greatlearning
Learning for Life

Machine Learning

Introduction to machine learning

greatlearning
Learning for Life

Machine Learning & Data Science

1. Machine learning is part of a larger discipline called Data Science

2. Data science is the process of applying science and domain expertise to data to extract useful information from data.

3. It includes application of all the statistical and mathematical tools and techniques to glean out the useful information from data using machine learning

Introduction to machine learning

greatlearning
Learning for Life

"[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed." Arthur Samuel 1959:

Image Source : https://www.icecreamlabs.com

Introduction to machine learning

greatlearning
Learning for Life

What is machine learning?

1. Is a process of enabling automated systems to learn to do tasks based on well defined statistical and mathematical methods

2. The ability to do the tasks is embodied in form of a model which is the result of the learning process

3. The model represents the process which generated the data used to build the model

4. The data used is expected to represent the long term behaviour of the process

5. The more representative data is of the real world in which the process is executed, the better the model would be

6. The challenge is how to get make true representative data sets

Introduction to machine learning

greatlearning
Learning for Life

What do machine learning algorithms do?

1. Search through data to look for patterns
2. Patterns in form of trends, cycles, associations, classes etc.
3. Express these patterns as mathematical structures such as probability equations or polynomial equations
4. These expressions (end result of running the algorithms) is broadly called models

Introduction to machine learning

greatlearning
Learning for Life

When is machine learning useful ?

1. Cannot express our knowledge about patterns as a program. For e.g. Character recognition or natural language processing
2. Do not have an algorithm to identify a pattern of interest. For e.g. In spam mail detection
3. Too complex and dynamic. For e.g. Weather forecasting
4. Too many permutations and combinations possible. For e.g. Genetic code mapping
5. No prior experience or knowledge. For e.g. Mars rover
6. Patterns hidden in humongous data. For e.g. Recommendation system

Introduction to machine learning

greatlearning
Learning for Life

Where are machine learning based systems used (examples only)

1. Fraud detection
2. Sentiment analysis
3. Credit risk management
4. Prediction of equipment failures
5. New pricing models / strategies
6. Network intrusion detection
7. Pattern and image recognition
8. Email spam filtering

Introduction to machine learning

greatlearning
Learning for Life

Machine Learning Pre-requisites

1. Rich set of data representing the real world

2. Knowledge and skills in
   a. Maths and statistics
   b. Programming (Python, R, Java, Go)
   c. Domain knowledge

greatlearning
Learning for Life

Real World as Mathematical Space

Introduction to machine learning

greatlearning
Learning for Life

Machine learning happens in mathematical space / feature space:

1. A data set representing the real world, is a collection attributes that define an entity

2. Each entity is represented as one record / line in the data set

Attributes / Dimensions

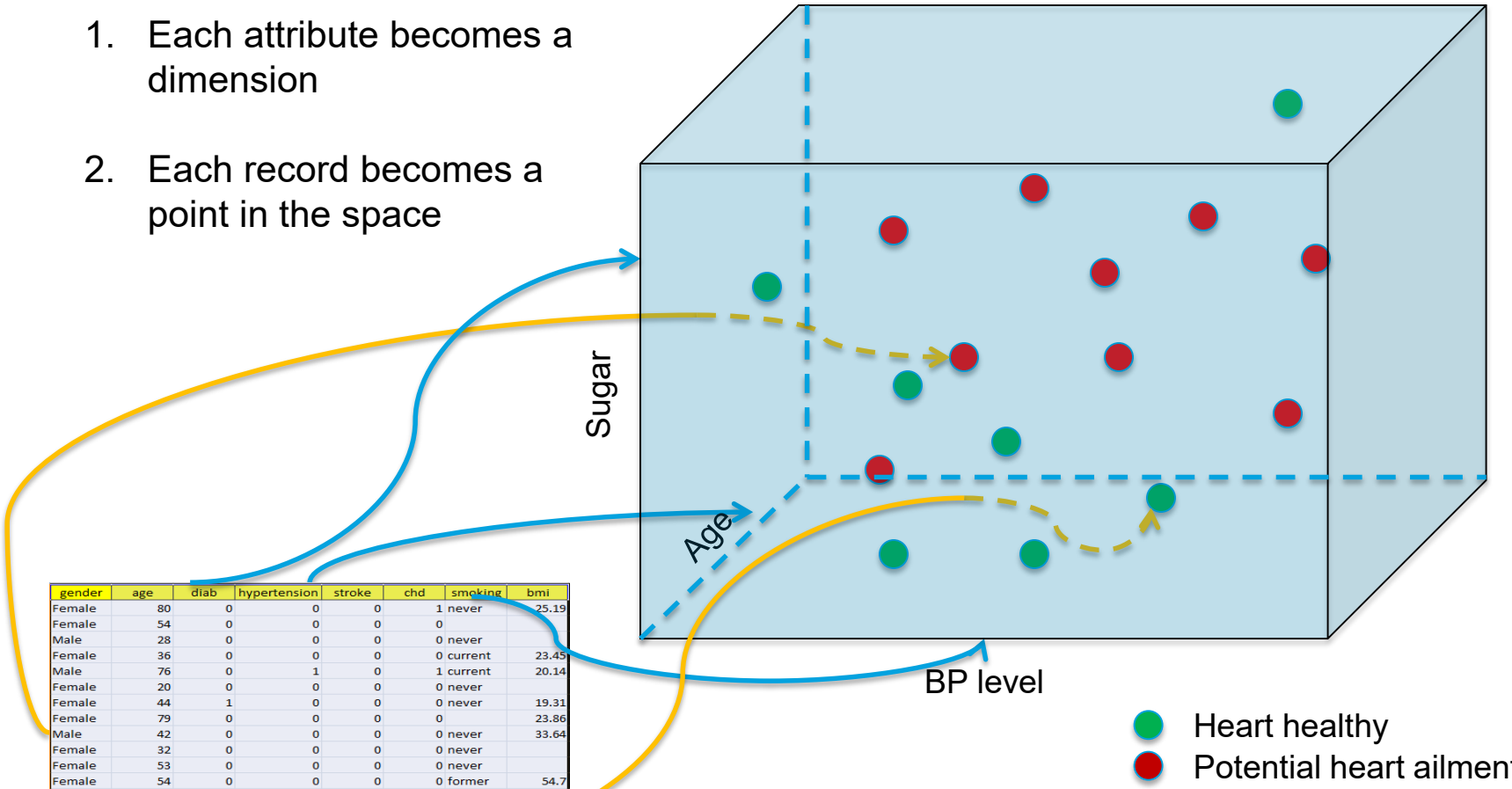
gender	age	diab	hypertension	stroke	chd	smoking	bmi
Female	80	0	0	0	1	never	25.19
Female	54	0	0	0	0		
Male	28	0	0	0	0	never	
Female	36	0	0	0	0	current	23.45
Male	76	0	1	0	1	current	20.14
Female	20	0	0	0	0	never	
Female	44	1	0	0	0	never	19.31
Female	79	0	0	0	0		23.86
Male	42	0	0	0	0	never	33.64
Female	32	0	0	0	0	never	
Female	53	0	0	0	0	never	
Female	54	0	0	0	0	former	54.7
Female	78	0	0	0	0	former	36.05
Female	67	0	0	0	0	never	25.69
Female	76	0	0	0	0		
Male	78	0	0	0	0		
Male	15	0	0	0	0	never	30.36
Female	42	0	0	0	0	never	24.48

