VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA

AIML231/DATA302 —Techniques in Machine Learning

Week 11 - Advanced Regression and Clustering Algorithms

Dr Qi Chen

School of Engineering and Computer Science

Victoria University of Wellington

Qi.Chen@vuw.ac.nz

Outline

- Introduction to Advanced Regression
  - Overview of different regression techniques
  - Importance and applications in various domains

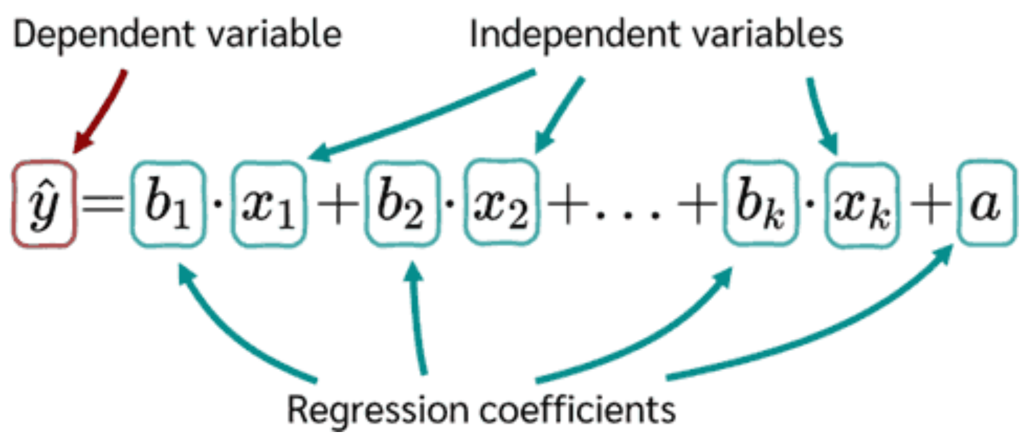
- Advanced Regression Techniques
  - Logistic Regression*
  - Polynomial Regression
  - Genetic Programming for Symbolic Regression

- Introduction to Advanced Clustering
  - Overview of different regression techniques
  - Importance and applications in various domains

- Advanced Clustering Techniques
  - Mean Shift Clustering
  - DBSCAN
  - BIRCH

Advanced Regression

- Regression analysis is a machine learning technique used to examine the relationship between one or more independent variables and a dependent variable



$$\hat{y} = b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_k \cdot x_k + a$$

- different types of regression analysis techniques get used when the target and input variables show a linear or non-linear relationship with the target variable contains continuous values

- advanced regression techniques enhance traditional regression methods by addressing various limitations e.g., overfitting, feature selection, and handling non-linear relationships.

Advanced Regression Applications

- regression technique is used mainly to determine the predictor strength, forecast trend, time series, and cause & effect relation
- advanced regression analysis a powerful tool for data-driven decision-making in various fields

