1. Machine learning. The scheme and an example in details

Machine learning (ML) is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computers to learn from data and improve their performance on a specific task without being explicitly programmed. Instead of relying on explicit programming, machine learning systems use statistical techniques to identify patterns in data and make predictions or decisions.

There are various types of machine learning approaches, including supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, the algorithm is trained on labeled data, where the input data is paired with corresponding output labels. Unsupervised learning involves working with unlabeled data, and the algorithm learns to identify patterns or structures in the data on its own. Reinforcement learning involves training a model to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties.

Machine learning finds applications in a wide range of domains, including image and speech recognition, natural language processing, recommendation systems, autonomous vehicles, and many more. The field continues to evolve, driven by advancements in algorithms, data availability, and computational power, leading to improved capabilities and applications.

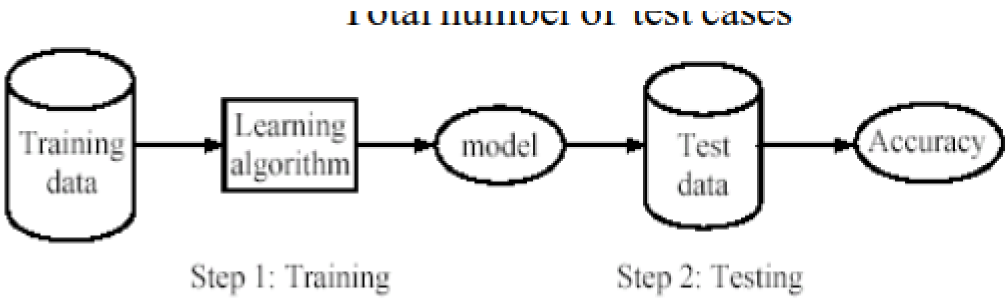
Types of Learning • Supervised (inductive) learning – Training data includes desired outputs • Unsupervised learning – Training data does not include desired outputs • Semi-supervised learning – Training data includes a few desired outputs • Reinforcement learning – Rewards from sequence of actions

2. Supervised Learning. The scheme and an example in details

Supervised learning is a type of machine learning paradigm where the algorithm is trained on a labeled dataset, which means that each input data point is associated with a corresponding target or output label. The goal of supervised learning is to learn a mapping from inputs to outputs, allowing the algorithm to make predictions or classifications on new, unseen data.

During the training phase, the algorithm is exposed to a set of input-output pairs, and it adjusts its internal parameters to minimize the difference between its predictions and the actual target labels. This process involves iteratively refining the model's performance until it can generalize well to new, unseen data.

Supervised learning is commonly used for tasks such as classification and regression. In classification, the algorithm learns to assign input data points to predefined categories, while in regression, the goal is to predict a continuous output. Examples of supervised learning applications include image recognition, speech recognition, and predicting house prices.



An **example** • An emergency room in a hospital measures 17 variables (e.g., blood pressure, age, etc) of newly admitted patients. • A decision is needed: whether to put a new patient in an intensive-care unit. • Due to the high cost of ICU, those patients who may survive less than a month are given higher priority. • Problem: to predict high-risk patients and discriminate them from low-risk patients.

3. Unsupervised Learning. The scheme and an example in details

Unsupervised learning is a machine learning approach where the algorithm is provided with input data without explicit output labels. The goal of unsupervised learning is to identify patterns, structures, or relationships within the data without guidance on what to look for. Unlike supervised learning, there is no predefined correct answer during the training phase.

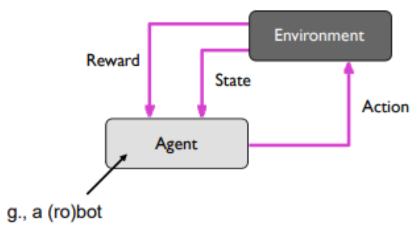
Clustering and dimensionality reduction are common tasks in unsupervised learning. In clustering, the algorithm groups similar data points together based on inherent patterns, while dimensionality reduction aims to simplify the dataset by capturing its essential features and reducing the number of variables.

Unsupervised learning is valuable for exploring and uncovering hidden patterns in data, and it is used in various applications such as customer segmentation, anomaly detection, and data compression. Examples of unsupervised learning algorithms include k-means clustering, hierarchical clustering, and principal component analysis (PCA).

Example: Let's say that a customer goes to a supermarket and buys bread, milk, fruits, and wheat. Another customer comes and buys bread, milk, rice, and butter. Now, when another customer comes, it is highly likely that if he buys bread, he will buy milk too. Hence, a relationship is established based on customer behavior and recommendations are made.