Introduction to machine learning

greatlearning
Learning for Life

Machine learning happens in mathematical space / feature space:

1. Each attribute becomes a dimension
2. Each record becomes a point in the space



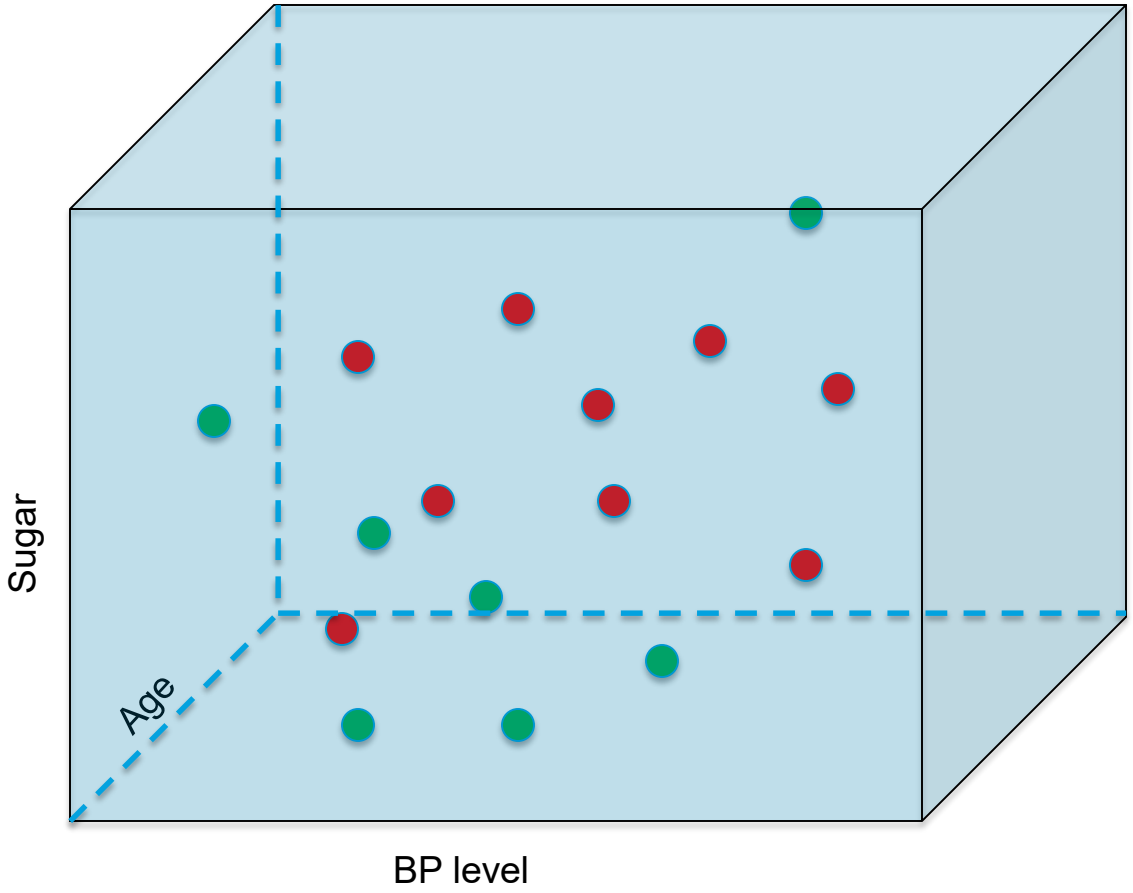
gender	age	diab	hypertension	stroke	chd	smoking	bmi
Female	80	0	0	0	1	never	25.19
Female	54	0	0	0	0		
Male	28	0	0	0	0	never	
Female	36	0	0	0	0	current	23.45
Male	76	0	1	0	1	current	20.14
Female	20	0	0	0	0	never	
Female	44	1	0	0	0	never	19.31
Female	79	0	0	0	0		23.86
Male	42	0	0	0	0	never	33.64
Female	32	0	0	0	0	never	
Female	53	0	0	0	0	never	
Female	54	0	0	0	0	former	54.7
Female	78	0	0	0	0	former	36.05
Female	67	0	0	0	0	never	25.6
Female	76	0	0	0	0		
Male	78	0	0	0	0		
Male	15	0	0	0	0	never	30.36
Female	42	0	0	0	0		

