Machine Learning Algorithms

Supervised Learning		Unsupervised Learning	
Tree-Based Models	**Linear Models**	**Clustering**	**Association**
Decision Tree	Linear Regression	K-Means	Apriori Algorithm
Random Forest	Logistic Regression	Hierarchical	
Gradient Boosting Regresstion	Ridge Regression	Gaussian Mixture Models	
XGBoost	Lasso Regression		
LightGBM Regressor			

Chai Knows

Tech series

Supervised

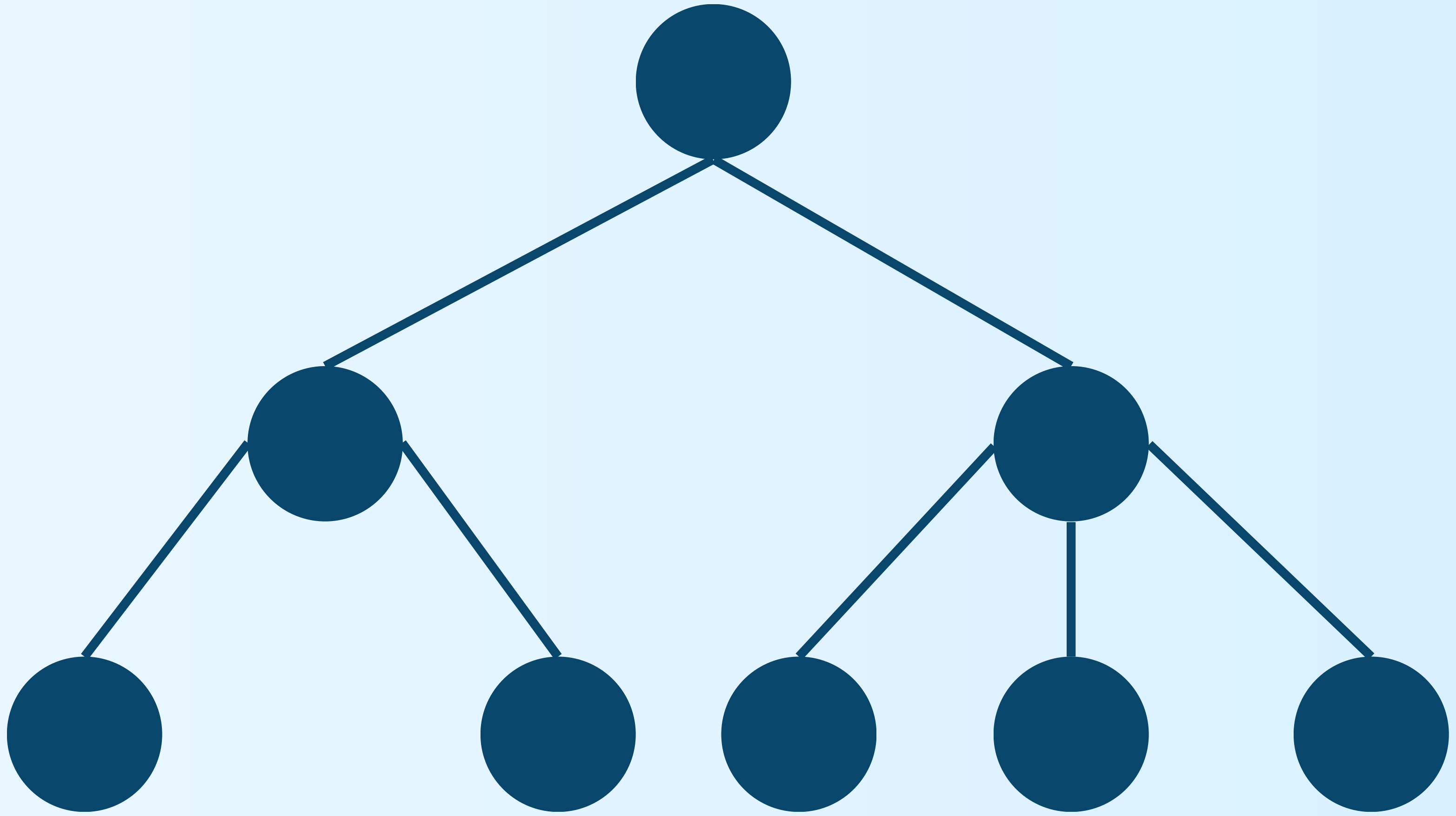
Tree-Based Models

Decision Tree

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences. It is a supervised learning algorithm that can be used for classification or regression. In a decision tree, an internal node represents a feature or attribute, and the branches from that node represent the possible values of that feature.

Applications / Use Cases

1. Classification and regression tasks
2. Predict stock prices
3. Quality control in manufacturing
4. Predict student performance
5. Predict the likelihood of certain environmental events



👍

1. Explainable & Interpretable
2. Can handle large datasets with many features
3. Can handle missing values
4. Can be used for both classification and regression tasks
5. Relatively fast to train and to use for making predictions
6. e less prone to overfitting than some other ML Algorithms

👎

1. If the training data is biased, the decision tree may learn from the biases and produce poor results when applied to new, unseen data.
2. Can be sensitive to the choice of hyperparameters:
3. Prone to overfitting
4. May not be the best choice for certain tasks
5. Sensitive to outliers

Supervised

Tree-Based Models

Random Forest

Random forests are a type of ensemble machine learning algorithm that is made up of multiple decision trees. The goal of a random forest is to make predictions by combining the predictions of many individual decision trees, each of which is trained on a different random subset of the training data.

Applications / Use Cases

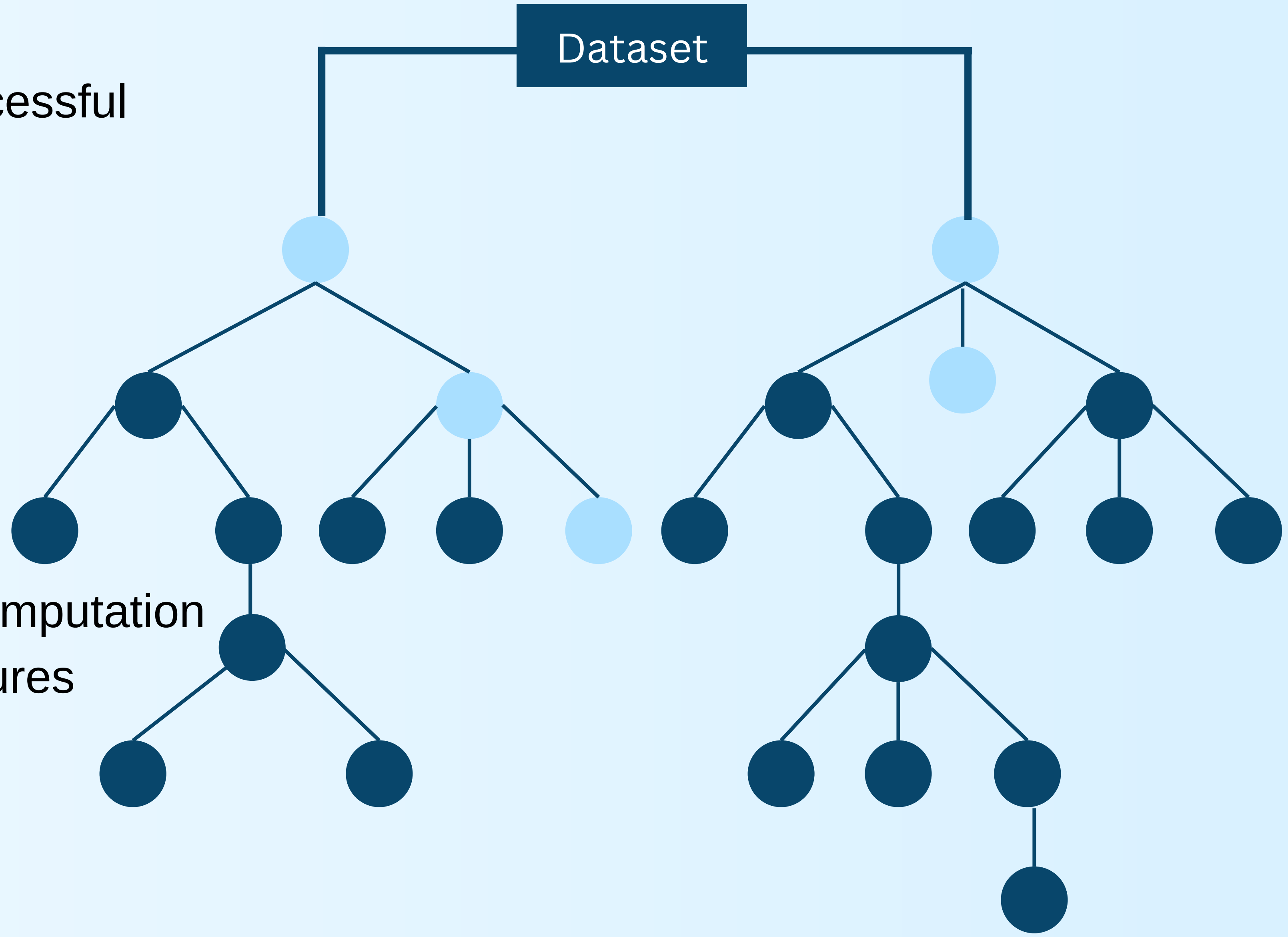
1. Predicting house prices
2. Predict score modeling
3. Predict which marketing campaigns are most likely to be successful
4. Optimize production processes
5. To identify factors that affect crop growth.