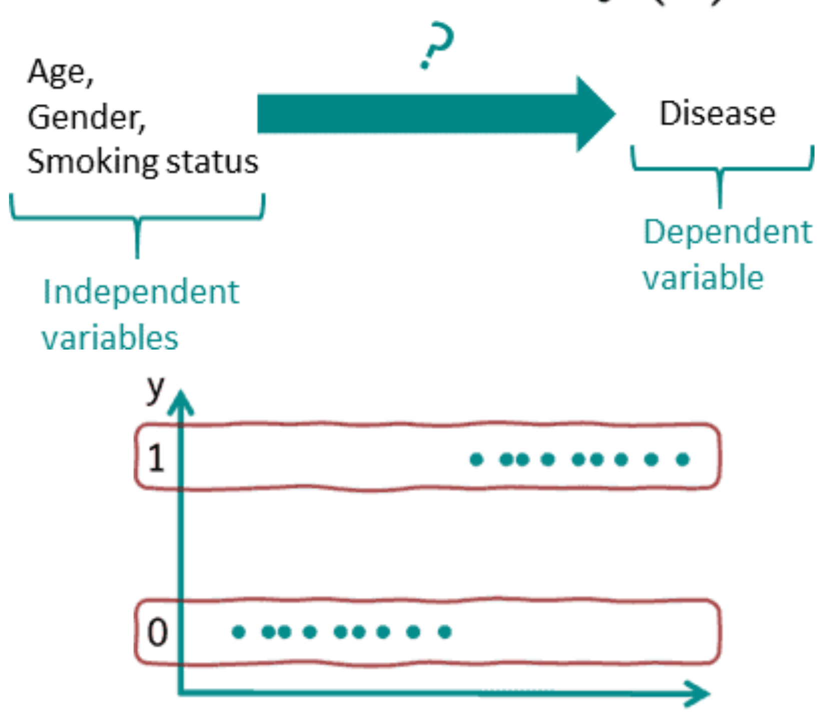
AI Generated

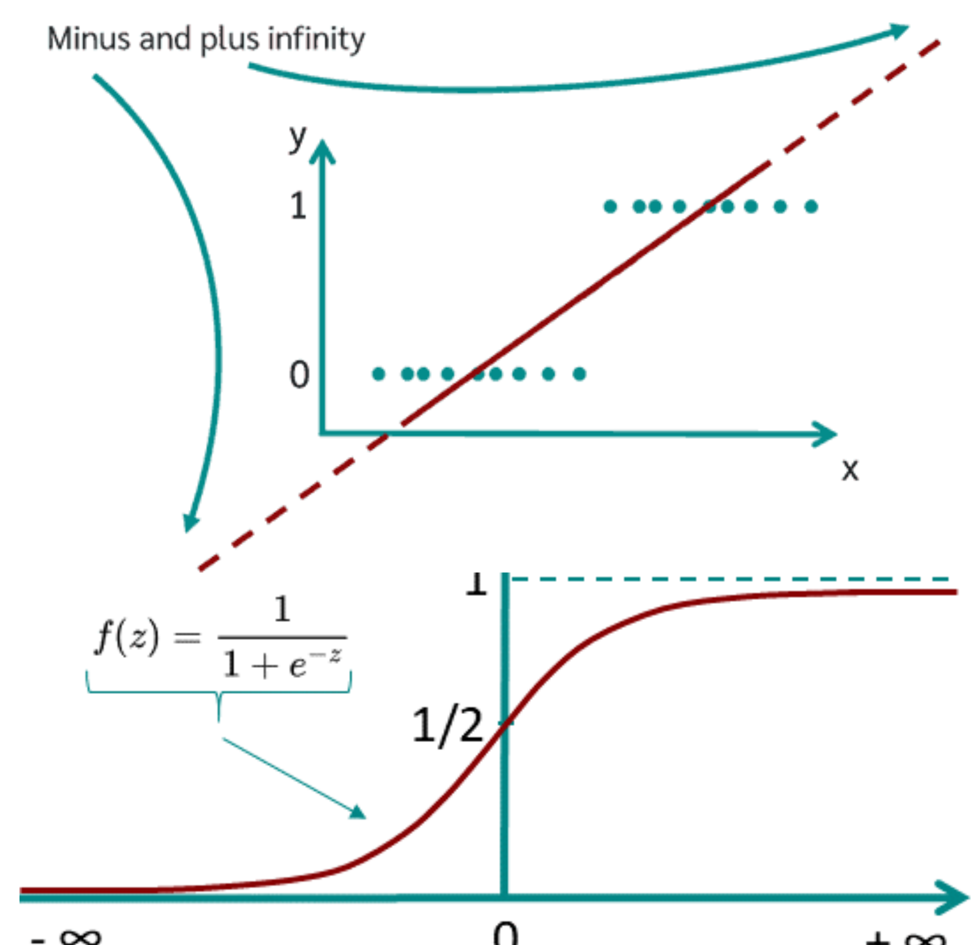
Logistic Regression

- Logistic regression is a regression analysis technique used for when the target variable is discrete (the basic form, 0 or 1)
- the probability of the occurrence of value 1 is estimated

$$\hat{y} = b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_k \cdot x_k + a$$

$$f(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(b_1 \cdot x_1 + \ldots + b_k \cdot x_k + a)}}$$





$$P(\text{Diseased}) = \frac{1}{1+e^{-(b_1 \text{Age} + b_2 \text{Gender} + b_3 \text{Smoking status} + a)}}$$

Coefficient Learning in Logistic Regression

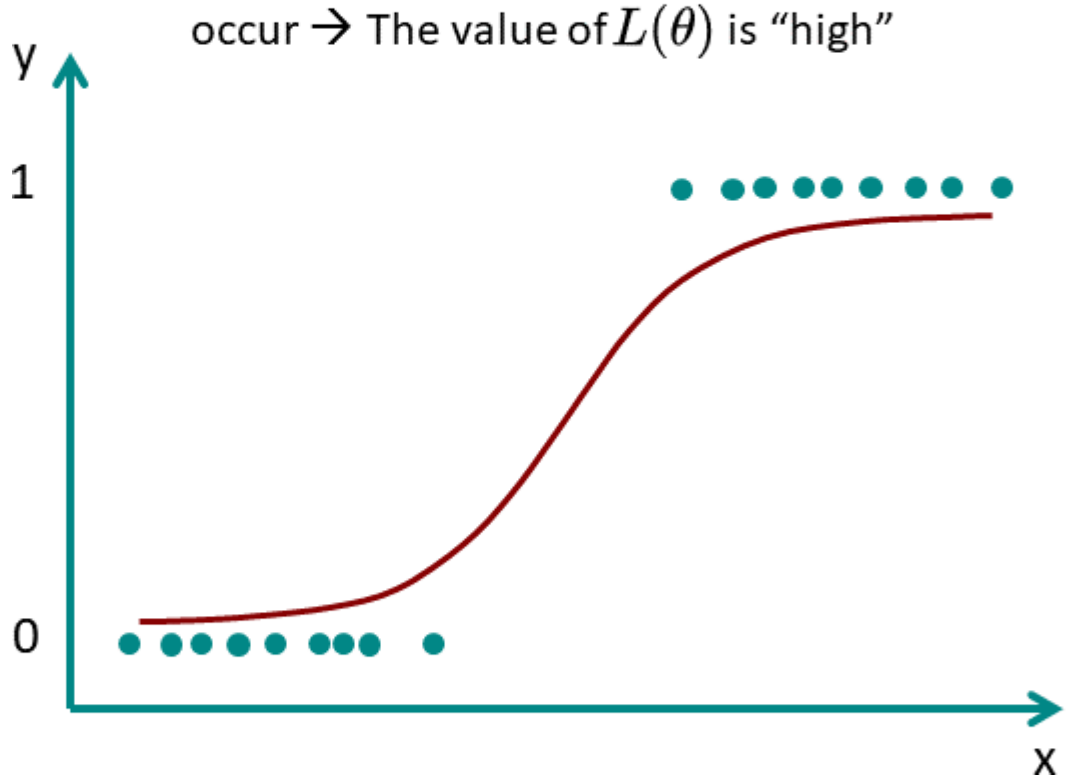
- the Maximum Likelihood Method is applied to determine the model parameters for the logistic regression equation

- introduce the likelihood function *L(b1,... bn, a)* or $L(\theta)$, indicates how probable it is that the observed data occur