Introduction to machine learning

greatlearning
Learning for Life

Machine learning happens in mathematical space / feature space:

1. Position of a point in space is defined with respect to the origin
2. The position is decided by the values of the attributes for a point
3. We believe there is a process in nature that based on the given parameters makes some healthy and others not so



Introduction to machine learning

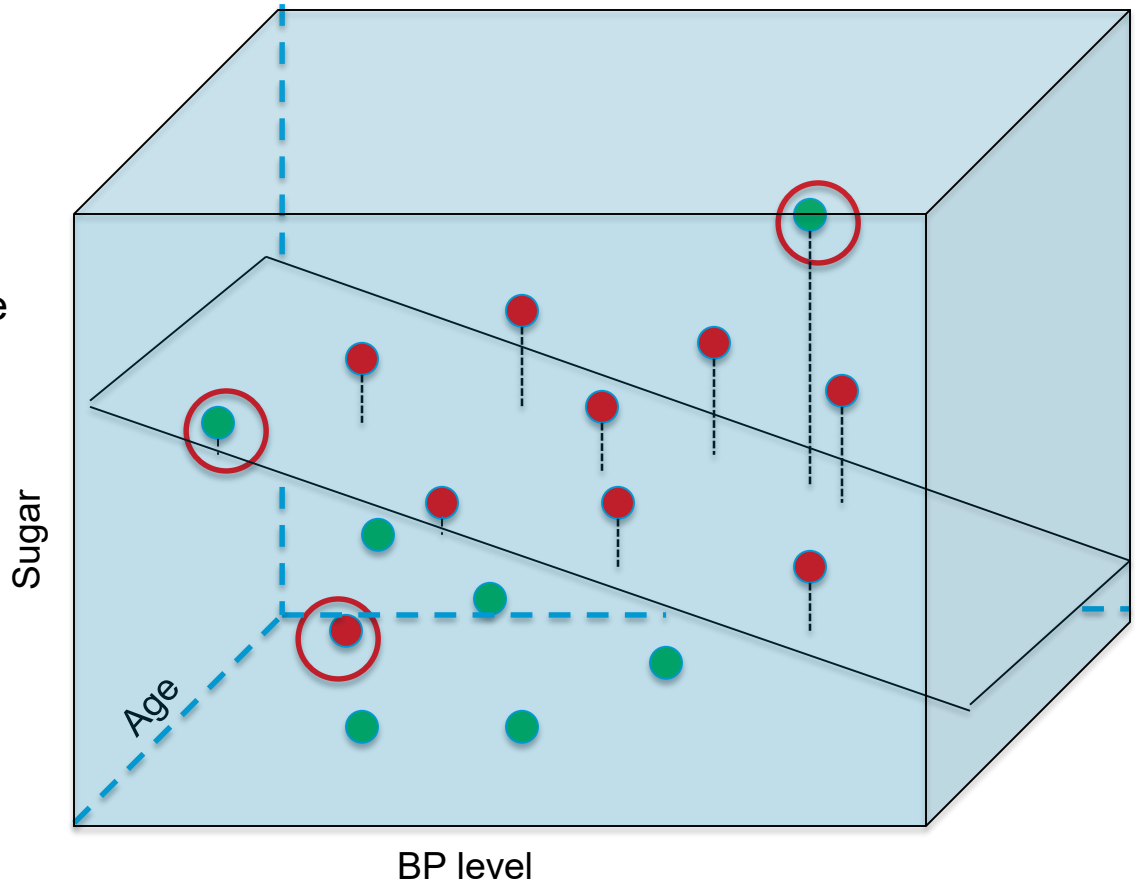
greatlearning
Learning for Life

Machine learning happens in mathematical space / feature space:

4. We wish to build a model that would represent the natural process

5. The model could be a simple plane, complex plane, hyper plane

6. The machine learning algorithm helps us find the model



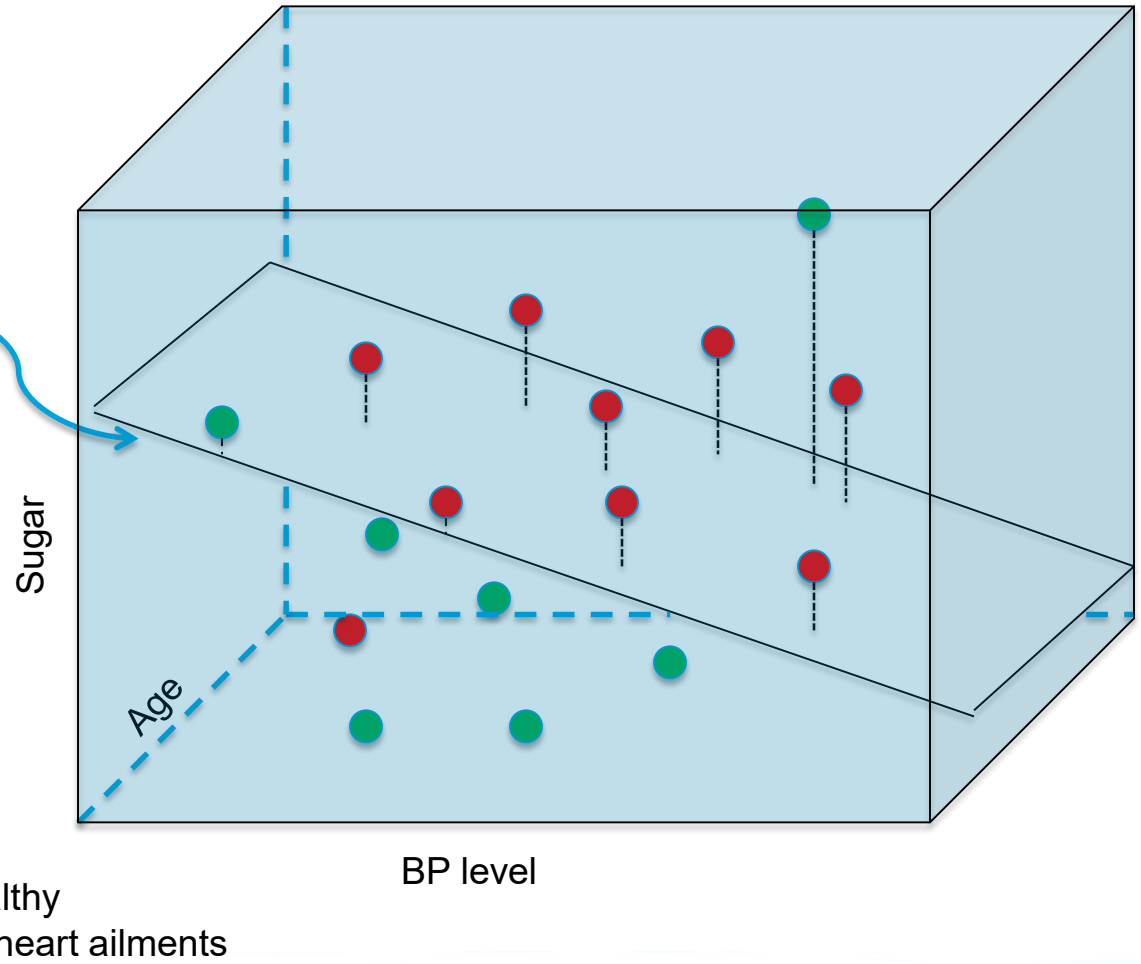
Introduction to machine learning

Machine learning happens in mathematical space / feature space:

7. In the figure, since the separator is a plane, the model will be the equation representing the plane

   ax + by + cz = d

8. x , y, z represent the three dimensions i.e. BP, Age, Sugar while d represents the color i.e. healthy or ailing heart