👍

1. Highly accurate in many different applications
2. Less prone to overfitting than individual decision trees
3. Relatively fast to train and to use for making predictions
4. Can handle large datasets with many features
5. Can handle missing values in training data without need for imputation
6. Can handle a mix of categorical, ordinal, and numerical features

👎

1. Can be difficult to understand and interpret
2. Training a random forest can take a long time
3. Difficult to interpret
4. Requires a large amount of memory and computational resources,
5. Predictions can be unstable
6. Prone to overfitting, especially when you have a large number of trees



Supervised

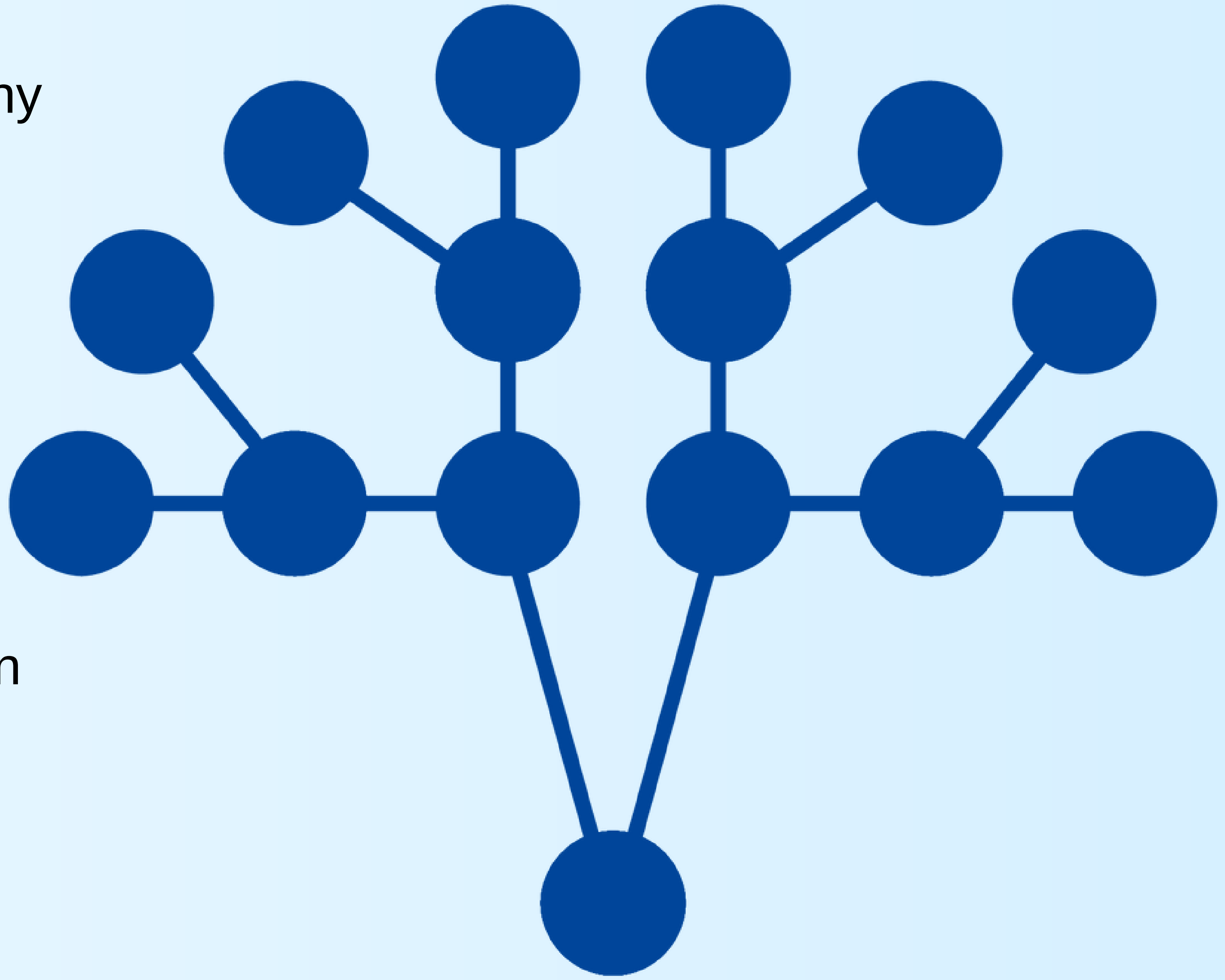
Tree-Based Models

Gradient Boosting Regression

Gradient Boosting Regression is an ensemble machine learning technique that combines the predictions of multiple smaller models to create a final prediction. It is used for regression problems, which involve predicting a continuous value output, rather than a class label.

Applications / Use Cases

1. Predicting car emissions
2. Predicting ride hailing fare amount
3. Forecast sales for a business by analyzing historical sales data
4. Build models that predict which customers are likely to leave a company
5. Predict the traffic on website



👍

1. Can produce highly accurate predictions even with non-linear data
2. Can handle missing values in the input data
3. Can handle mixed data types - continuous and categorical variables
4. Can handle highly imbalanced datasets
5. Computationally efficient and can handle large datasets
6. Flexible algorithm that can be used for both classification & regression

👎

1. Can be computationally intensive
2. Requires the selection of several hyperparameters
3. Creates a complex model that is difficult to interpret,
4. Can be sensitive to noisy data and may not perform well
5. Can be memory intensive, especially when dealing with large datasets

Supervised

Tree-Based Models

XGBoost

XGBoost is an optimized implementation of the Gradient Boosting algorithm. It stands for "Extreme Gradient Boosting" and is an open-source software library for gradient boosting on decision trees. The algorithm was developed and open-sourced by Tianqi Chen at the University of Washington.