4. Reinforcement Learning. The scheme and an example in details

Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with an environment. The agent receives feedback in the form of rewards or punishments based on the actions it takes. The goal of



reinforcement learning is to find the optimal strategy or policy that maximizes the cumulative reward over time.

Here's a brief overview of the reinforcement learning process:

1. **Agent:** The entity making decisions and taking actions within an environment.
2. **Environment:** The external system with which the agent interacts.
3. **State:** A specific situation or configuration of the environment.
4. **Action:** The decision or move made by the agent.
5. **Reward:** A numerical feedback signal indicating the immediate benefit or cost of the action taken by the agent.
6. **Policy:** The strategy or set of rules that the agent follows to make decisions.

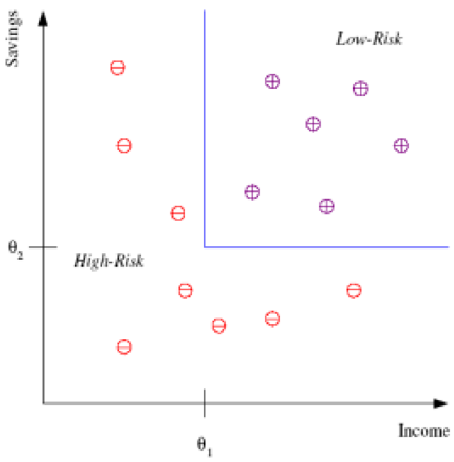
Example: Consider training a computer program to play a game, such as chess. The agent (the program) plays moves against the environment (the game board). After each move, the agent receives a reward or punishment based on the outcome (winning, losing, or drawing). Over time, the agent learns the best strategy (policy) to make optimal moves and increase its chances of winning. Reinforcement learning is widely applied in areas like game playing, robotics, and autonomous systems.

5. Ordinal Regression. The scheme and an example in details

Ordinal regression is a type of statistical and machine learning method used when the target variable is ordinal, meaning it has a meaningful order but the intervals between the categories are not necessarily uniform. Unlike in nominal regression, where categories have no inherent order, ordinal regression accounts for the ordinal nature of the target variable.

In ordinal regression, the goal is to predict the relative ordering of the categories of the dependent variable. The model assumes that there is a meaningful sequence in the categories, but it doesn't assume that the differences between the categories are equal.

Example: Suppose you are working on a survey where respondents rate a product on a scale of "Poor," "Fair," "Good," and "Excellent." The ratings form an ordinal scale because there is a clear order from worst to best, but the intervals between the categories are not necessarily uniform. Ordinal regression can be employed to build a model that predicts the likelihood of a higher rating based on certain features, such as product characteristics or customer demographics.



Ordinal regression methods include proportional odds models and continuation ratio models, which are designed to handle the ordinal nature of the target variable appropriately. These models are valuable in scenarios where the outcome variable represents ordered categories, and preserving the order information is crucial.

Example. A regression procedure can be used to test a patient's drug dosage. Possible solutions can be described as none, mild, prescribed or severe. The difference between a mild and moderate reaction is possible or imminent and is based on perception. Moreover, the difference between mild and moderate reactions may be greater or limited than between moderate and severe reactions. The same can be said in other states, such as good bad medium. or high risk or low risk

6. Logistic Regression. The scheme and an example in details

Logistic Regression is a statistical method used for binary classification tasks. Despite its name, it's primarily used for classification rather than regression. The algorithm models the probability of an instance belonging to a particular category, with the output being a value between 0 and 1. This output is then transformed using a logistic function (sigmoid) to make predictions.

Example: Email Spam Classification.

Suppose you have a dataset of emails, and for each email, you have various features like the frequency of certain words, presence of links, and other characteristics. The target variable is binary, indicating whether an email is spam (1) or not spam (0).

7. Dimensionality Reduction. The scheme and an example in details

Dimensionality reduction is a technique used in machine learning and data analysis to reduce the number of features or variables in a dataset. This is often done to simplify the dataset, speed up training of models, and mitigate the curse of dimensionality. One common method for dimensionality reduction is Principal Component Analysis (PCA).

Example: Facial Recognition with PCA

Let's consider an example of using PCA for dimensionality reduction in the context of facial recognition.