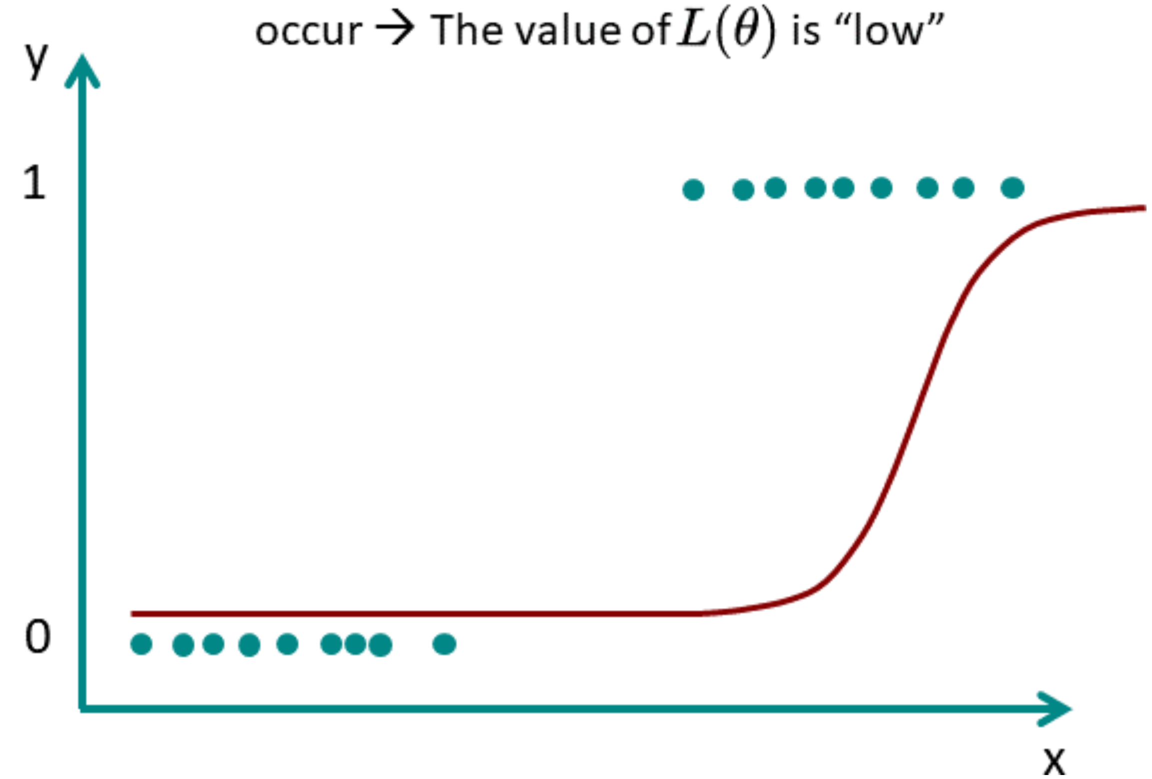
$$L(\theta) = \prod_{i=1}^{n} P(y_i|x_i;\theta)$$

- Stochastic gradient descent to maximize the log likelihood function $Log(L(\theta))$

With the given logistic function, the probability is "high" that the given points occur → The value of $L(\theta)$ is "high"



With the given logistic function, the probability is "low" that the given points occur → The value of $L(\theta)$ is "low"



Polynomial Regression

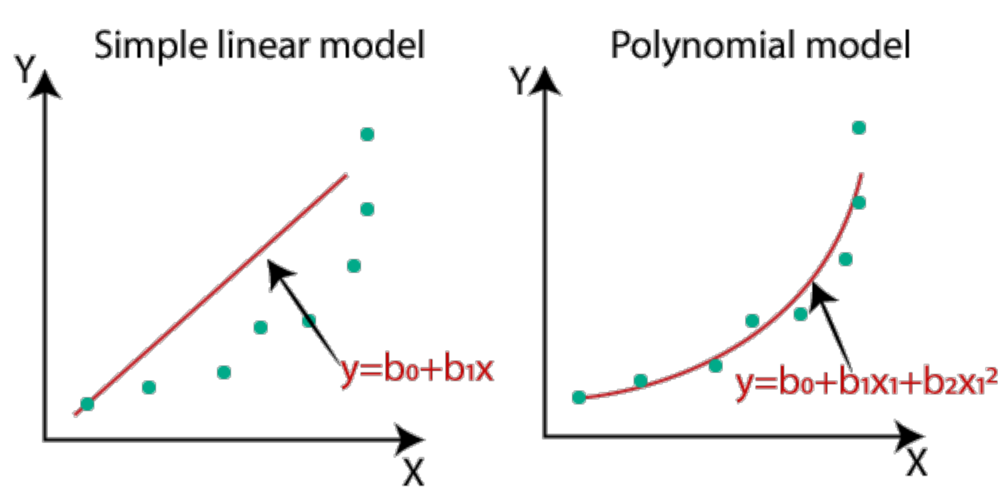
- Polynomial Regression is a regression analysis in which the relationship between the independent variables and dependent variables are modeled in the n^{th} degree polynomial

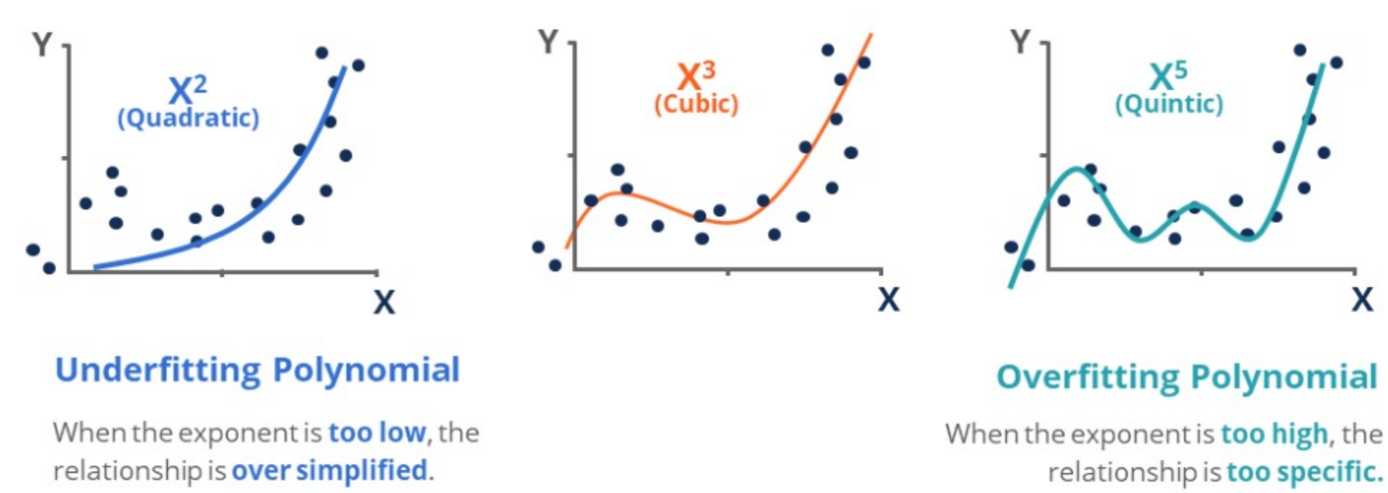
$$y = b_0 + b_1 x_1 + b_2 x_1^2 + \ldots + b_n x_1^n$$

can have multivariate polynomial regression, but You don't see this often

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{i=1}^{p}\sum_{j=i}^{p} \beta_{ij} x_i x_j + \sum_{i=1}^{p}\sum_{j=i}^{p}\sum_{k=j}^{p} \beta_{ijk} x_i x_j x_k + \cdots + \beta_{1,2,\ldots,n} x_1 x_2 \cdots x_n + \varepsilon$$