Applications / Use Cases

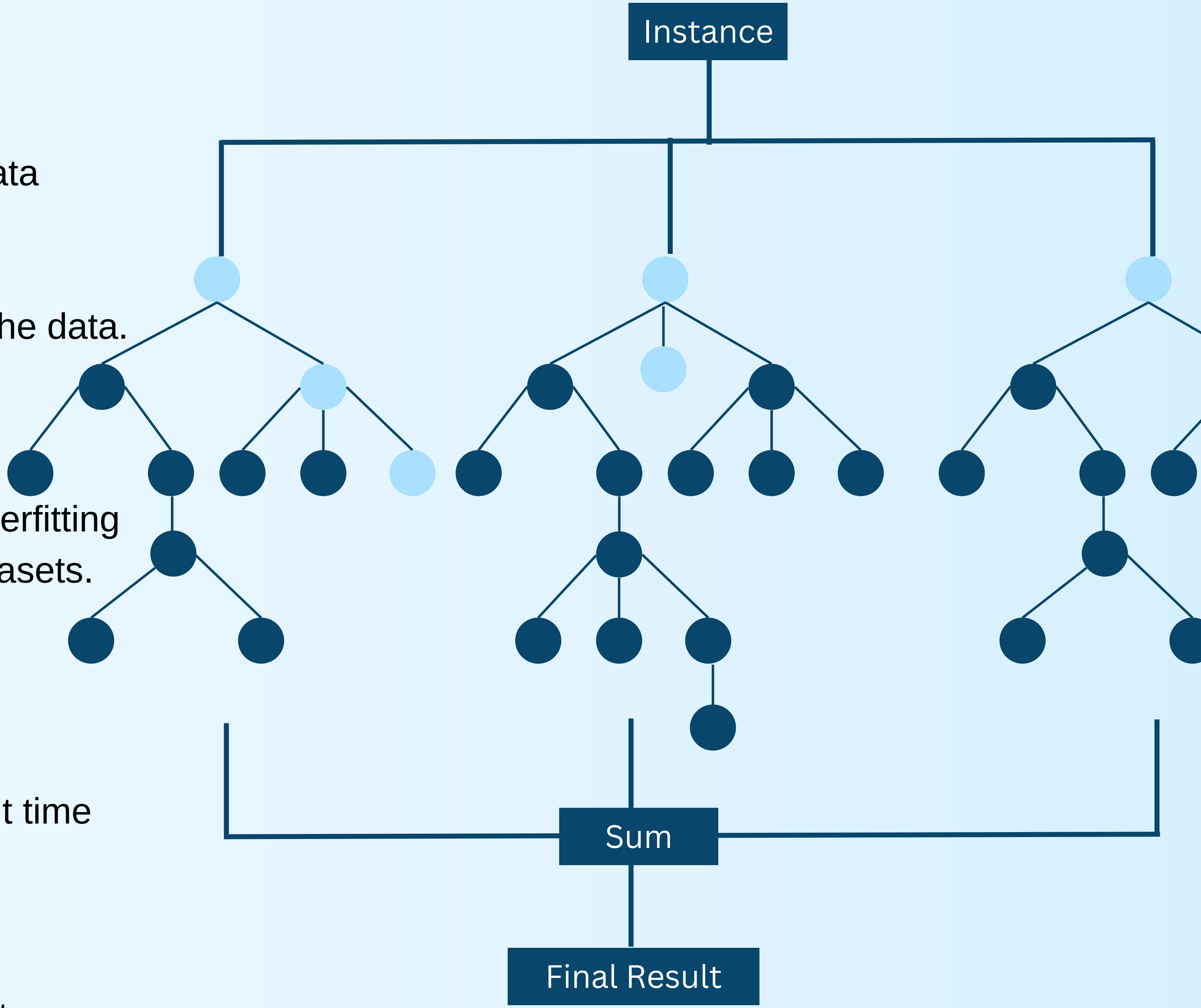
1. Predict the optimal price for a product or service
2. Identify anomalies or intrusions in network traffic data
3. Predict when a machine or equipment will fail
4. Predict the severity of disease
5. Identify fraudulent claims by analyzing patterns in the data.

👍

1. Provides feature importance
2. Has built-in support for cross-validation
3. Introduces regularization parameters to reduce overfitting
4. computationally efficient and can handle large datasets.
5. Can handle mixed data types
6. Can handle missing values in the input data

👎

1. Limited to gradient boosting algorithm
2. Training and tuning the model can take a significant time
3. Can be memory intensive
4. Heavily depends on the parameters settings
5. Can be sensitive to noisy data
6. Creates a complex model that is difficult to interpret



Instance

Sum

Final Result

Supervised

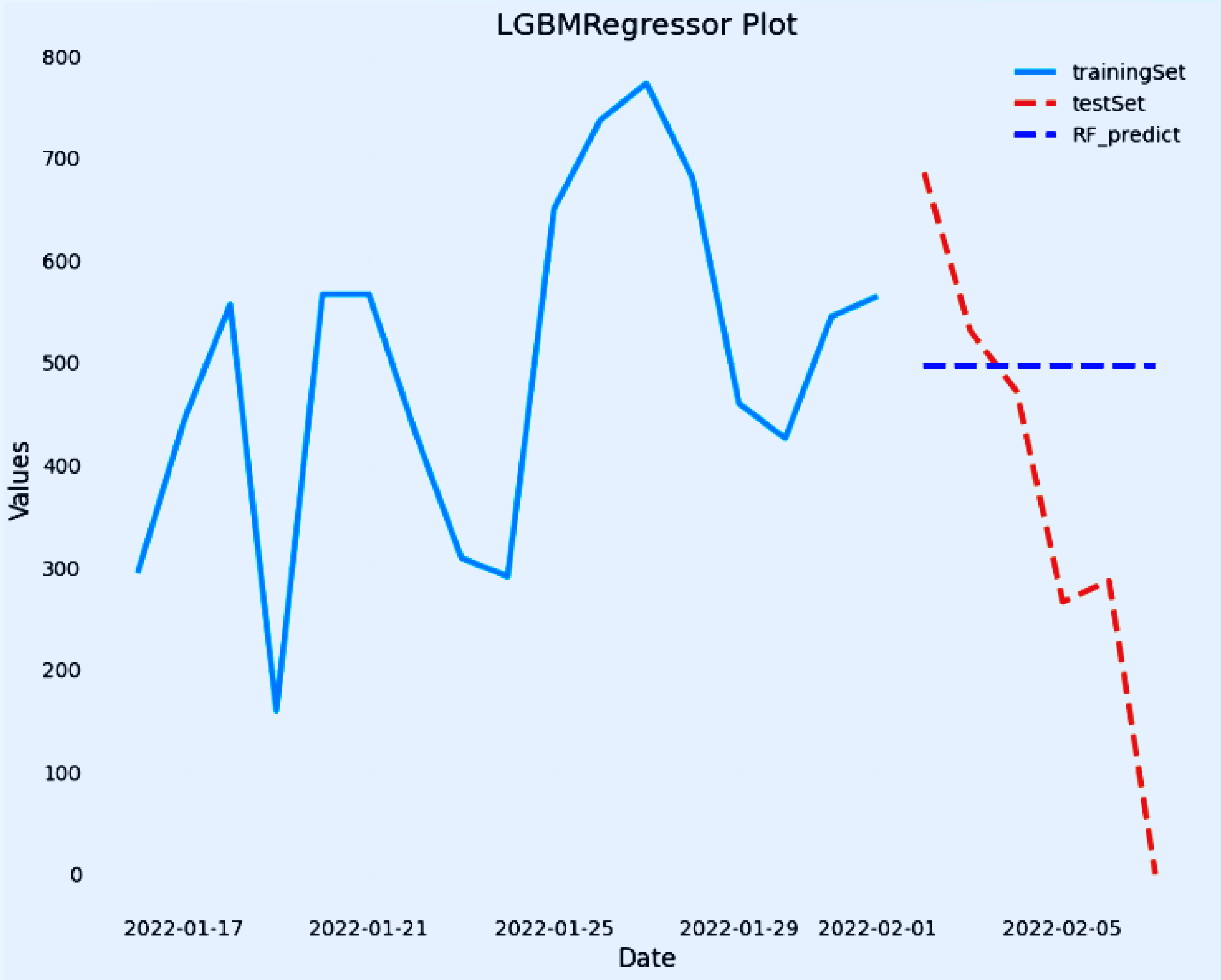
Tree-Based Models

LightGBM Regressor

LightGBM is an open-source gradient boosting framework developed by Microsoft that is designed to be efficient and scalable. LightGBM uses a technique called gradient-based one-side sampling (GOSS) to randomly sample a subset of the data, which reduces the number of data points that need to be used to build each tree, making the training process faster.