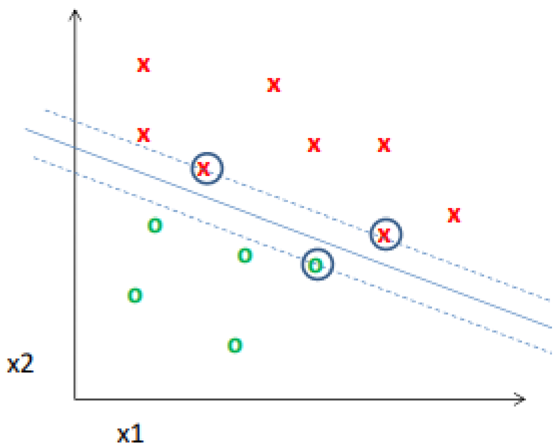
1. Data Collection: Collect a dataset of facial images where each image is represented by a large number of pixels. Each pixel can be considered as a feature, making the dataset high-dimensional.

2. High Dimensional Data: Each facial image can have thousands or even millions of pixel features, making the dataset computationally expensive and potentially prone to overfitting.
3. Feature Extraction with PCA: Apply Principal Component Analysis (PCA) to the facial image dataset. PCA identifies the principal components, which are linear combinations of the original features, capturing the most significant variance in the data.
4. Dimensionality Reduction: Retain only the top principal components that capture a significant percentage of the total variance in the data. This reduces the dimensionality of the data while retaining as much information as possible.
5. Model Training: Train a facial recognition model using the reduced-dimensional dataset. The model might be a classifier that learns to recognize individuals based on facial features.

By reducing the dimensionality of the facial image dataset with PCA, you can achieve a more computationally efficient model while preserving the essential information needed for facial recognition. The reduced set of features may also help the model generalize better to new, unseen data.

8. Classifiers. The scheme and an example in details

Classifiers are machine learning algorithms used to assign data to specific categories or classes based on input features. An example of a classifier would be a Support Vector Machine (SVM). Suppose we have a dataset containing technical characteristics of cars (such as engine horsepower, mileage, year of manufacture) and whether the car is used or new. Using the SVM, we can create a model that will classify new cars as used or new based on these characteristics.



Another example of a classifier is the Random Forest algorithm. We can use it to classify email messages as spam or non-spam based on the characteristics of the message content (e.g. keywords, message length, presence of links).

Classifiers are widely used in various fields, such as medicine, finance, natural language processing, to predict or identify specific events, group data or make decisions based on analysis of input features.

9. Nearest Neighbor Classifier. The scheme and an example in details

The nearest neighbor classifier (kNN) is a simple classification algorithm in machine learning. It assumes that a new data point belongs to the class to which most of its nearest neighbors from the training set belong. For example, if we have a dataset containing information about a fruit (e.g., its color, shape, size) and its classification as an apple, banana or orange, a nearest-neighbor classifier can determine which class the new fruit belongs to by analyzing the features of the fruit around it.

In practice, to classify a new data point using kNN, the distance between that point and all points in the training set is calculated. The k nearest points (neighbors) are then selected and the new point is classified as belonging to the class to which most of these neighbors belong.

The nearest-neighbor classifier is used in various fields, such as image recognition, recommender systems and biomedical data analysis, because of its simplicity and flexibility in classifying data based on its similarity to existing patterns.

10. Naïve Bayes. The scheme and an example in details

The Naive Bayes classifier is based on Bayes' theorem and assumes independence between predictor variables, which may be a simplification, but often works effectively in practice. An example of the application of the Naive Bayes algorithm could be the classification of emails as spam or non-spam based on words occurring in the body of the email.

For example, given a collection of emails marked as spam or non-spam and analyzing the occurrence of certain words (e.g. "offer," "win," "discount" vs. "job," "meeting," "information"), the Naive Bayes algorithm is able to predict whether a new email is spam or not by calculating the probability based on the occurrence of these words.

The algorithm uses Bayes' rule to calculate the probability that a new data point belongs to a given class based on previous observations. Despite its simplicity, Naive Bayes can be an effective classifier in text analysis, pattern recognition in data, email spam filtering and many other applications, especially when the data has a large number of features.

11. Linear SVM. The scheme and an example in details

Certainly! A Linear Support Vector Machine (SVM) is a machine learning algorithm used for classification tasks by finding the optimal hyperplane that best separates different classes in a dataset. Let's consider an example where we have a dataset of flower samples with features like petal length and width. We want to classify these flowers into two categories: roses and daisies.