- increase the degree in the model, it tends to increase the performance of the model





Coefficient learning in Polynomial Regression

- Same as linear regression, we can use least squares estimation to learn coefficients by minimizing the sum of the squared residuals in the model

$$RSS = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_i + b_2 {x_i}^2 + \cdots + b_n {x_i}^2))^2$$

- Ordinary Least Squares: using matrix algebra by solving the normal equations

$$Y = X\beta + \epsilon \qquad X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

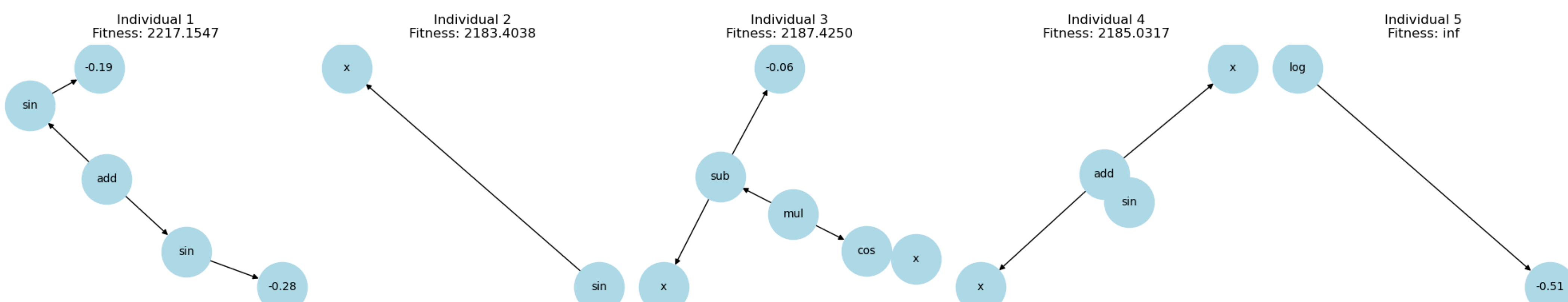
$$\text{RSS} = (Y - X\beta)^T(Y - X\beta) \Rightarrow \beta = (X^T X)^{-1} X^T Y$$

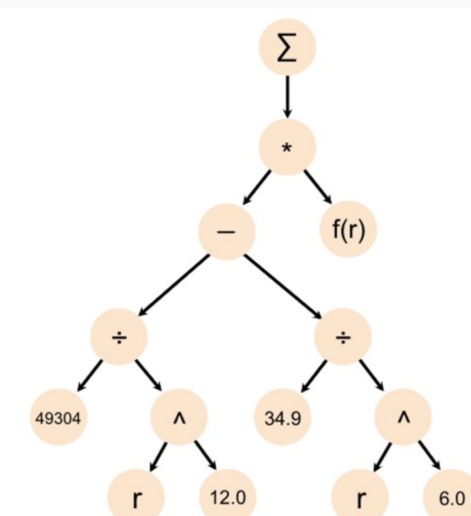
- Gradient Descent: iteratively updating the coefficients in the direction of the negative gradient
- Regularization techniques are used to prevent overfitting in polynomial regression, especially when dealing with high-degree polynomials

Symbolic Regression

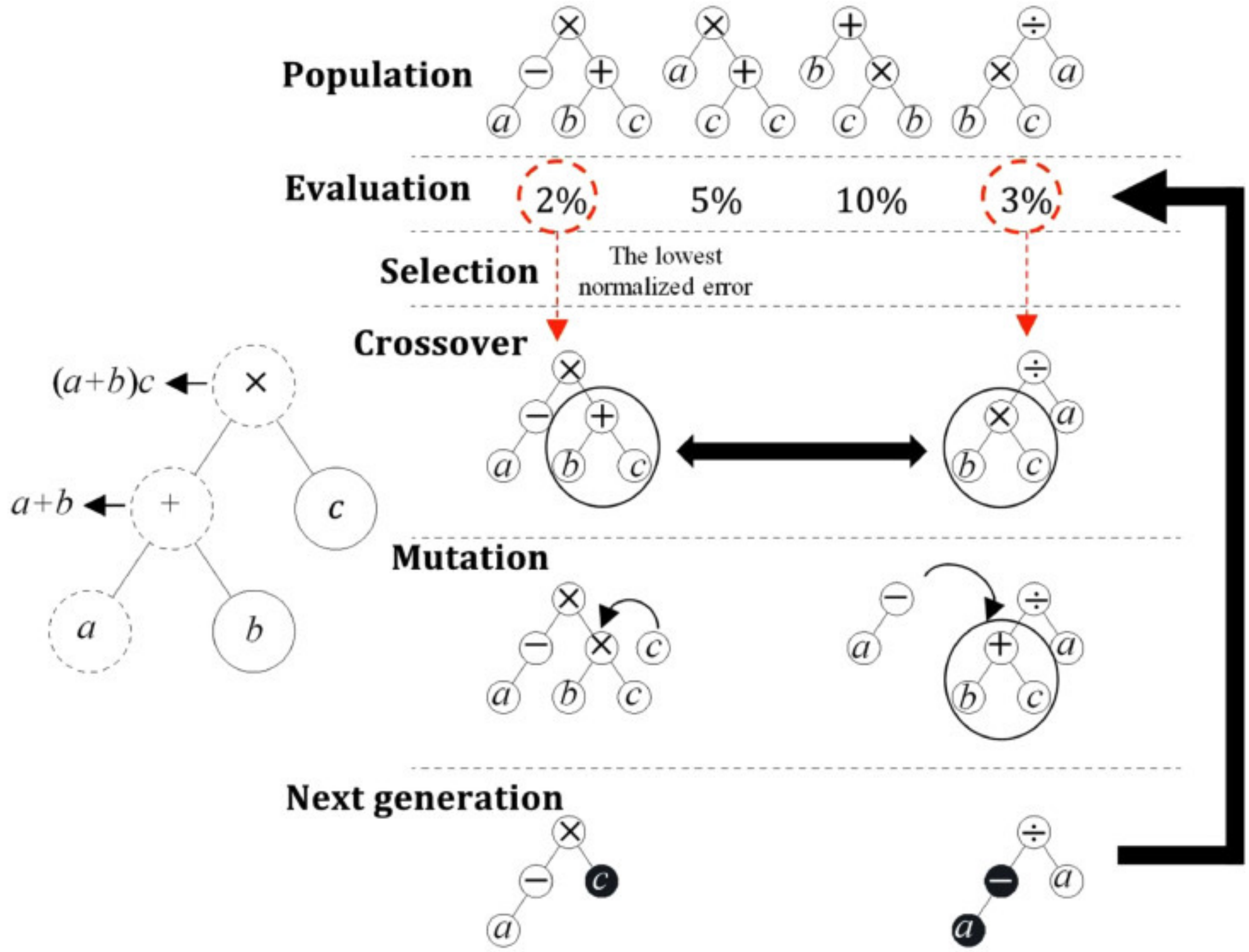
- Symbolic regression is a type of regression analysis that searches for mathematical expressions that best fit a given dataset
- Unlike traditional regression methods, which fit data to a predefined model, symbolic regression explores a space of mathematical expressions to find the most suitable model for the data