Applications / Use Cases

1. Predict which customers are likely to leave a company
2. Forecast sales for a business by analyzing historical data
3. Predict the energy consumption
4. Predict the severity of disease
5. Identify fraudulent claims by analyzing patterns in the data



1. Uses subsampling to keep the algorithm fast
2. Uses a histogram-based algorithm
3. Uses a novel technique called gradient-based one-side sampling which reduces the amount of data needed to be processed
4. Can take advantage of multiple CPU cores and GPUs to speed up the training process.

1. Susceptible to overfitting.
2. Generates complex trees which can make it difficult to interpret decision
3. It may not always perform well for small datasets, High-dimensional datasets with a small number of observations may be problematic.
4. does not support online learning like some other libraries, such as xgboost.

Supervised

Linear models

Linear Regression

Linear regression is a statistical model used to predict a continuous outcome variable (y) based on one or more predictor variables (x). The model is called "linear" because it assumes that the relationship between the predictor variables and the outcome is linear.

Applications / Use Cases

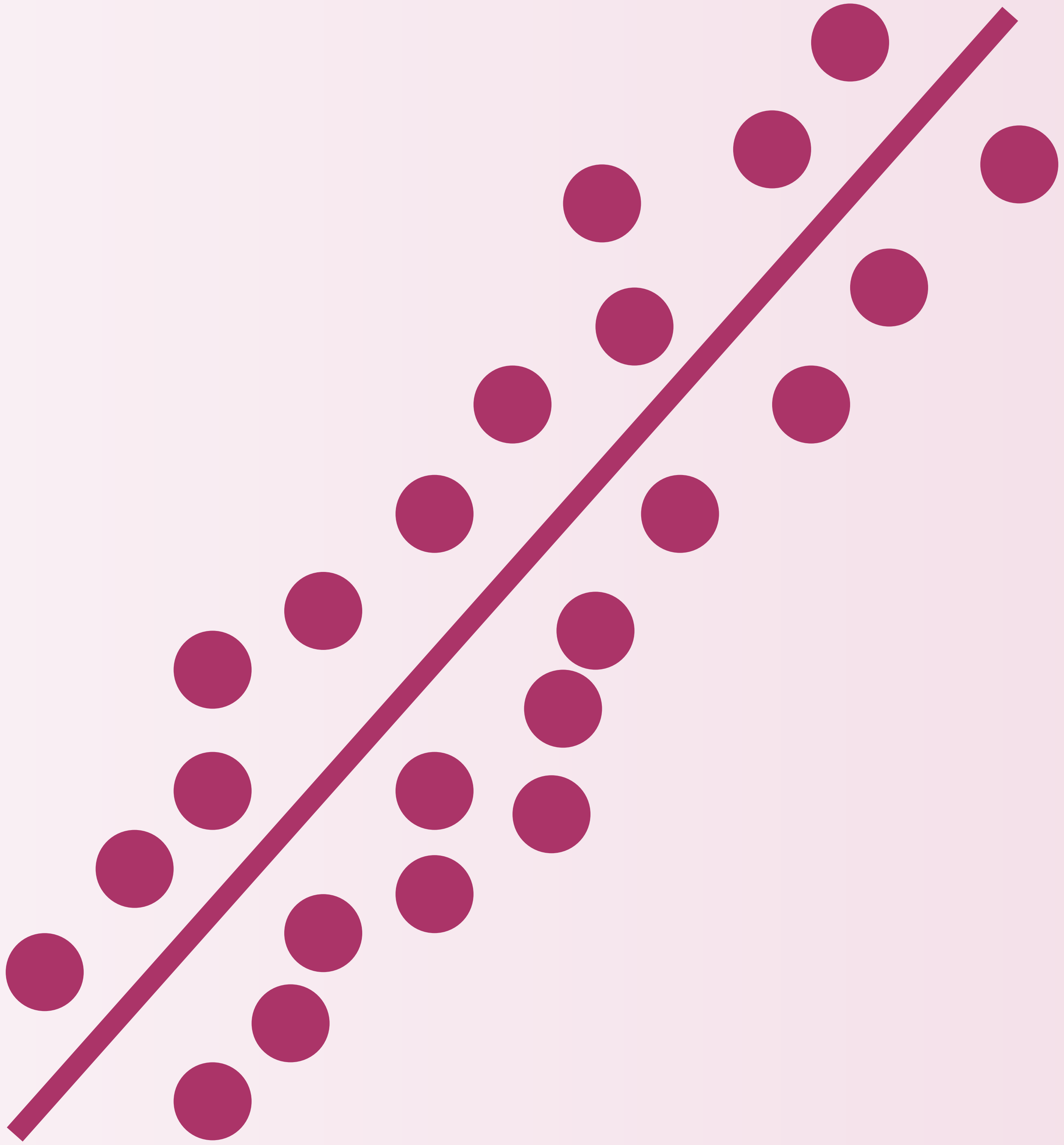
1. Predicting house prices
2. Forecasting stock prices
3. Estimating crop yields
4. Determining the impact of advertising on sales
5. Customer lifetime value prediction

👍

1. Simple to implement and efficient to train
2. Regularization can help to reduce overfitting.
3. Faster to train
4. Performs well when the dataset is linearly separable
5. Explainable method
6. Output coefficients produce interpretable findings.

👎

1. Assumes that the data is independent which is uncommon in real life.
2. Assumes Linearity between input & output
3. Prone to noise & overfitting
4. Underfitting is possible with tiny high-dimensional data.
5. Sensitive to outliers