Introduction to machine learning

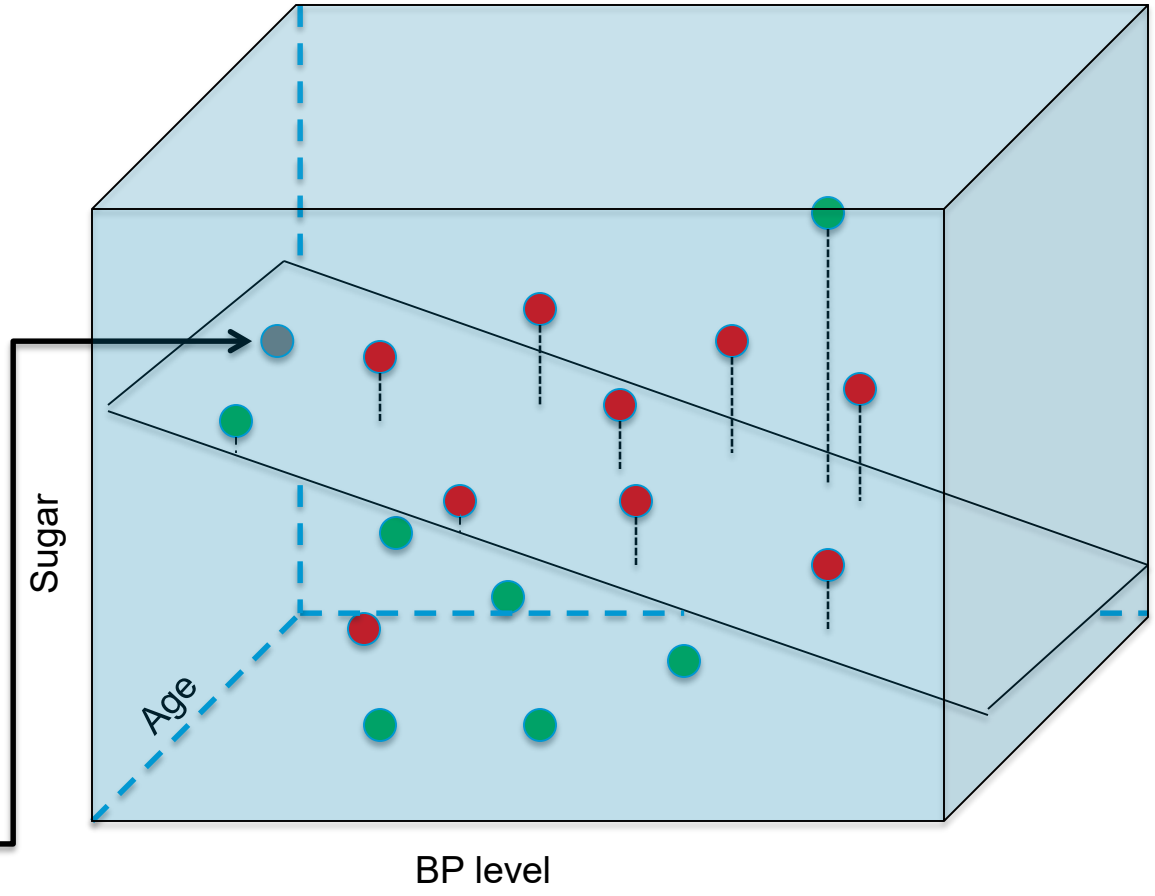
greatlearning
Learning for Life

Machine learning happens in mathematical space / feature space:

