PARAMETRIC AND NON-PARAMETRIC LEARNING METHODS

SUBMITTED BY:

SRISHTI SINGH-A023119820035

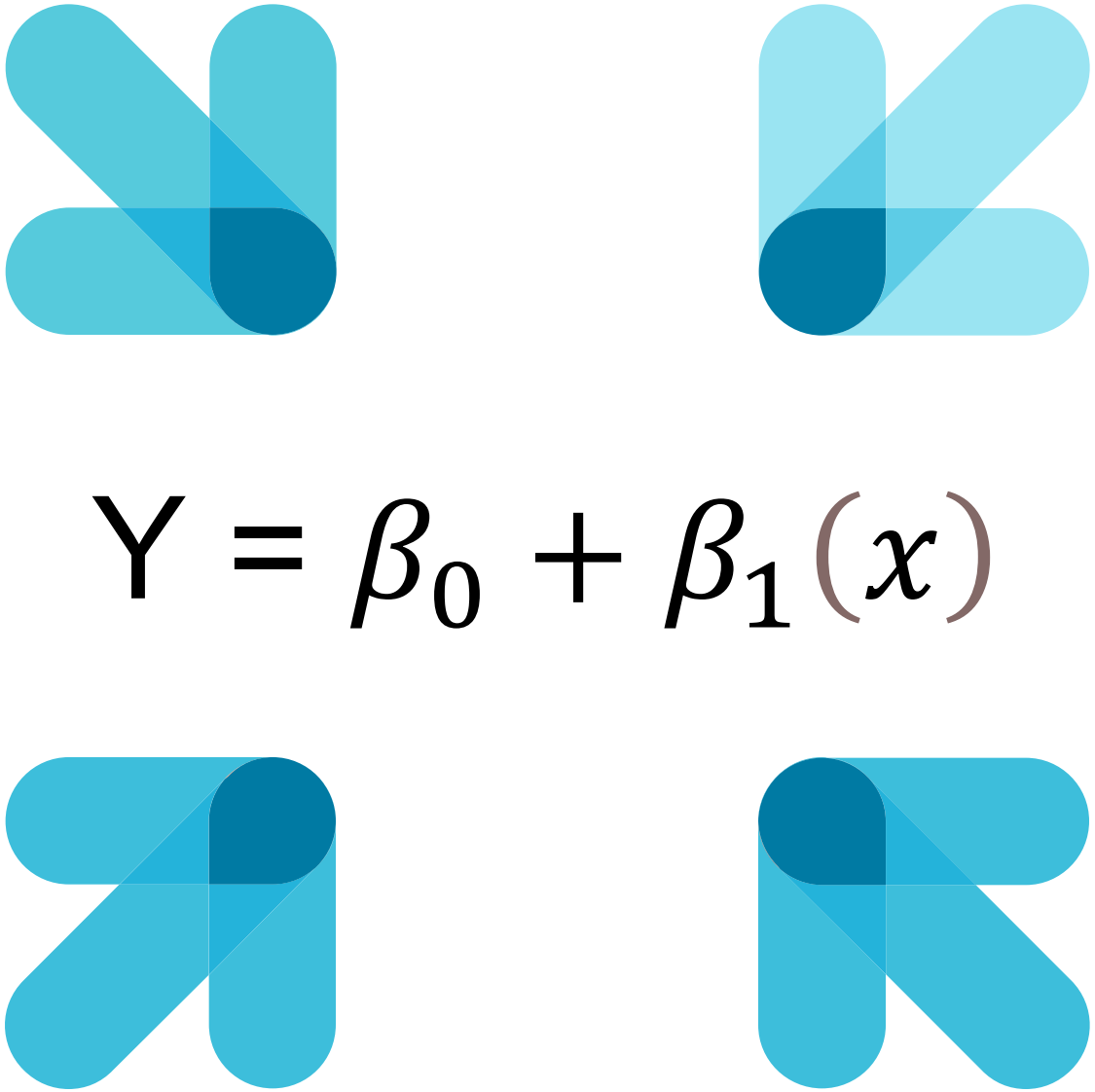
NIKITA-A023119820034

PARAMETRIC MODEL OF MACHINE LEARNING

A learning model that summarizes data with a set of fixed-size parameters

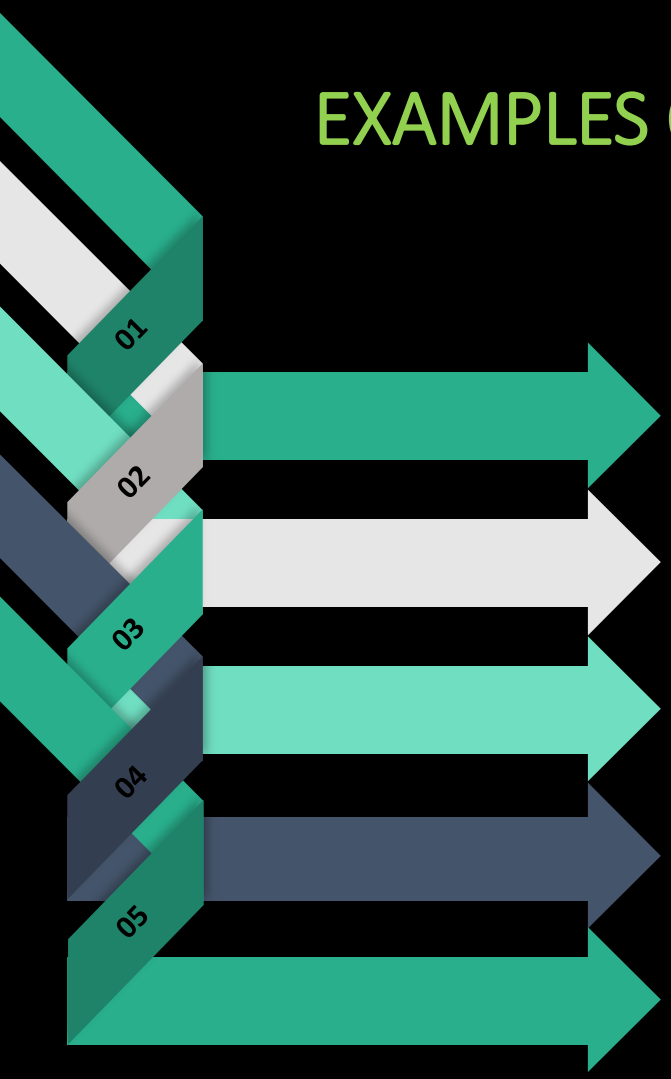
(independent on the number of instances of training).

Parametric machine learning algorithms are which optimizes the function to a known form.