Supervised

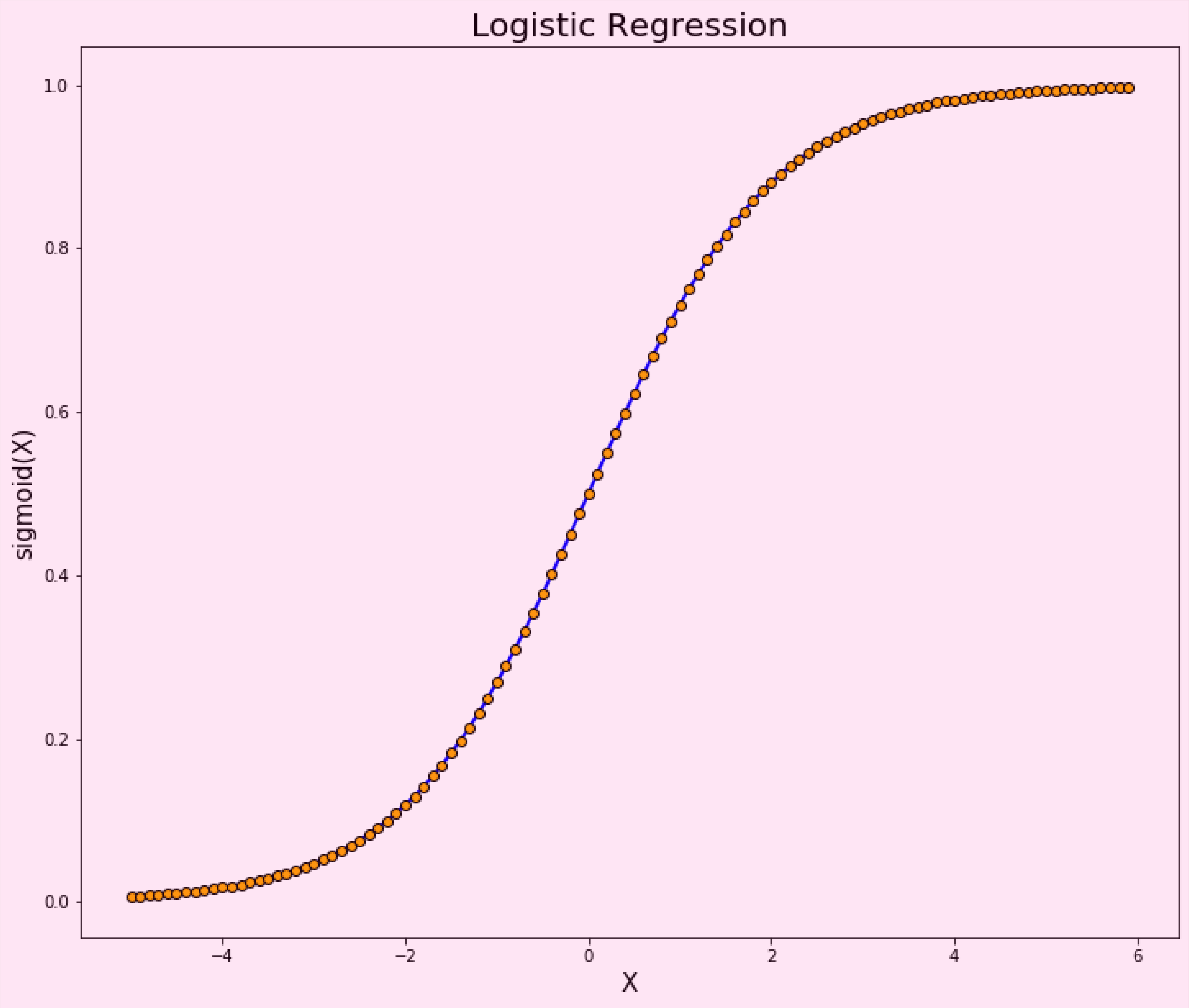
Linear models

Logistic Regression

Logistic Regression is a type of generalized linear model that is commonly used for binary classification, but can also be used for multi-class classification as well. The goal of logistic regression is to model the probability of an instance belonging to a certain class (e.g. a binary outcome of "1" or "0").

Applications / Use Cases

1. Determining whether an email is spam or not.
2. Predict the risk of default on a loan or credit card
3. Sentiment analysis
4. Determining the impact of advertising on sales
5. Multi-class classification problems



👍

1. Can be regularized to prevent overfitting
2. It is a probabilistic model (can provide a probability of outcome)
3. Can handle multiple predictor variables
4. Can handle both continuous & categorical variables as predictors
5. Very efficient and easy to implement
6. Easy to interpret

👎

1. Makes several assumptions about the data & If these assumptions are not met, the model's performance can be poor.
2. Requires a large sample size to achieve stable and meaningful results
3. Not powerful enough to model more complex relationships
4. Cannot provide prediction beyond the range

Supervised

Linear models

Ridge Regression

Ridge Regression is a variation of Linear Regression that adds a penalty term to the cost function to discourage the coefficients from becoming too large. The penalty term is a measure of the magnitude of the coefficients, and is controlled by a hyperparameter called the regularization strength, often represented by the Greek letter lambda (λ). When this parameter is set to zero, Ridge Regression becomes equivalent to Linear Regression.

Applications / Use Cases

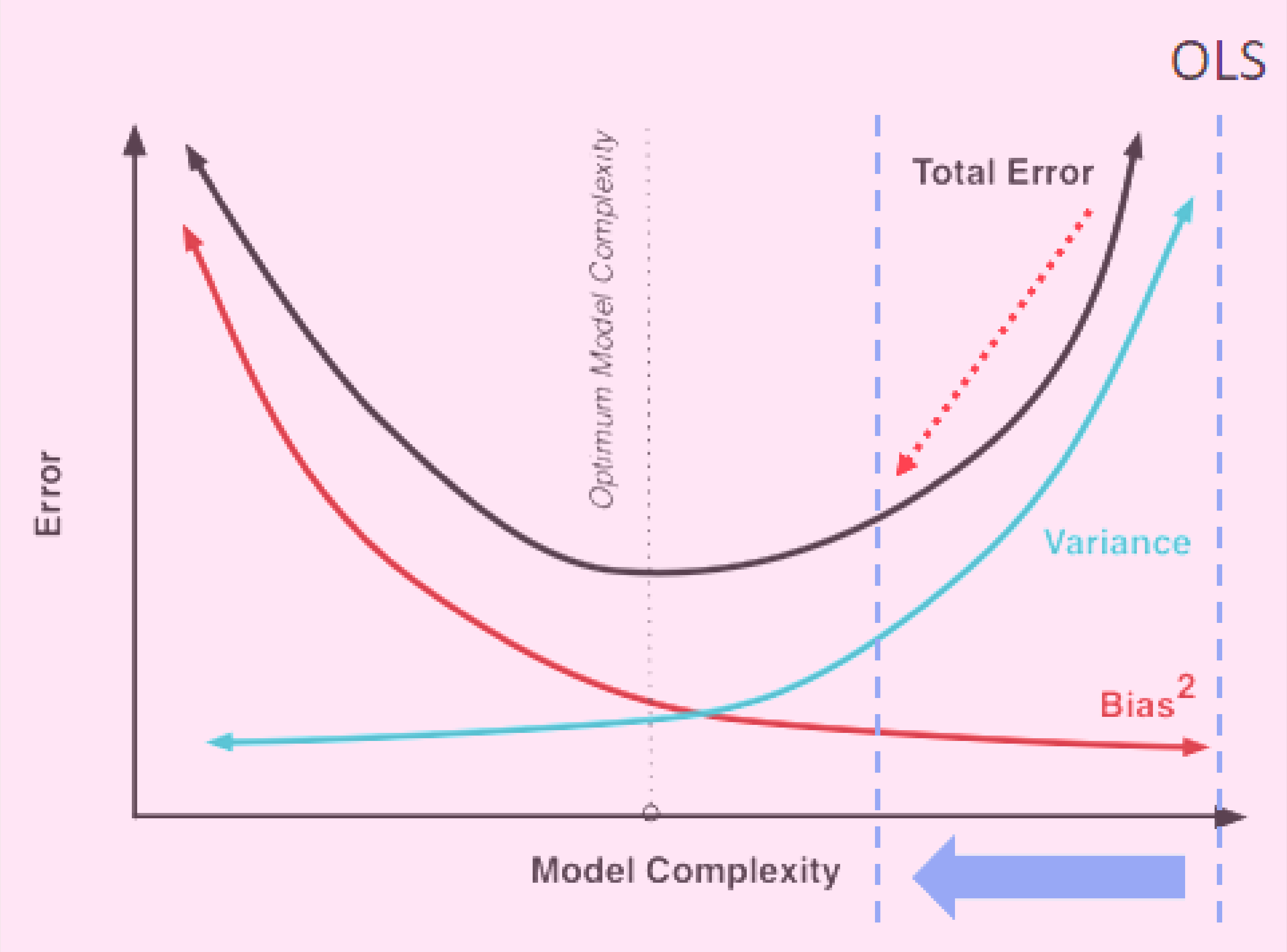
1. Predict about future economic trends.
2. Control the movement of robots
3. Object Detection/ image recognition/ image segmentation.
4. Predict future weather patterns.
5. Predict future health outcomes.

👍

1. Can handle high dimensional data well by shrinking coefficients
2. Helps to mitigate the problem of multicollinearity
3. Coefficients are bias towards zero
4. Can be used as a feature selection technique
5. Can be used as a regularization technique
6. Can work well with correlated predictor variables

👎

1. Not designed for classification problem
2. Not suitable for sparse data
3. Can be computationally expensive
4. Can be sensitive to the choice of lambda.
5. Can be sensitive to the scaling of the predictors



Lasso Regression

Lasso regression is a type of linear regression that uses regularization to reduce the number of features in a model by assigning zero coefficients to some of the features. Lasso stands for "Least Absolute Shrinkage and Selection Operator." It is similar to Ridge Regression, but uses the L1 norm of the coefficients to shrink them. The L1 norm tends to produce sparse solutions, where many coefficients are zero, while the L2 norm tends to produce solutions where the coefficients are small but non-zero.