Function Set and Terminal Set





Genetic Programming for Symbolic Regression

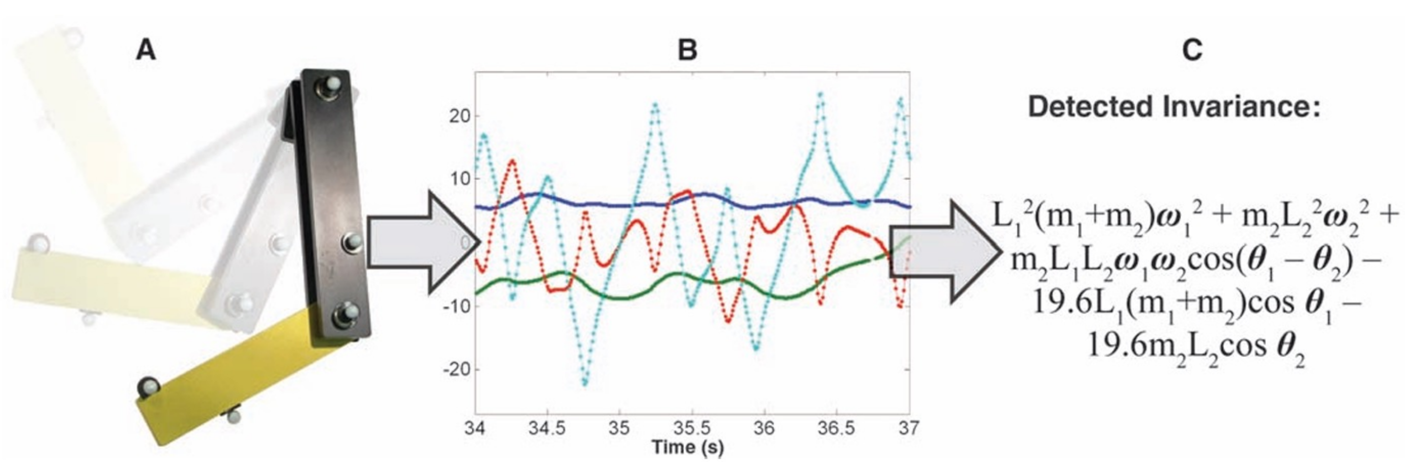


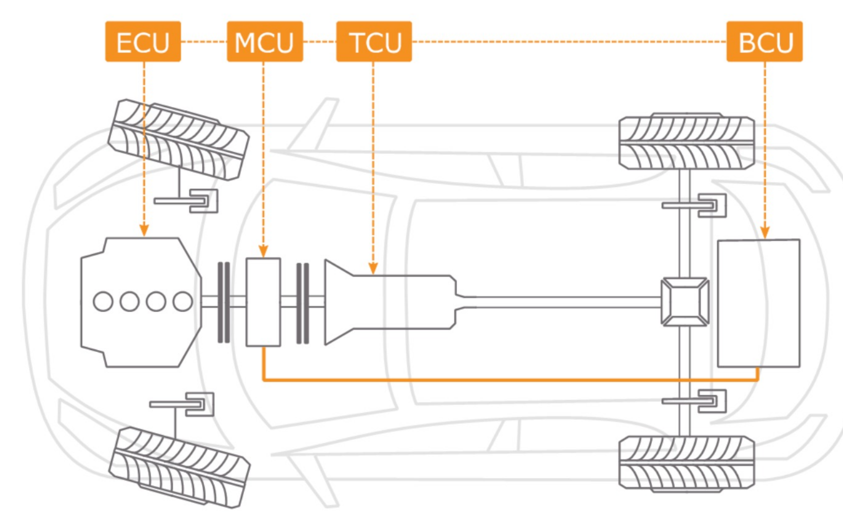
Symbolic Regression Applications

- discover both the form and parameters of the underlying model, making it highly flexible
- capture complex nonlinear relationships between variables
- models are often interpretable

Applications:

- Scientific Discovery: automatically identify mathematical formulas that explain experimental data
- Engineering: can be used to model systems where the underlying dynamics are complex or unknown
- Many application in Finance, Healthcare, Environmental Modeling, Robotics and Control Systems