$$Y = \beta_0 + \beta_1(x)$$

EXAMPLES OF PARAMETRIC MACHINE LEARNING ALGORITHMS



LINEAR REGRESSION

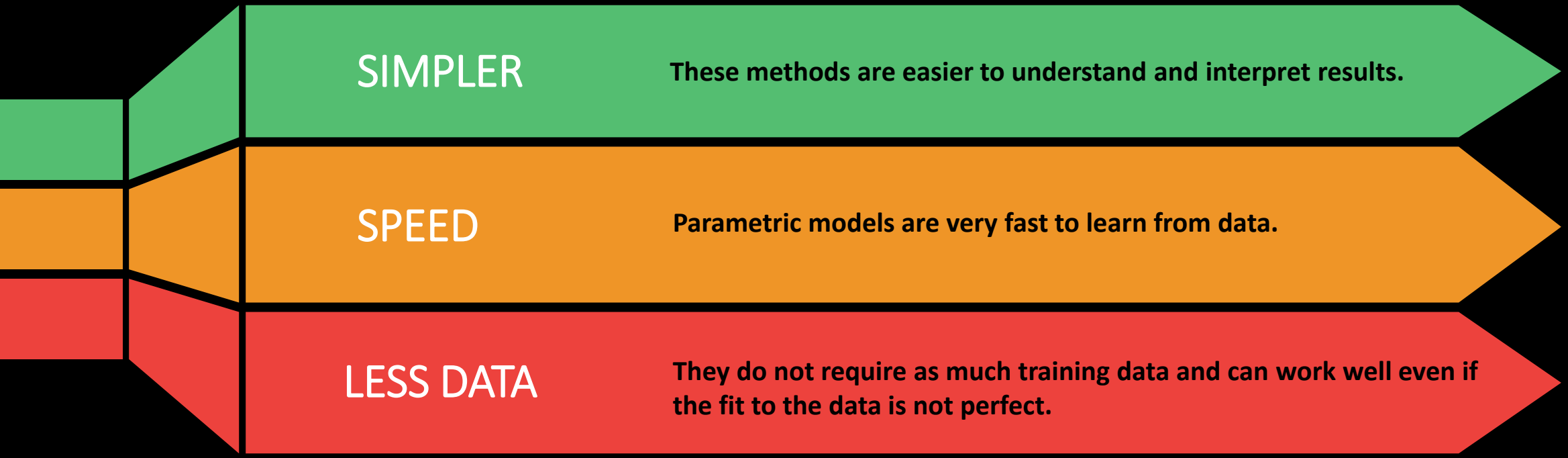
LINEAR DISCRIMINANT ANALYSIS

PERCEPTRON

NAIVE BAYES

SIMPLE NEURAL NETWORKS

BENEFITS OF PARAMETRIC MACHINE LEARNING ALGORITHM



LIMITATIONS OF PARAMETRIC MACHINE LEARNING ALGORITHM

CONSTRAINED

By choosing a functional form these methods are highly constrained to the specified form.