9. A new data point (a new customer) enters the system

10. It's x,y and z values will be fed into the model to get value of d (healthy or ailing)

11. The data point will be placed above or below the plane based on d



ax + by + cz = d,

Female	36	0	0	0	current	23.45

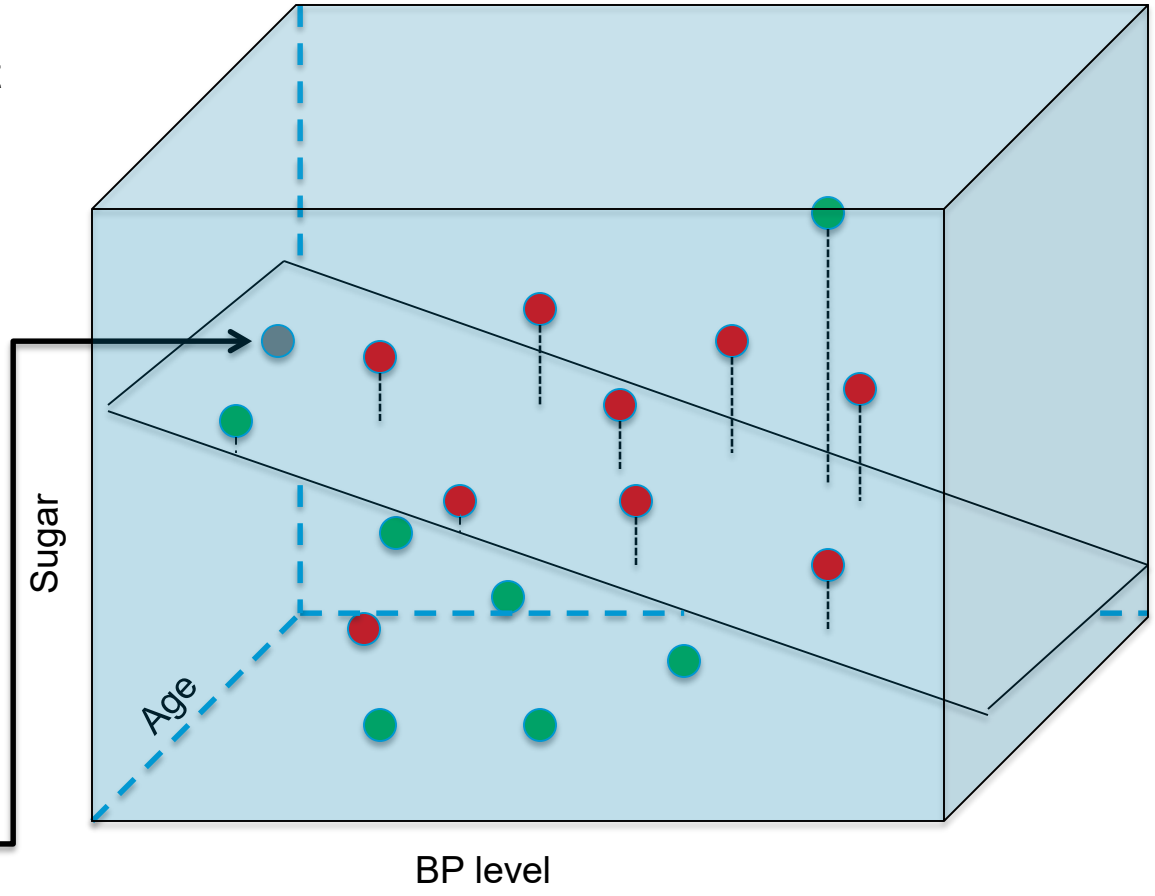
Heart healthy
Potential heart ailments

Introduction to machine learning

greatlearning
Learning for Life

Machine learning happens in mathematical space / feature space:

12. Whether the new data point is correctly placed (above or below the plane) i.e. correctly classified as ailing or healthy hear will be known only after direct observation



ax + by + cz = d,

Female	36	0	0	0	0	current	23.45

Heart healthy
Potential heart ailments

Introduction to machine learning

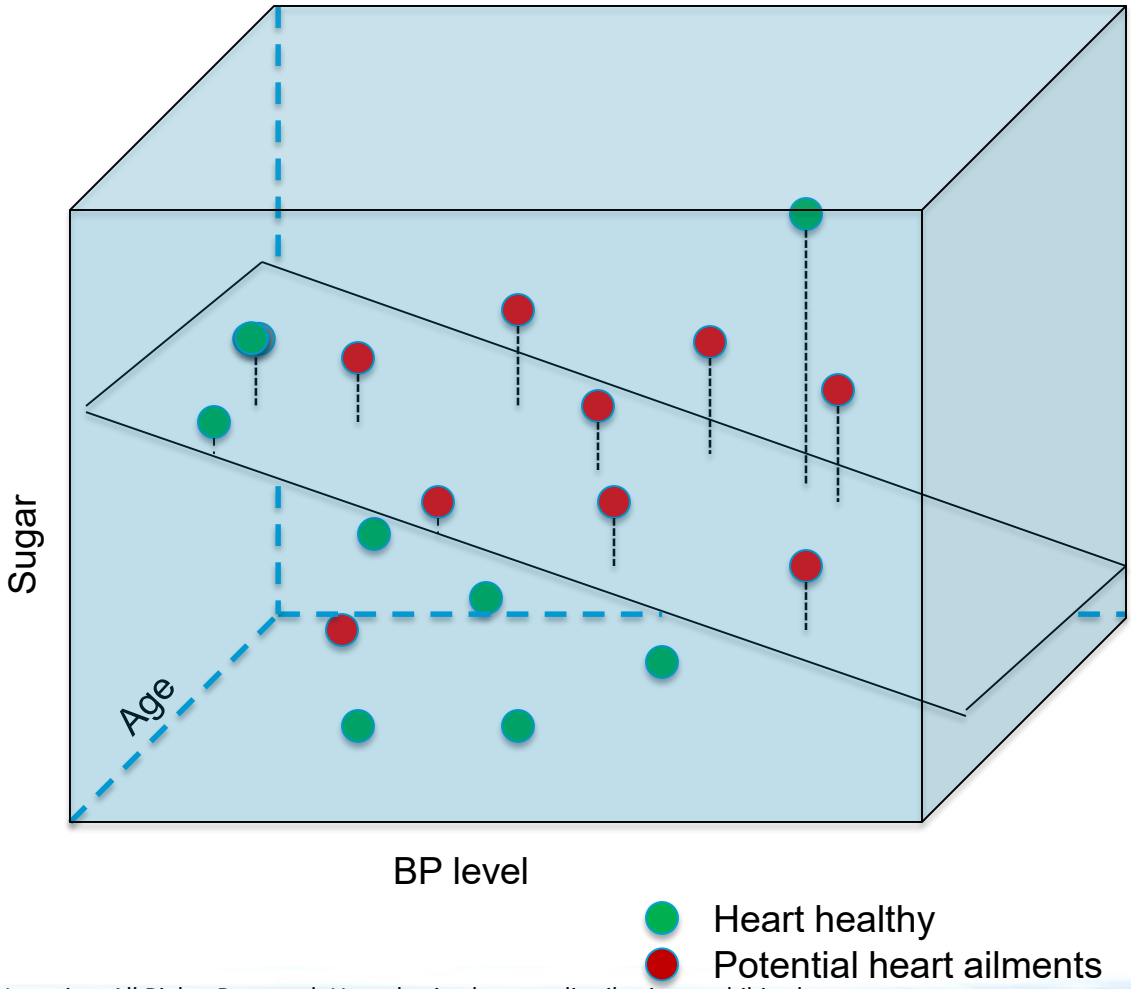
greatlearning
Learning for Life

Machine learning happens in mathematical space / feature space:

13. Only direct test on the object of interest will tell whether the classification is correct or not

ax + by + cz = d,

14. If majority of new data points are correctly classified, the model is good else not



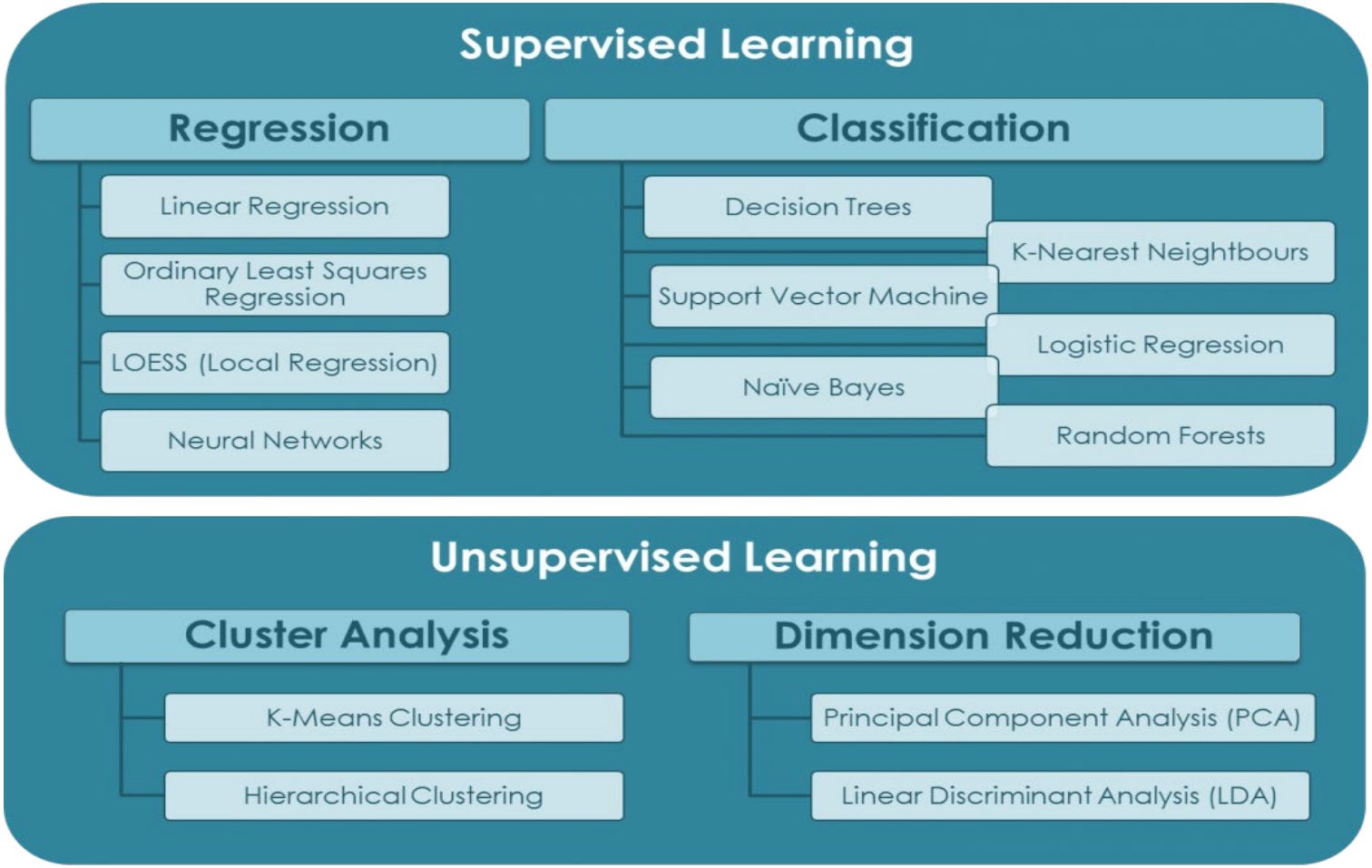
greatlearning
Learning for Life

Machine Learning Categories

Introduction to machine learning

greatlearning
Learning for Life

Machine learning categories:



Source: https://quantdare.com/machine-learning-a-brief-breakdown/

Introduction to machine learning

greatlearning
Learning for Life

Characteristics of Supervised Machine Learning -

a. Class of machine learning algorithms that work on externally supplied instances (data) in form of predictor attributes and associated target values

b. Process of model building involves training and testing stages

c. Training stage involves use of training data (a subset of the externally supplied data) supplied in form of independent and target values

d. They produce a model which is supposed to represent the real process that generated the data.

e. The model is tested for it's performance in test stage using test data. If satisfactory the model is implemented (productionized)

f. The model is used to predict target values for new data points

greatlearning
Learning for Life

Python for Machine Learning

Introduction to machine learning

greatlearning
Learning for Life

Python For Machine Learning

1. Python and R are suited for data science functions. Recently, Go has emerged as an up and coming alternative to the three major languages, but is not yet as well supported as Python.

2. In practice, data science teams use a combination of languages to play to the strengths of each one, with Python and R used in varying degrees

3. Python stands out as the language best suited for all areas of the data science and machine learning framework.

4. Refer : https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis#gs.qkenPdo

Introduction to machine learning

greatlearning
Learning for Life

Python packages for Data Science and Machine Learning

NumPy for scientific computing, many other libraries use NumPy arrays to operate efficiently. It also supports multidimensional arrays and matrices, as well as mathematical and statistical functions that need little code

SciPy builds on NumPy by adding a collection of algorithms and functions for computing integrals numerically, solving differential equations, optimization, and more.