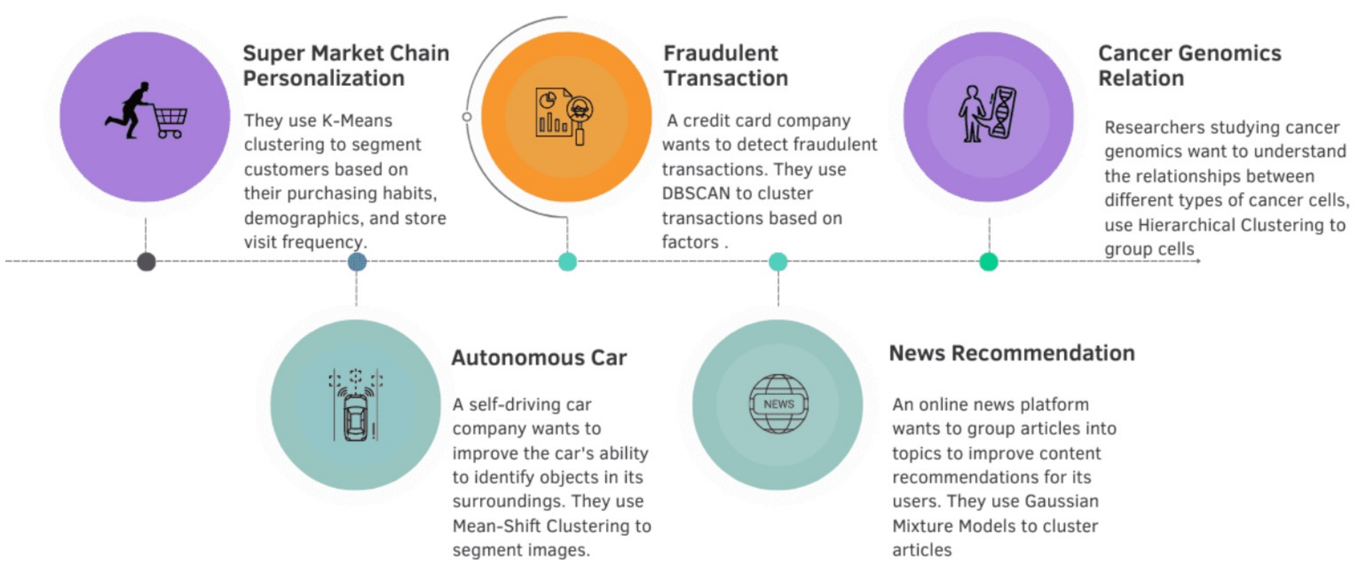




https://heal.heuristiclab.com/projects/jrc-symreg

Advanced Clustering

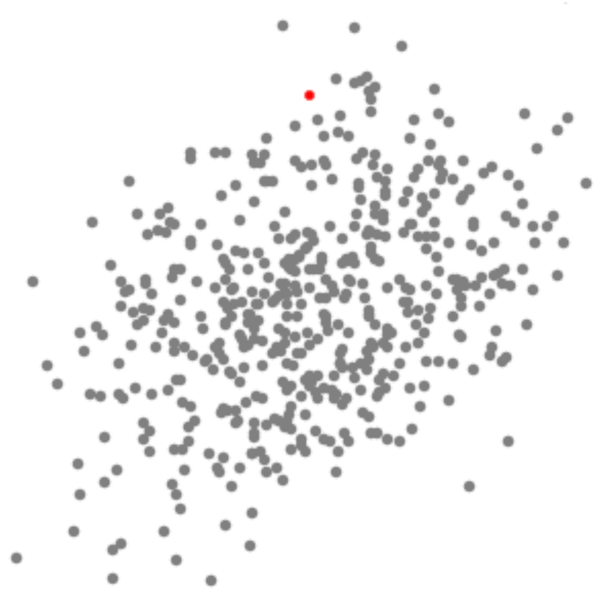
- Clustering techniques: partitioning methods, hierarchical methods, density-based methods, distribution-based Methods, …
- to handle more complex data distributions and structures, e.g. clusters of arbitrary shapes
- to efficiently handle large datasets, suitable for big data applications
- effective at identifying and handling noise and outliers
- Advanced clustering techniques are used in a variety of real-world applications such as market segmentation, anomaly detection, bioinformatics, social network analysis, and document clustering

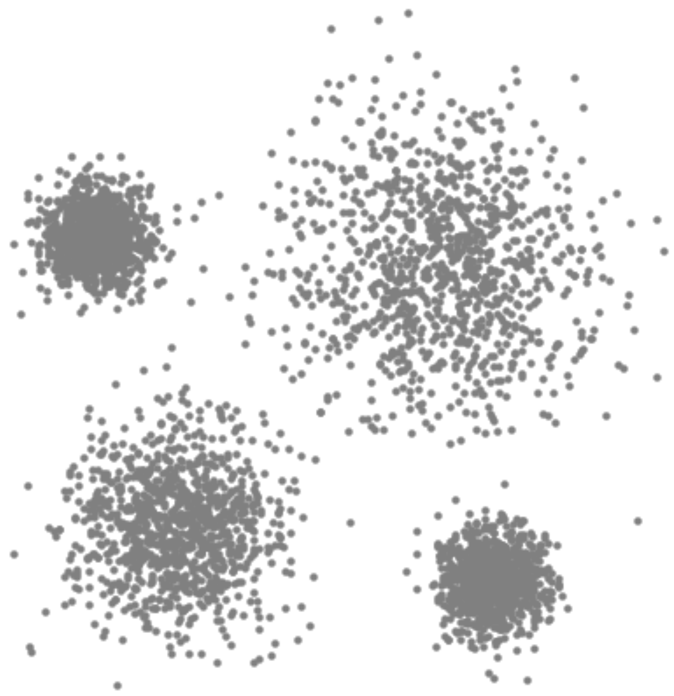


Mean Shift Clustering

a sliding-window-based algorithm that attempts to find dense areas of data points