LIMITED COMPLEXITY

The methods are more suited to simpler problems.

POOR FIT

In practice the methods are unlikely to match the underlying mapping function.

WHAT IS LINEAR REGRESSION

Linear regression is a statistical model that examines the linear relationship between two (Simple Linear Regression) or more (Multiple Linear Regression) variables — a dependent variable and independent variable(s).

EXAMPLE

The weight of the person is linearly related to their height. So, this shows a linear relationship between the height and weight of the person. According to this, as we increase the height, the weight of the person will also increase.

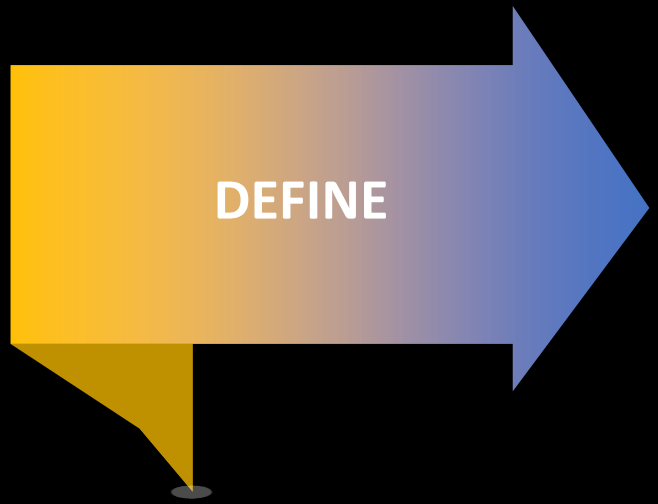
LINEAR REGRESSION

It is not necessary that one variable is dependent on others, or one causes the other, but there is some critical relationship between the two variables. In such cases, we use a scatter plot to simplify the strength of the relationship between the variables.

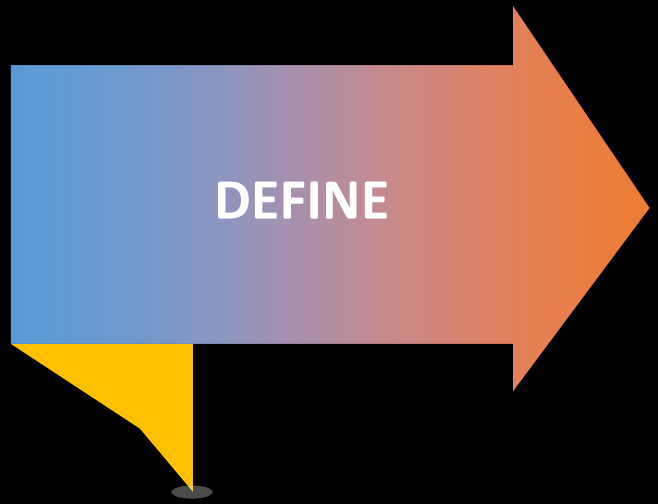
LINEAR REGRESSION

If there is no relation or linking between the variables then the scatter plot does not indicate any increasing or decreasing pattern. In such cases, the linear regression design is not beneficial to the given data.

LOGISTIC REGRESSION



Logistic regression is a statistical classification model, it gives you the likelihood that an observation belongs to a specific class. Logistic regression predicts the output of a categorical dependent variable.



It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, **it gives the probabilistic values which lie between 0 and 1**.

1) The dependent variable must be categorical in nature.
2) The independent variable should not have multi-collinearity.