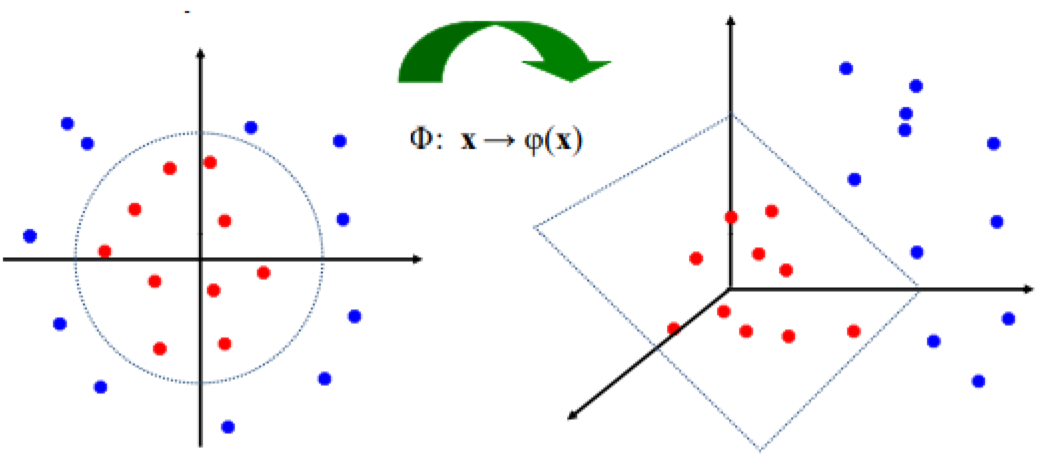
Using a Linear SVM, the algorithm tries to find the best straight line (in two dimensions) that can separate these two types of flowers based on their petal measurements. The goal is to maximize the margin, which is the distance between the hyperplane and the nearest data points from each class.

For instance, if we have data points representing petal measurements of different flowers and we label them as roses or daisies, the Linear SVM will determine the hyperplane that effectively distinguishes between these two classes. This hyperplane serves as a decision boundary, allowing us to predict the class of new flowers based on their petal measurements; whether they fall on one side of the hyperplane or the other.

Linear SVMs are widely used in various fields such as image classification, text classification, and biological data analysis due to their effectiveness in finding linear decision boundaries in high-dimensional spaces. They are especially useful when the data is linearly separable, making them a powerful tool for classification tasks.

12. Nonlinear SVM. The scheme and an example in details

Nonlinear Support Vector Machines (SVMs) use kernel functions to handle data that can't be separated by a straight line. For example, consider classifying fruits based on features like color and size. If the fruit data is not separable by a linear boundary, a nonlinear SVM with a kernel (like the radial basis function - RBF) transforms the data to a higher-dimensional space. This transformation helps find a curved or nonlinear boundary that separates different fruit types. Nonlinear SVMs are effective for complex classifications where the relationship between features isn't linear.
General idea: the original input space can always be mapped to some higher- dimensional feature space where the training dimensional feature space where the training set is separable:

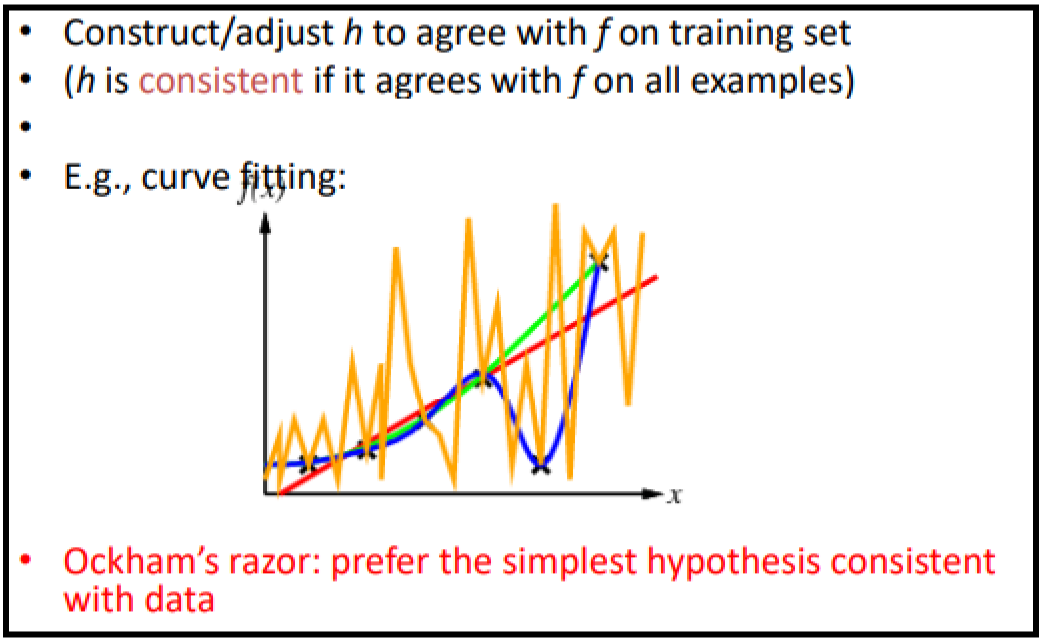


13. Inductive learning. The scheme and an example in details

A machine learning technique called inductive learning trains a model to generate predictions based on examples or observations. During inductive learning, the model takes knowledge from specific examples or cases and generalises it in such a way that it can predict outcomes for completely new data. The model is trained to detect trends and relationships in the input data and then uses this knowledge to predict outcomes from fresh data. The aim of inductive learning is to create a model that can accurately predict the outcome of subsequent cases.

Advantages:

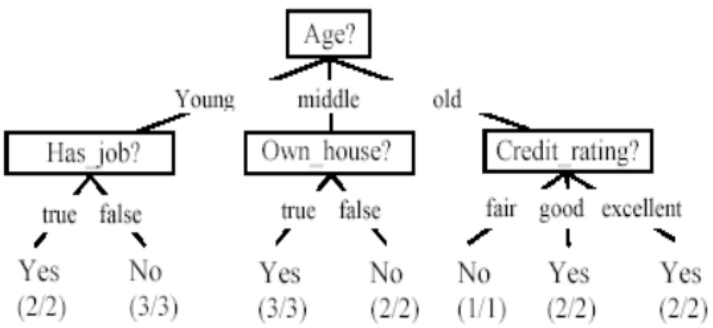
are suitable for processing difficult, complex and dynamic information. -are suitable for tasks such as pattern recognition and classification, as they enable the identification of relationships and patterns in the data, -are suitable for applications requiring the processing of huge amounts of data, -are suitable for situations where the rules are not precisely described or understood beforehand.



14. Decision tree. The scheme and an example in details

A decision tree from the loan data

- Decision nodes and leaf nodes (classes)



Decision tree learning is one of the most widely used techniques for classification.
– Its classification accuracy is competitive with other methods, and
– it is very efficient.
• The classification model is a tree, called decision tree

15. Precision and recall measures. The scheme and an example in details

Classification measures
• Accuracy is only one measure (error = 1-accuracy).
• Accuracy is not suitable in some applications.
• In text mining, we may only be interested in the documents of a particular topic, which are only a small portion of a big document collection.
• In classification involving skewed or highly imbalanced data, e.g., network intrusion and financial fraud detections, we are interested only in the minority class.
– High accuracy does not mean any intrusion is detected.
– E.g., 1% intrusion. Achieve 99% accuracy by doing nothing.
• The class of interest is commonly called the positive class, and the rest negative cla
Precision and recall measures.
• Used in information retrieval and text classification.
• We use a confusion matrix to introduce them.

16. Rule evaluation in learn-one-rule-2. The scheme and an example in details

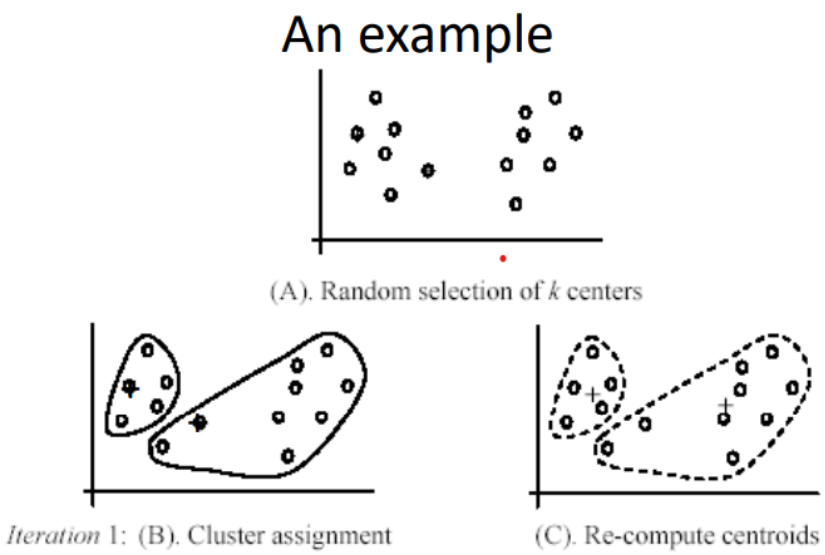
The "Learn-One-Rule-2" algorithm creates rules to classify data. For instance, in sorting fruits, it might start with a rule like "If the fruit is red, then it's an apple." It checks if this rule accurately separates apples from other fruits. It refines rules based on different features like size or taste, choosing the most accurate rule, such as considering both color and size together, for classifying apples.