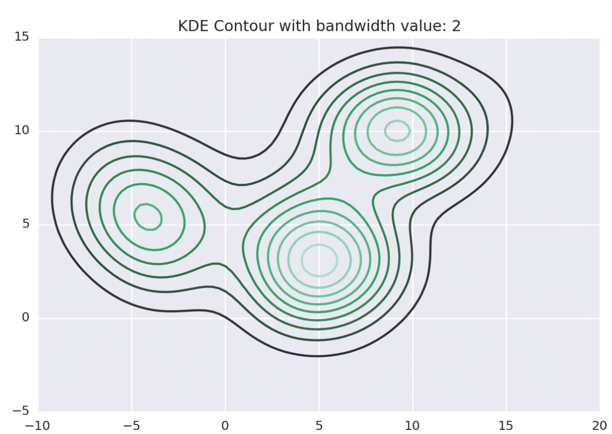
- the goal is to locate the center points of each group
- candidates for centroids to be the mean of the point in the sliding-window
- eliminate near-duplicates candidate windows
- need to define "bandwidth" but not number of clusters

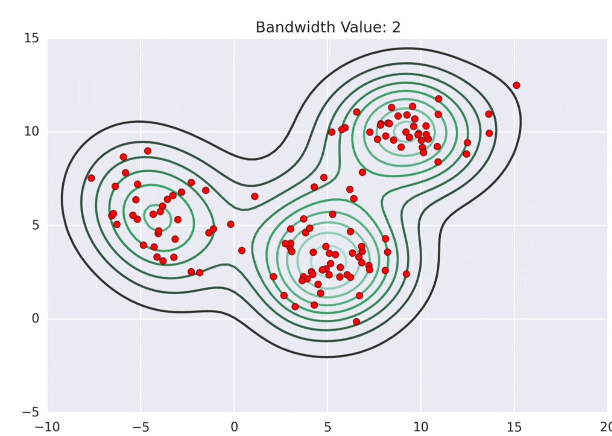


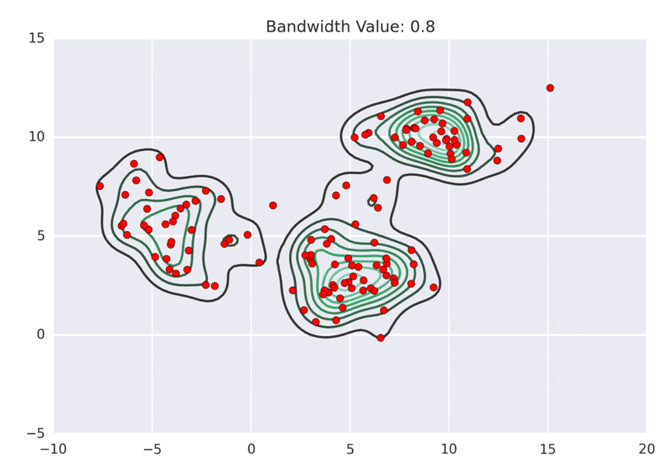


Mean Shift Clustering

- Initialise centers: start with the initial centers of the clusters, can be randomly selected or, typically, every data point is considered as an initial center.
- Calculate mean shift vectors: for each center, perform the following:
  - identify points within bandwidth
  - compute weighted mean: calculate the weighted mean of these points using a kernel function
- Update centers: update the position of each center to the computed weighted mean
- Check for convergence: measure the amount each center moved since the last iteration. If all centers move less than a predefined small threshold, then assume convergence and stop the iteration
- Assign clusters: assign each data point to the cluster of the nearest center
- Finalize clusters: Optionally, you can merge centers that are very close to each other to reduce the number of clusters



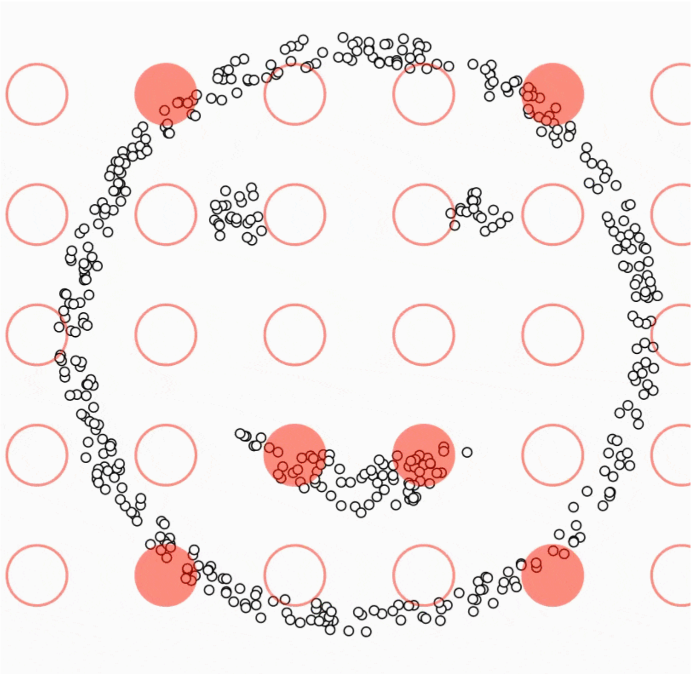




DBSCAN

Density-based spatial clustering of applications with noise

- Basic idea, identify clusters as sets of core samples that can be built by recursively
- Start with no cluster, and mark all points as unvisited, go through each point in the dataset that hasn't been visited
  - find all nearby points within a certain distance ('eps').
    - if there aren't enough nearby points (MinPts), mark it as noise.
    - if there are enough nearby points, start a new cluster:
      - add the point and its nearby points to the cluster.
      - for each point in the cluster:
        - if it hasn't been visited, mark it as visited and find its nearby points.
        - if there are enough nearby points, add them to the cluster.