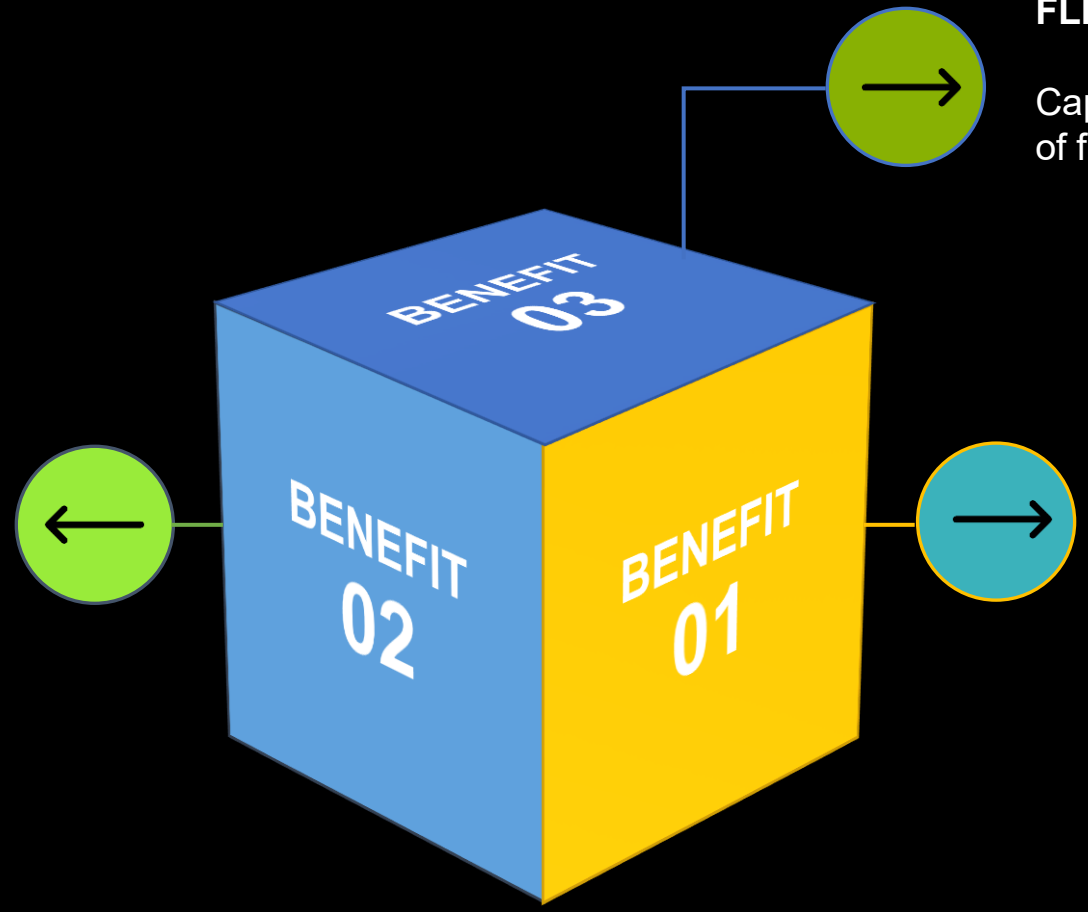
NON-PARAMETRIC METHOD OF MACHINE LEARNING

Non-parametric methods are good when you have a lot of data and no prior knowledge, and when you don’t want to worry too much about choosing just the right features. Nonparametric methods seek to best fit the training data in constructing the mapping function, whilst maintaining some ability to generalize to unseen data. As such, they are able to fit a large number of functional forms

BENEFITS OF NON-PARAMETRIC LEARNING ALGORITHM:

FLEXIBILITY:

Capable of fitting a large number of functional forms.



PERFORMANCE:

Can result in higher performance models for prediction.

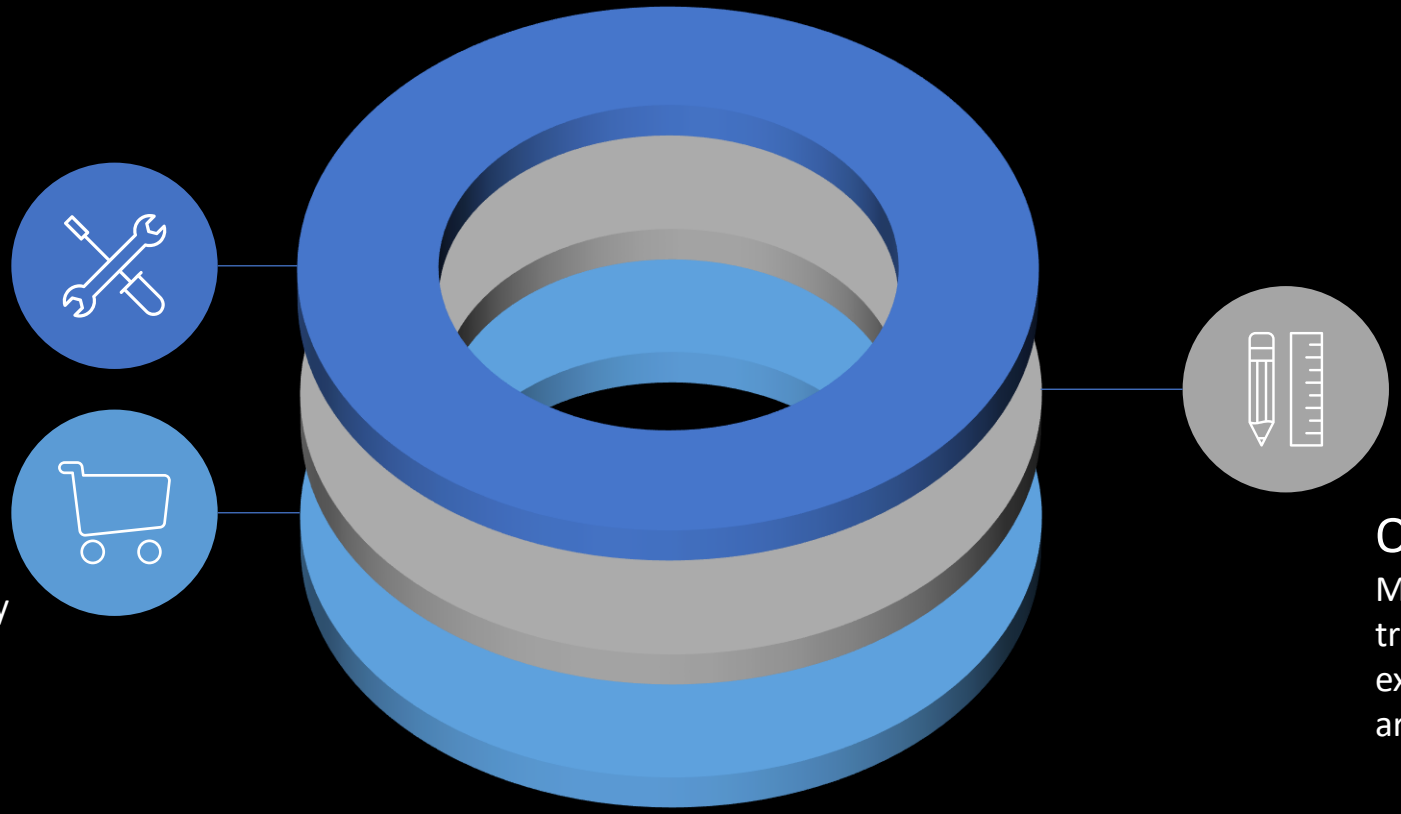
POWER:

No assumptions (or weak assumptions) about the underlying function.

LIMITATIONS OF NON-PARAMETRIC LEARNING ALGORITHM:

MORE DATA:

Require a lot more training data to estimate the mapping function.



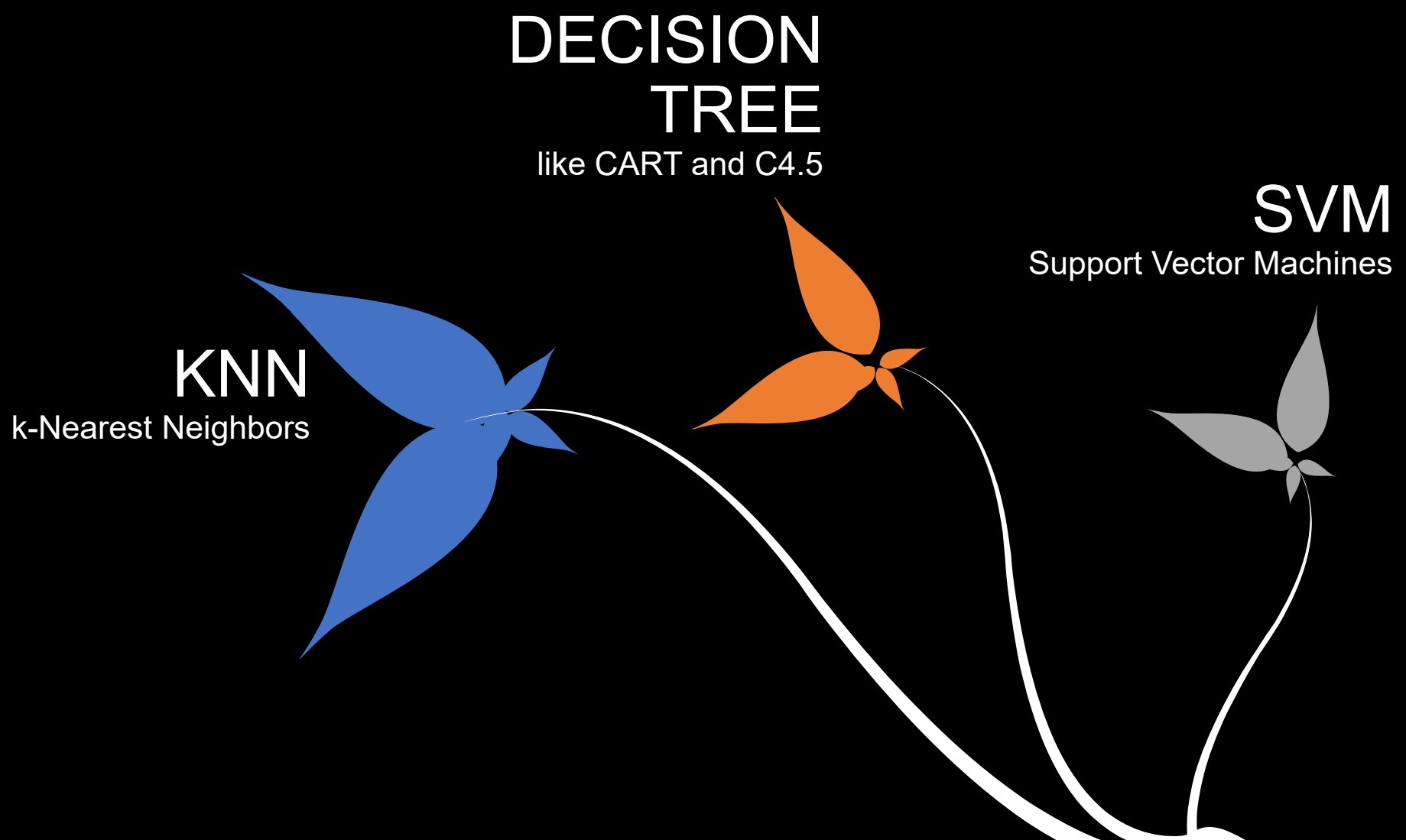
SLOWER:

A lot slower to train as they often have far more parameters to train.

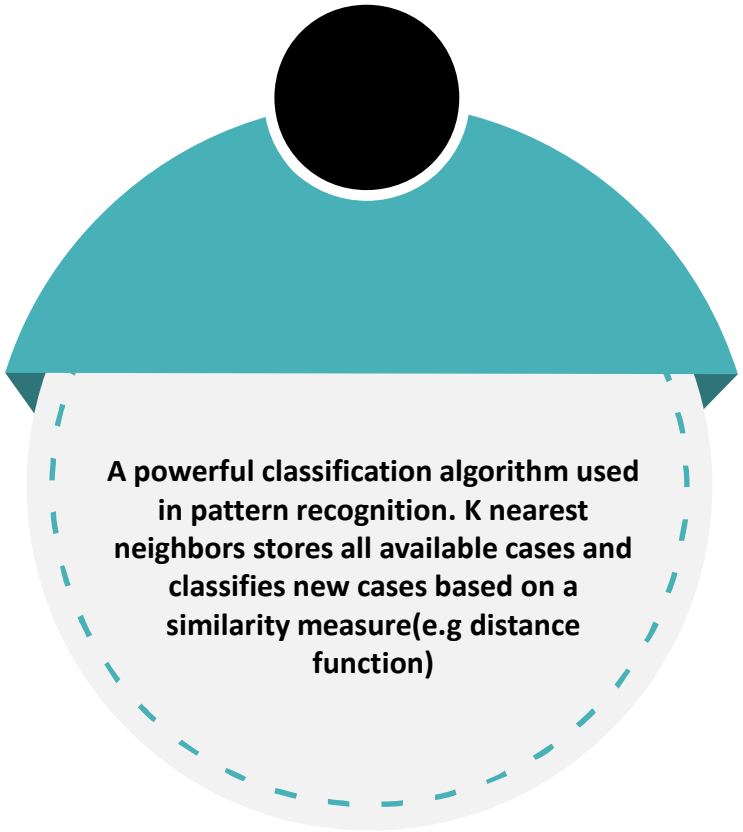
OVERFITTING:

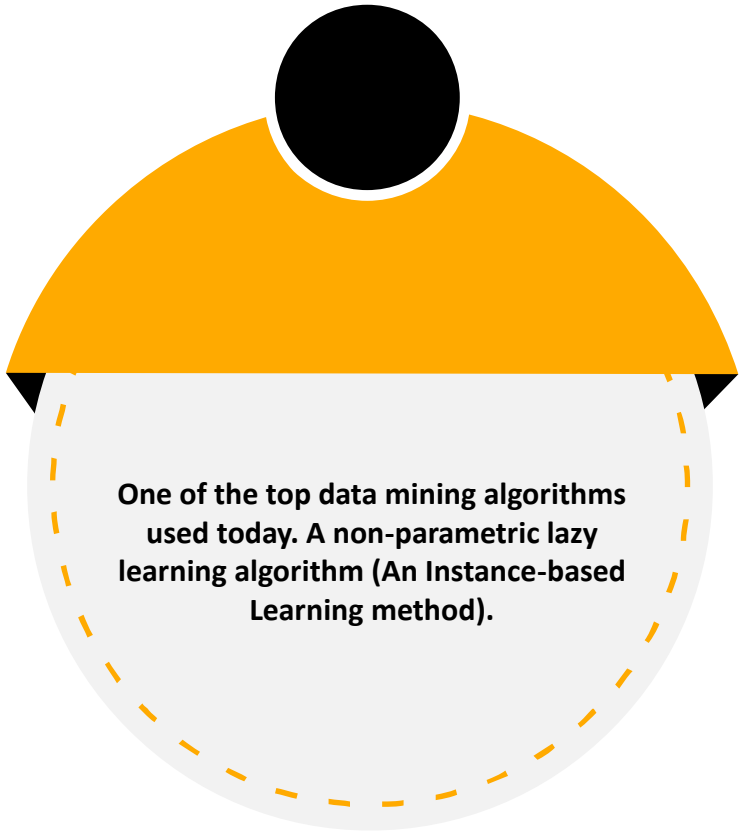
More of a risk to overfit the training data and it is harder to explain why specific predictions are made.