17. Class association rules. The scheme and an example in details

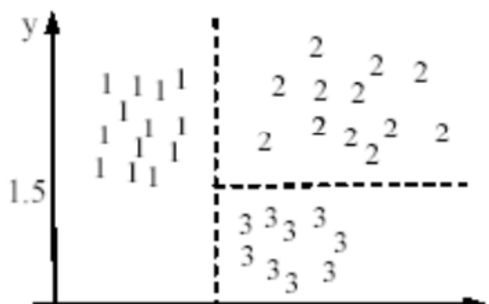
Three approaches
• Three main approaches of using association rules classification.
– Using class association rules to build classifiers
– Using class association rules as attributes/features
– Using normal association rules for classification.
Using Class Association Rules
• Classification: mine a small set of rules existing in the data to form a classifier or predictor.
– It has a target attribute: Class attribute
• Association rules: have no fixed target, but we can fix a target.
• Class association rules (CAR): has a target class attribute. E.g.,
Own_house = true -> Class =Yes [sup=6/15, conf=6/6]
– CARs can obviously be used for classification.

18. K-means clustering. The scheme and an example in details

K-means clustering groups similar data points together. For example, in customer data, it can group customers based on their spending habits. If we want to form three clusters, K-means starts by choosing three random points as cluster centers. It then assigns customers to the nearest cluster center based on their buying behavior. After several iterations of updating cluster centers and reassigning customers, it forms clusters where customers within each group have similar purchasing patterns. This helps businesses understand different customer segments for targeted marketing or analysis.



19. Representation of clusters. The scheme and an example in details

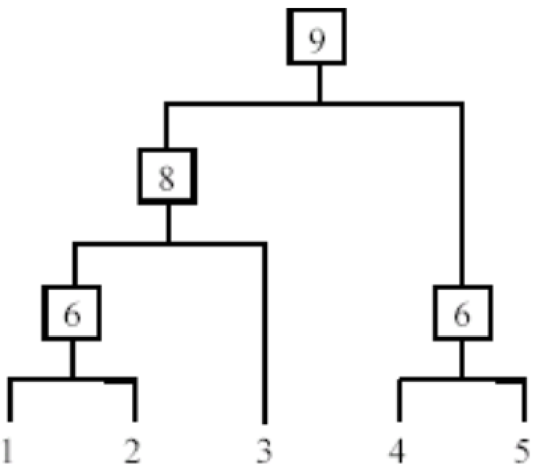


Common ways to represent clusters.
• Use the centroid of each cluster to represent the cluster.
– compute the radius and
– standard deviation of the cluster to determine its spread in each dimension
– The centroid representation alone works well if the clusters are of the hyper spherical shape.
If clusters are elongated or are of other shapes, centroids are not sufficient.

20. Hierarchical Clustering. The scheme and an example in details

Hierarchical clustering groups data into a tree-like structure. For example, with animal data, it starts by considering each animal as a separate group. Then, it merges similar animals together step by step until they form clusters based on their features. This method creates a tree diagram (dendrogram) showing how animals are grouped based on their similarities.

20. Hierarchical Clustering. The scheme and an example in details



Hierarchical clustering groups data into a tree-like structure. For example, with animal data, it starts by considering each animal as a separate group. Then, it merges similar animals together step by step until they form clusters based on their features. This method creates a tree diagram (dendrogram) showing how animals are grouped based on their similarities.

21. Distance functions. The scheme and an example in details

Distance functions
• Key to clustering. “similarity” and “dissimilarity” can also commonly used terms.
• There are numerous distance functions for
– Different types of data
• Numeric data
• Nominal data
– Different specific applications.

Distance functions for numeric attributes
• Most commonly used functions are
– Euclidean distance and
– Manhattan (city block) distance

22. Data standardization. The scheme and an example in details

Data standardization makes data have a common scale. For instance, in income and age data, standardization ensures both features have a similar scale by adjusting their means and spreads. This helps algorithms work better and prevents certain features from having undue influence due to their scale differences.