DBSCAN

- a density-based clustering techniques, views clusters as areas of high density separated by areas of low density

- Points are identified as: core samples/points, which are samples that are in areas of high density; or outliers

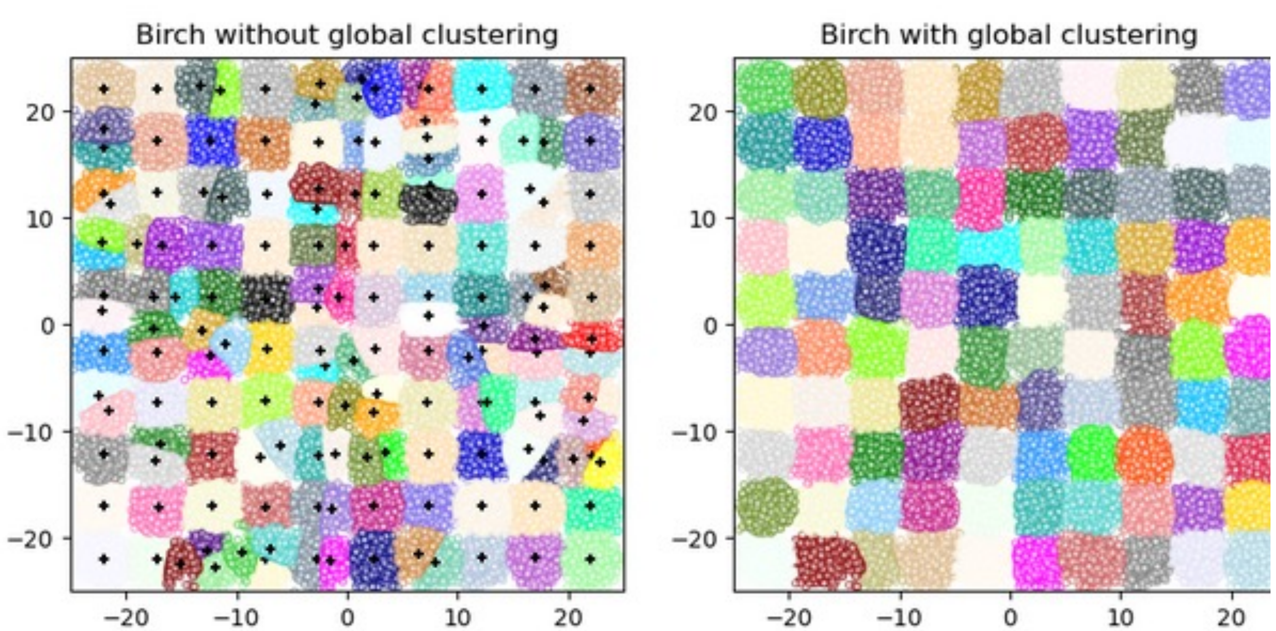
- primary strengths lies in its ability to identify clusters of arbitrary shapes
- No need to specify the number of clusters beforehand

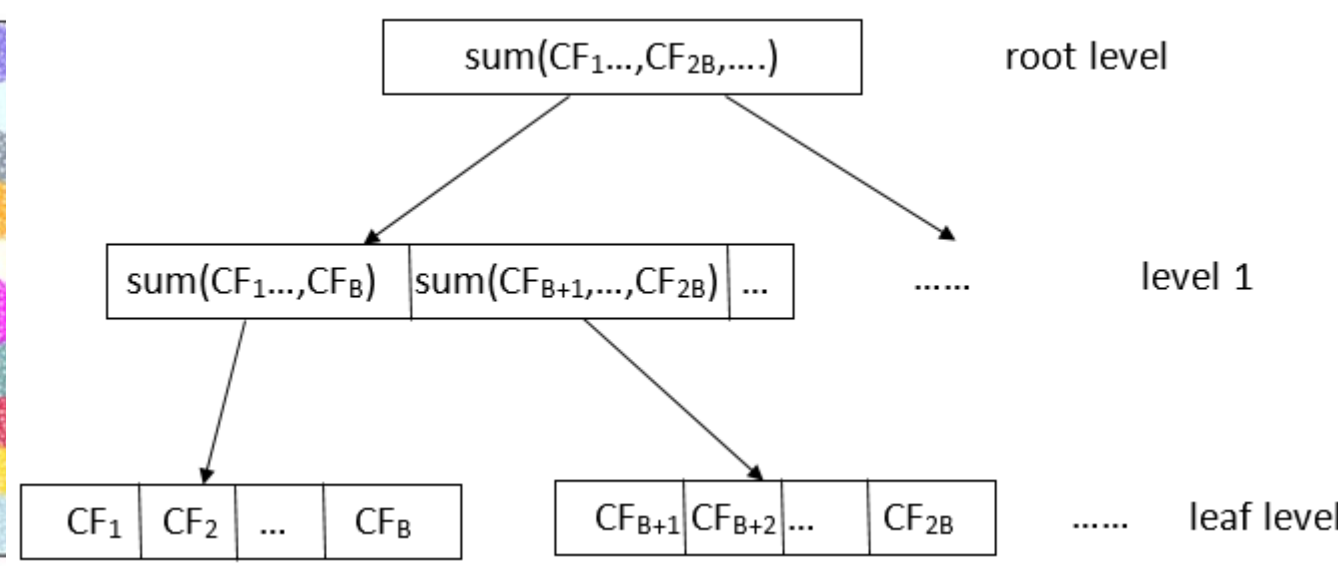
- Sensitive to two key parameters:
  - eps - the distance that specifies the neighborhoods
  - minPts - minimum number of data points to define a cluster
- struggle to handle datasets with varying densities

BIRCH Clustering

Balanced Iterative Reducing Clusters using Hierarchies

- Hierarchical clustering - builds a tree called the Clustering Feature Tree (CFT) for the given data, uses CF to summarize a cluster
- often used to complement other clustering algorithms cluster large datasets
- creating a summary of the dataset that the other clustering algorithm can now use





Summary

- Advanced regression techniques: Logistic, Polynomial, and Symbolic Regression
  - Logistic Regression for binary outcomes
  - Polynomial Regression for capturing non-linear relationships with a specified degree
  - Symbolic Regression for uncovering unknown mathematical relationships
  - Provide flexibility in modeling and predicting diverse types of data, making them valuable tools in various scientific and practical applications

- Advanced clustering techniques: Mean Shift, DBSCAN, and BIRCH
  - Mean Shift focuses on density peaks without predefined cluster numbers
  - DBSCAN excels at handling noise and finding clusters of arbitrary shape
  - BIRCH is optimized for large datasets with a hierarchical approach
  - enhance the flexibility and effectiveness of clustering in various fields