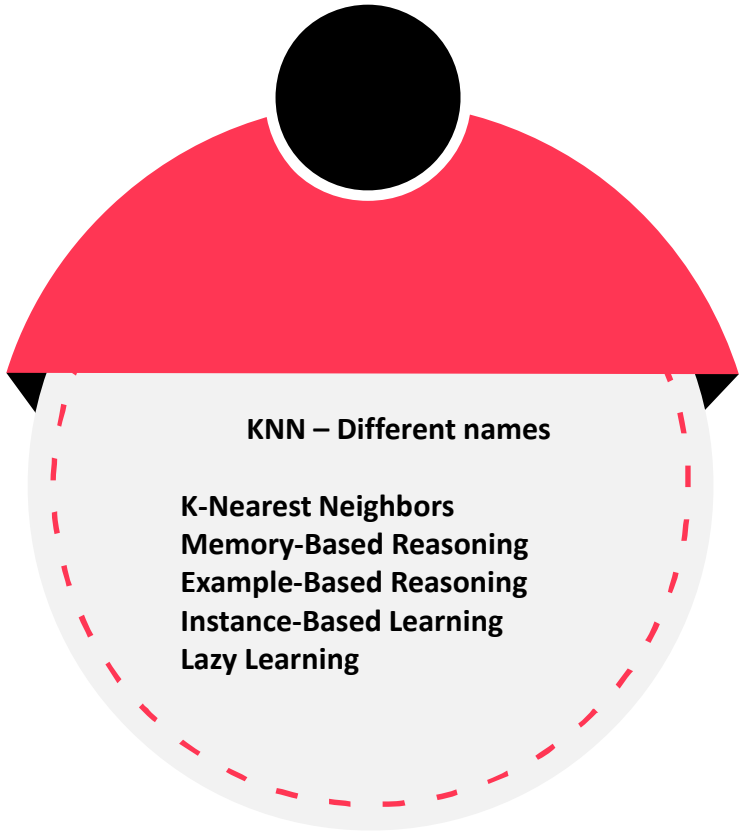
EXAMPLES OF NON-PARAMETRIC MACHINE LEARNING ALGORITHM



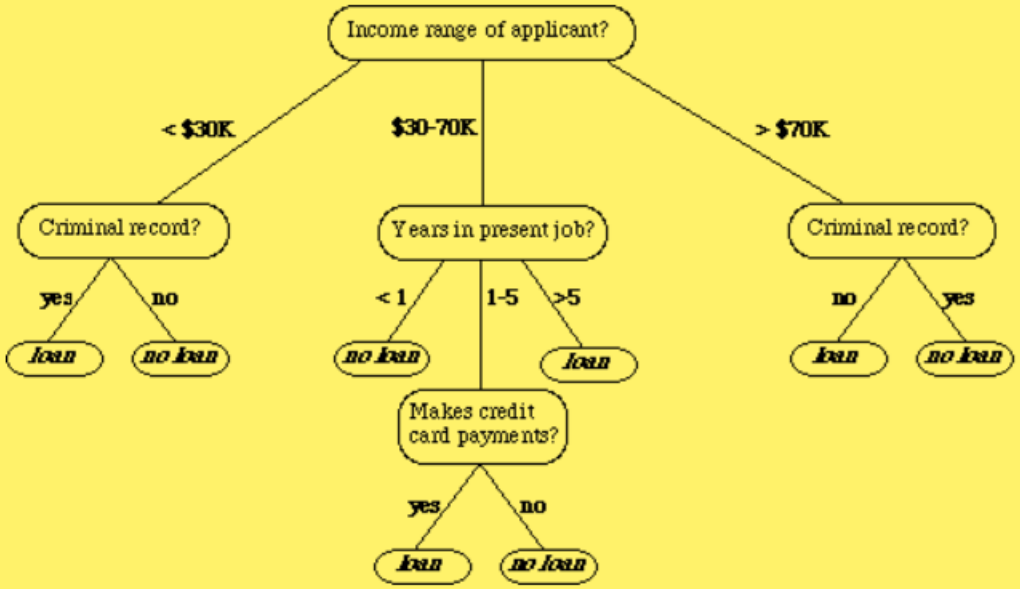
K-Nearest Neighbor







What is a Decision Tree?



A decision tree is a largely used non-parametric effective machine learning modeling technique for regression and classification problems. To find solutions a decision tree makes sequential, hierarchical decision about the outcomes variable based on the predictor data.