Applications / Use cases

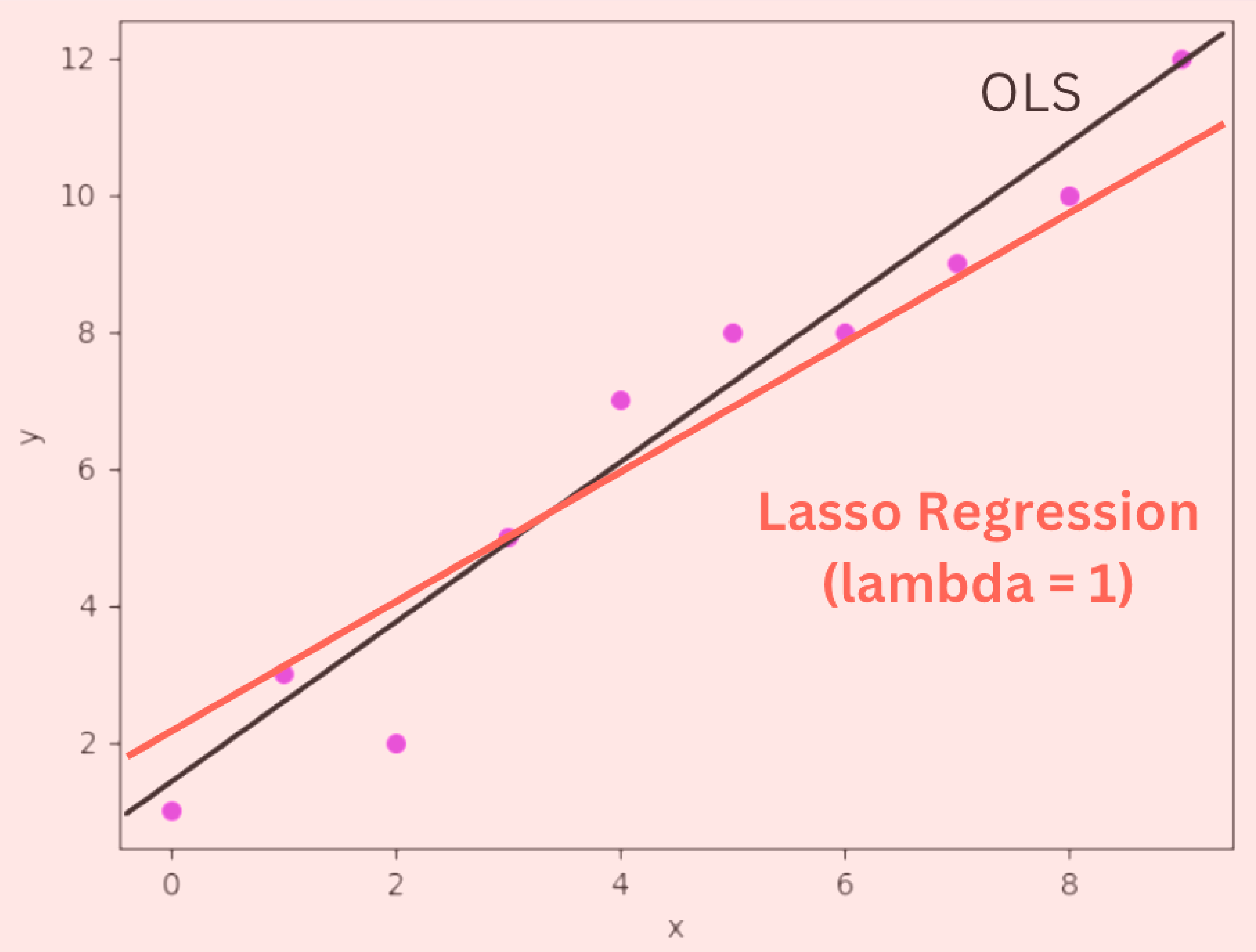
1. Identify the most relevant features in a dataset.
2. Time-series forecasting
3. Identify the most important pixels of an image
4. Predict customer behavior
5. engineering fields to predict a response variable

👍

1. Simple to implement and efficient to train
2. Regularization can help to reduce overfitting.
3. Faster to train
4. Performs well when the dataset is linearly separable
5. Explainable method
6. Output coefficients produce interpretable findings.

👎

1. Assumes that the data is independent which is uncommon in real life.
2. Assumes Linearity between input & output
3. Prone to noise & overfitting
4. Underfitting is possible with tiny high-dimensional data.
5. Sensitive to outliers



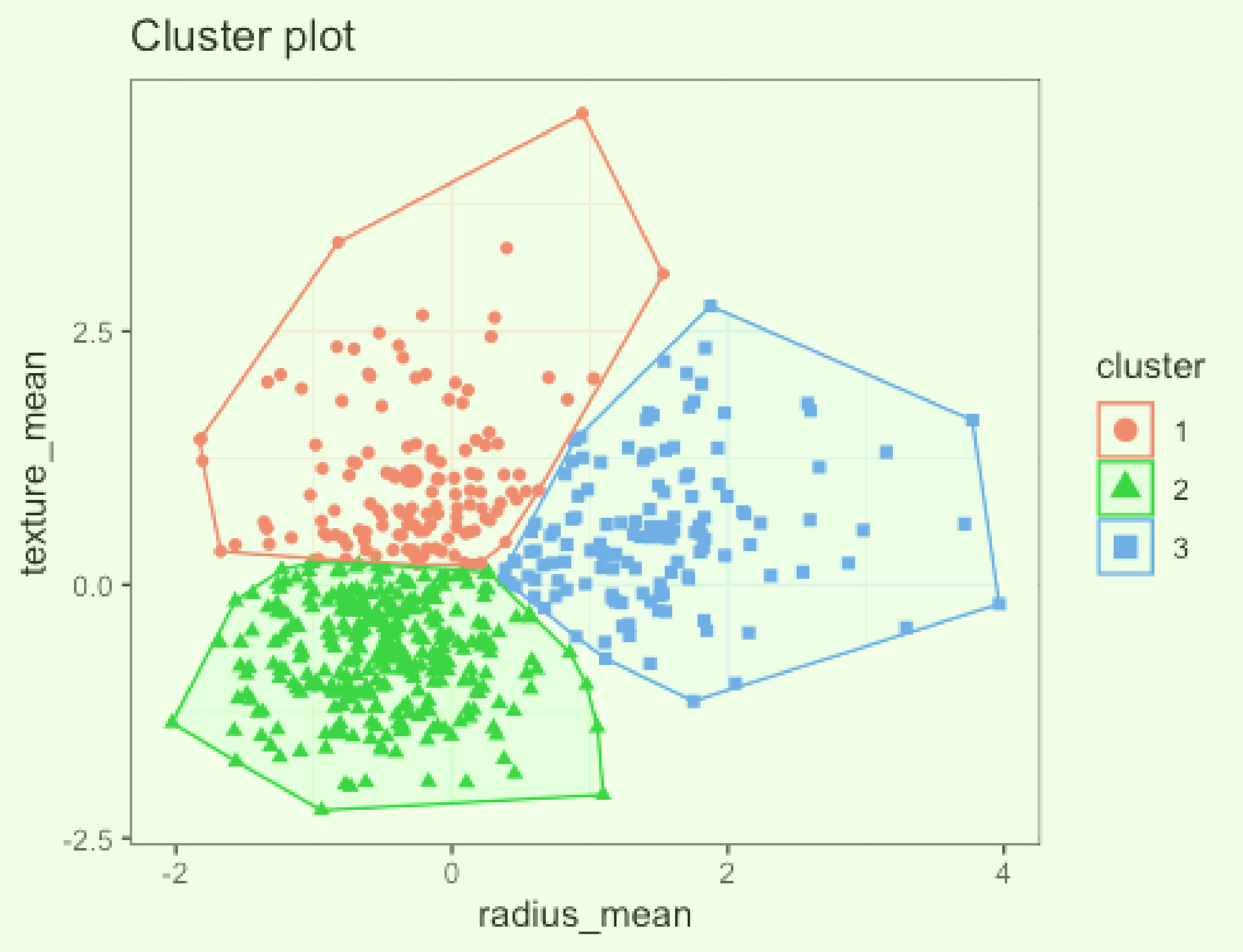
Unsupervised Clustering

K-means Clustering

The algorithm works by first randomly initializing k centroids, and then iteratively assigning each data point to the closest centroid and updating the centroid's position to be the mean of the points assigned to it. This process continues until the assignments of points to clusters no longer change or a maximum number of iterations is reached.

