal is to find joint values of the item variables that appear most frequently in the database.
- This is achieved obtaining sets of items that are closely correlated in the transaction database and with the notions of support.

The support of an itemset X is the fraction of transactions in T, of which X is a subset.

In the database example, the sets {Bread, Butter, Milk}, and {Fish, Beef, Ham} have the largest support.

Association Rules

- An association rule is denoted in the form X ⇒ Y , where the “⇒” is intended to give a direction to the nature of the correlation between the set of items X and Y.

  For example, the rule that relates {Butter, Milk}⇒{Bread}

Association Rules

- An association rule is denoted in the form X ⇒ Y , where the “⇒” is intended to give a direction to the nature of the correlation between the set of items X and Y.

  For example, the rule that relates {Butter, Milk}⇒{Bread}

- The strength of such a rule is measured by its confidence:

  The confidence of the rule X ⇒ Y is the conditional probability that a transaction in T contains Y, given that it also contains X.

Association Rules

A priori algorithm

A classical algorithm for association rules is the a priori algorithm that considers a threshold for the support. In addition, the lift metric could be employed:

- The lift metric measures how much more often the antecedent X and consequent of Y a rule X ⇒ Y occur together than we would expect if they were statistically independent. It is 1 if X and Y are independent.

Hands-On