DISADVANTAGES

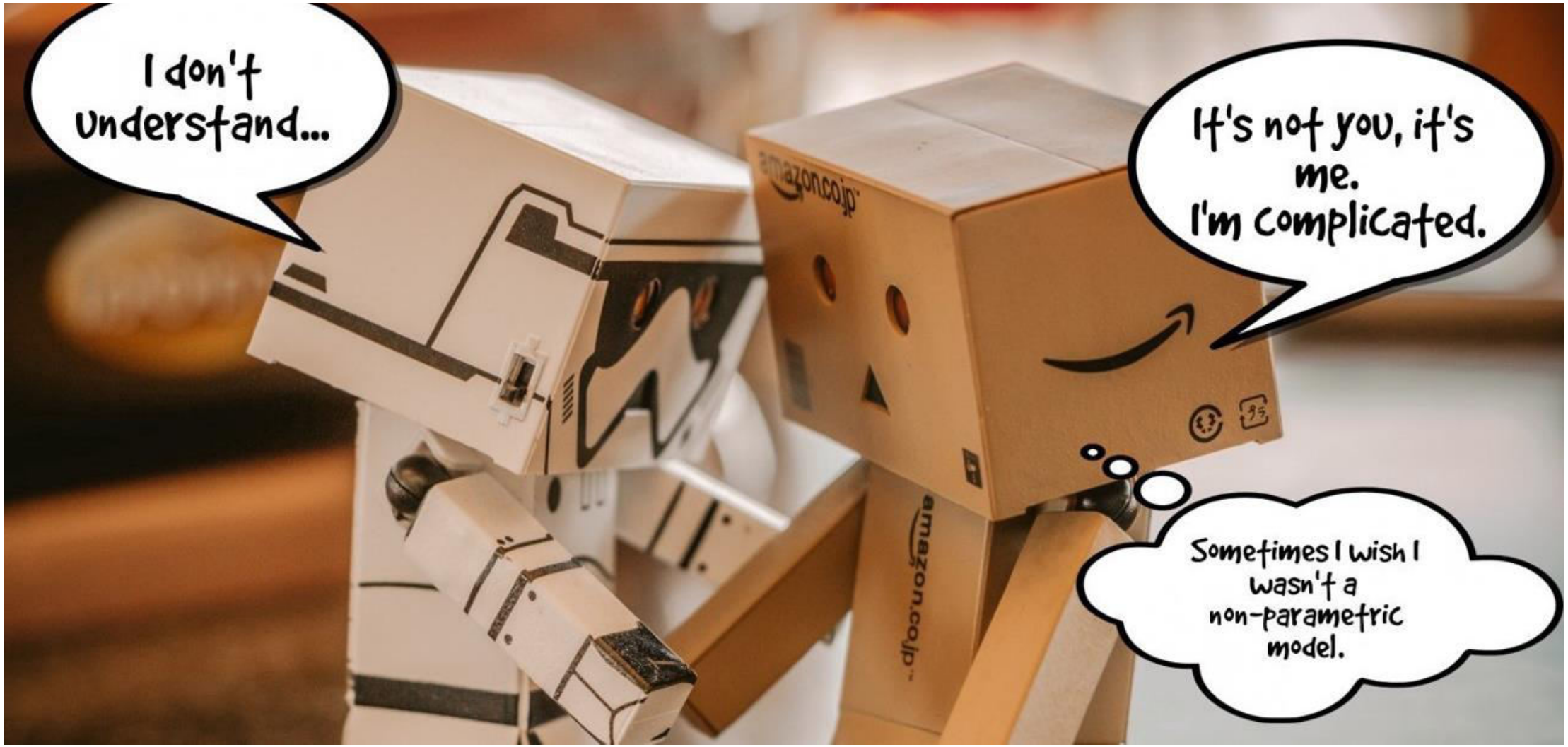
- Decision-tree learners can create over-complex trees that do not generalize well from the training data.
- It is locally optimized using a greedy algorithm where we cannot guarantee a return to the globally optimal decision tree.

ADVANTAGES

- They are incredibly simple to understand due to their visual representation, they require very little data, can handle qualitative and quantitative data,
- it can be validated using statistical sets,
- it can handle large amounts of data and it is quite computationally inexpensive.

SUMMARY

Parametric model	Non-parametric model
Constant number of parameters, independent of training data	Number of parameters grows with the number of training samples
Strong assumption about the training data	No assumption about the training data
Fewer training samples required	Many training samples required
Fast training, fast inference	Slow training and slow inference
Examples: Linear regression and logistic regression	Examples: Decision trees and k-nearest neighbors


I don't understand...
It's not you, it's me.
I'm complicated.
Sometimes I wish I wasn't a non-parametric model.
amazon.co.jp
amazon.co.jp