Applications / Use Cases

1. Reduce the number of colors in an image Compression
2. Segment customers in a market based on their purchasing habits
3. Group similar documents together
4. Anomaly detection
5. Gene expression analysis



---

1. Simple to understand and implement
2. Very fast, even for large datasets
3. Can handle various types of data
4. Cluster any kind of data
5. Global Convergence

---

1. Assumes that clusters are spherical in shape, which may not always be the case in real-world data sets.
2. Assumes that all clusters have same number of data points
3. Sensitive to initial starting point
4. Requires the number of clusters to be pre-specified
5. Doesn't handle missing data

Unsupervised Clustering

Hierarchical

Hierarchical clustering is a method of cluster analysis in which a hierarchy of clusters is constructed. The hierarchy is represented as a tree-like diagram called a dendrogram. There are two main types of hierarchical clustering: agglomerative and divisive.

Applications / Use Cases

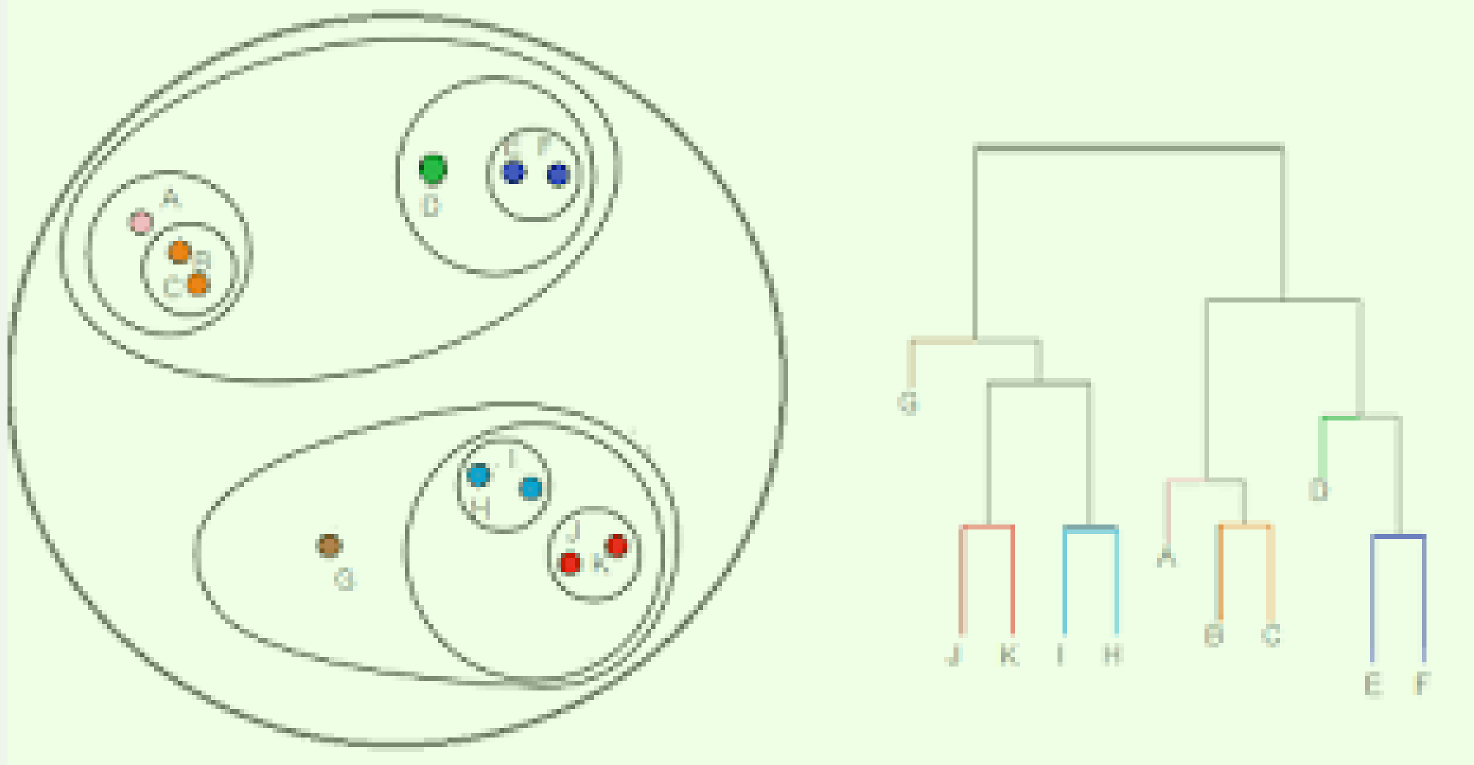
1. Segment an image into different regions of similar pixel intensity
2. Often used to group genes based on their expression patterns
3. Can be used to group individuals in a social network
4. Group documents based on their contents
5. Group customers based on their demographics

👍

1. Can be useful in situations where it is not possible or practical to use other methods
2. Provide great flexibility and ability to handle multiple types of data
3. Can handle large datasets and can be parallelizable.
4. Handles missing or noisy data
5. Does not require the no. of clusters to be specified in advance

👎

1. Can be computationally expensive for large datasets
2. Results can be sensitive to the linkage criterion used
3. Structure of the clusters can be difficult to interpret
4. Relies on metric distance measure between the elements of dataset
5. Doesn't give any quantitative results



Unsupervised

Clustering

Gaussian Mixture Models

A Gaussian Mixture Model (GMM) is a probabilistic model that assumes that the underlying data is generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Each Gaussian component in the mixture represents a cluster, and the task of GMM is to estimate the parameters of the Gaussian distributions that best fit the data.