Pandas adds data structures and tools that are designed for practical data analysis in finance, statistics, social sciences, and engineering. Pandas works well with incomplete, messy, and unlabeled data (i.e., the kind of data you're likely to encounter in the real world), and provides tools for shaping, merging, reshaping, and slicing datasets

matplotlib is the standard Python library for creating 2D plots and graphs.

Jupyter extends the functionality of Python's interactive interpreter with a souped-up interactive shell that adds introspection, rich media, shell syntax, tab completion, and command history retrieval.

Introduction to machine learning

greatlearning
Learning for Life

Python packages for Data Science and Machine Learning

scikit-learn builds on NumPy and SciPy by adding a set of algorithms for common machine learning and data mining tasks, including clustering, regression, and classification.

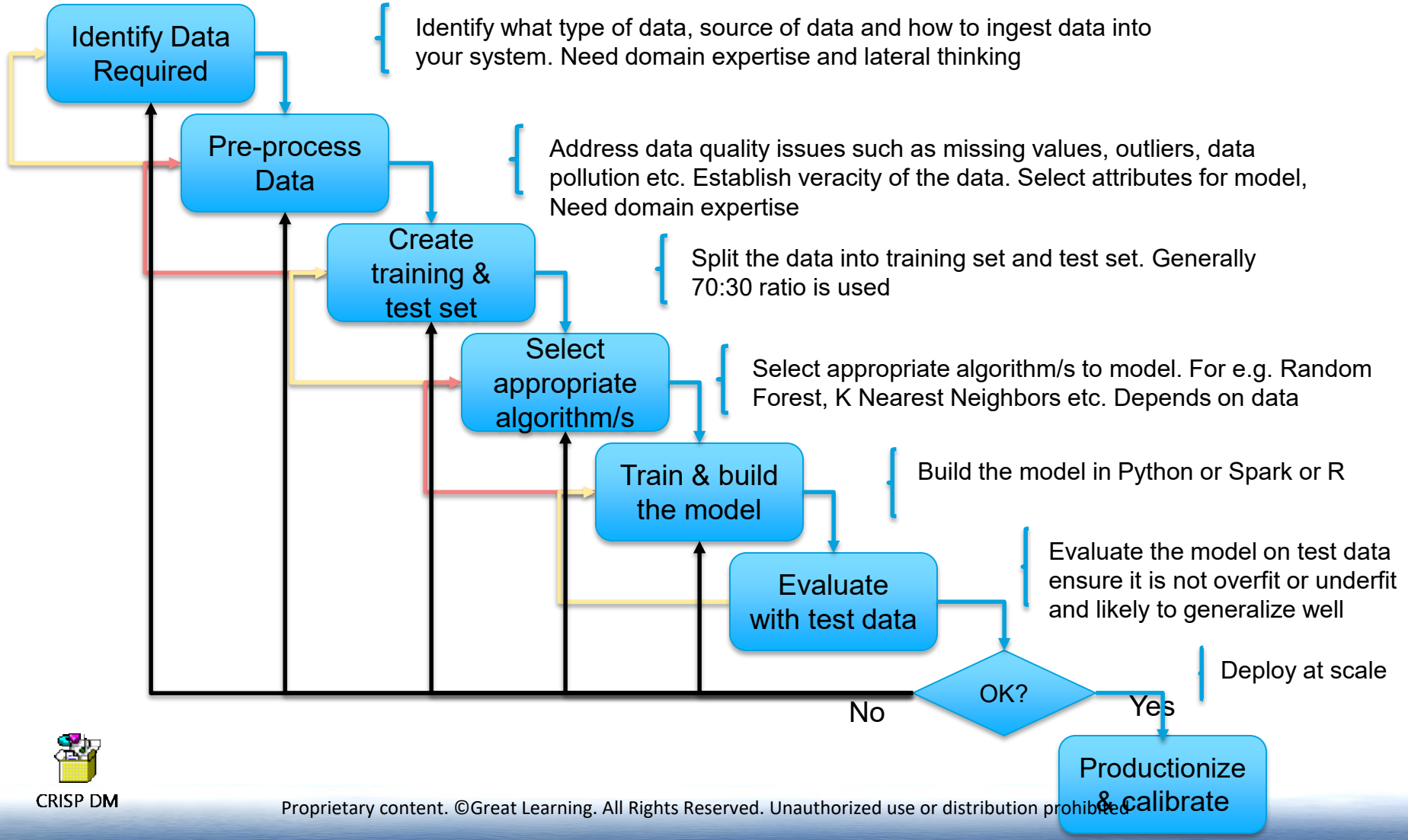
Theano uses NumPy-like syntax to optimize and evaluate mathematical expressions. What sets Theano apart is that it takes advantage of the computer's GPU in order to make data-intensive calculations up to 100x faster than the CPU alone. Theano's speed makes it especially valuable for deep learning and other computationally complex tasks.

TensorFlow Developed by Google as an open-source library for training neural networks.

Introduction to machine learning

Data Science Machine Learning Steps -

greatlearning
Learning for Life



CRISP DM