Applications / Use Cases

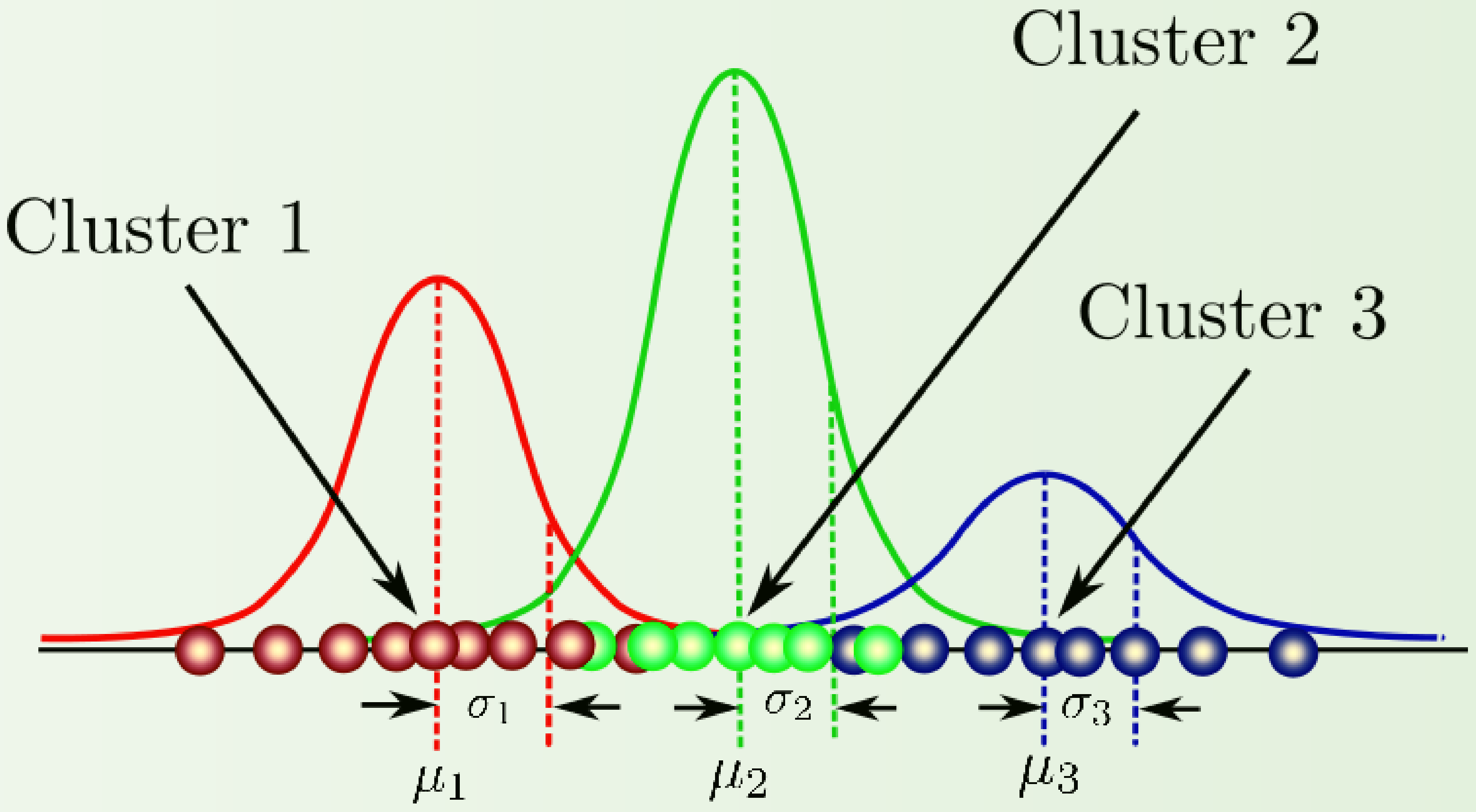
1. Gait recognition
2. Handwriting recognition
3. Can be used to segment an image into different regions
4. Determining the impact of advertising on sales
5. Object tracking in video sequences

👍

1. Trained using the Expectation-Maximization (EM) algorithm
2. Can handle high-dimensional data and large datasets
3. Assumes that the data is generated from a mixture of Gaussian distributions
4. Can handle high-dimensional data and large datasets,
5. Can model a wide range of data distributions

👎

1. Sensitive to the initial values of the parameters
2. Can converge to a local optimum solution instead of global one
3. Can be computationally intensive
4. Determining the optimal number of components can be difficult
5. Assumes that the covariance matrix of each Gaussian component is spherical



Unsupervised

Association

Apriori Algorithm

Apriori algorithm is an association rule mining algorithm that is used to find frequent item sets in a large dataset, with the goal of discovering underlying associations and relationships among the items. It is based on the "Apriori property", which states that any subset of a frequent item set must also be frequent. The Apriori algorithm works by first identifying all the item sets that have a support count above a certain threshold and then using these frequent item sets to generate a set of association rules.

Applications / Use Cases

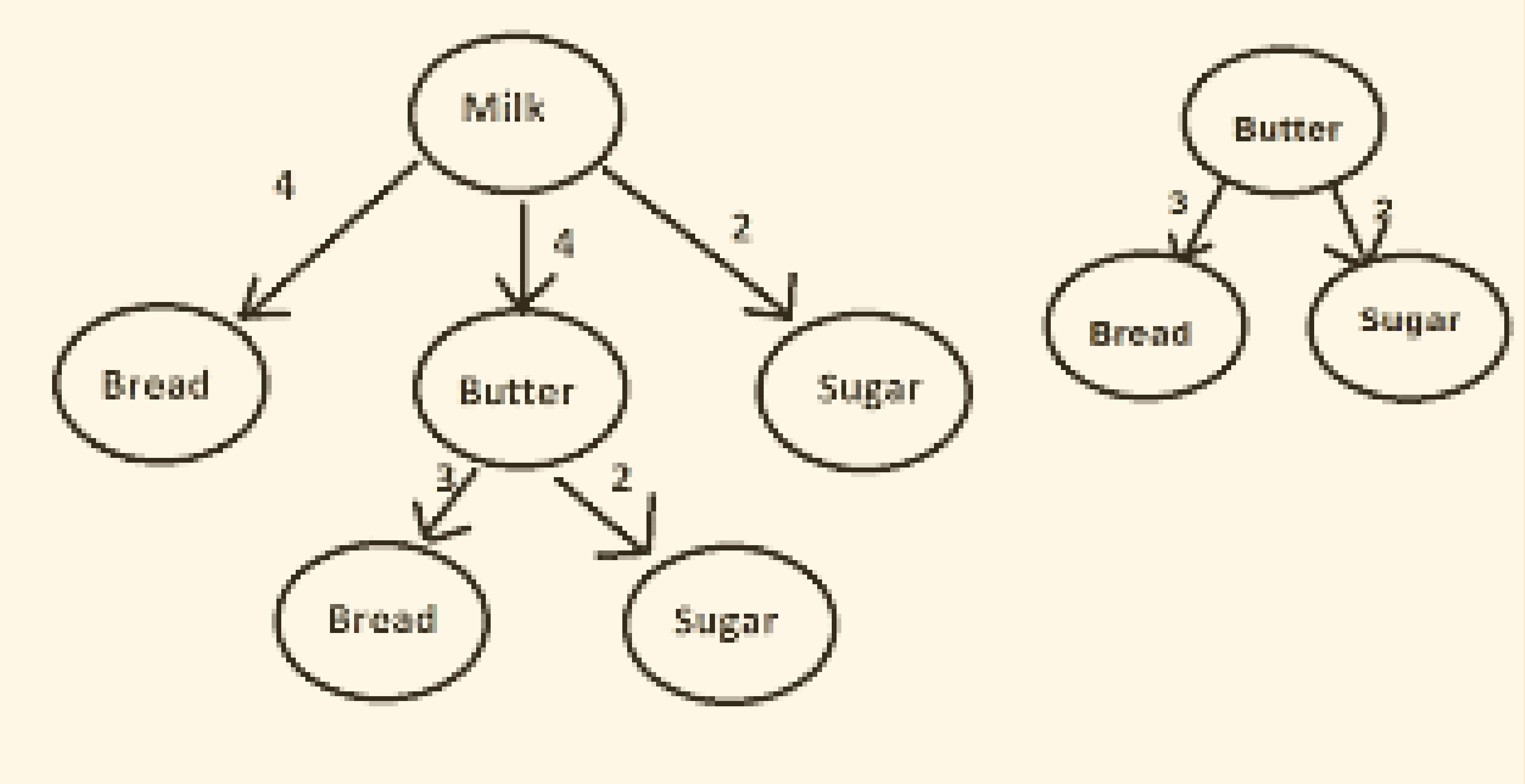
1. Can be used to analyze customer purchase data to identify items that are frequently bought together
2. Can be used to analyze customer call data to identify patterns of calling behavior
3. Can be used to analyze patient data

👍

1. Simple and intuitive algorithm
2. Can be used to find association rules and frequent item sets
3. Uses the "Apriori property" to significantly reduce space
4. Can easily handle both categorical and continuous data.
5. Generates simple association rules

👎

1. Requires a large amount of memory to store them
2. Does not take into account the correlation between the items
3. Only handles binary data, either 0 or 1
4. Not be suitable for continuous data
5. Generates too many association rules



Thanks for reading!

Chai Knows

in Tech series