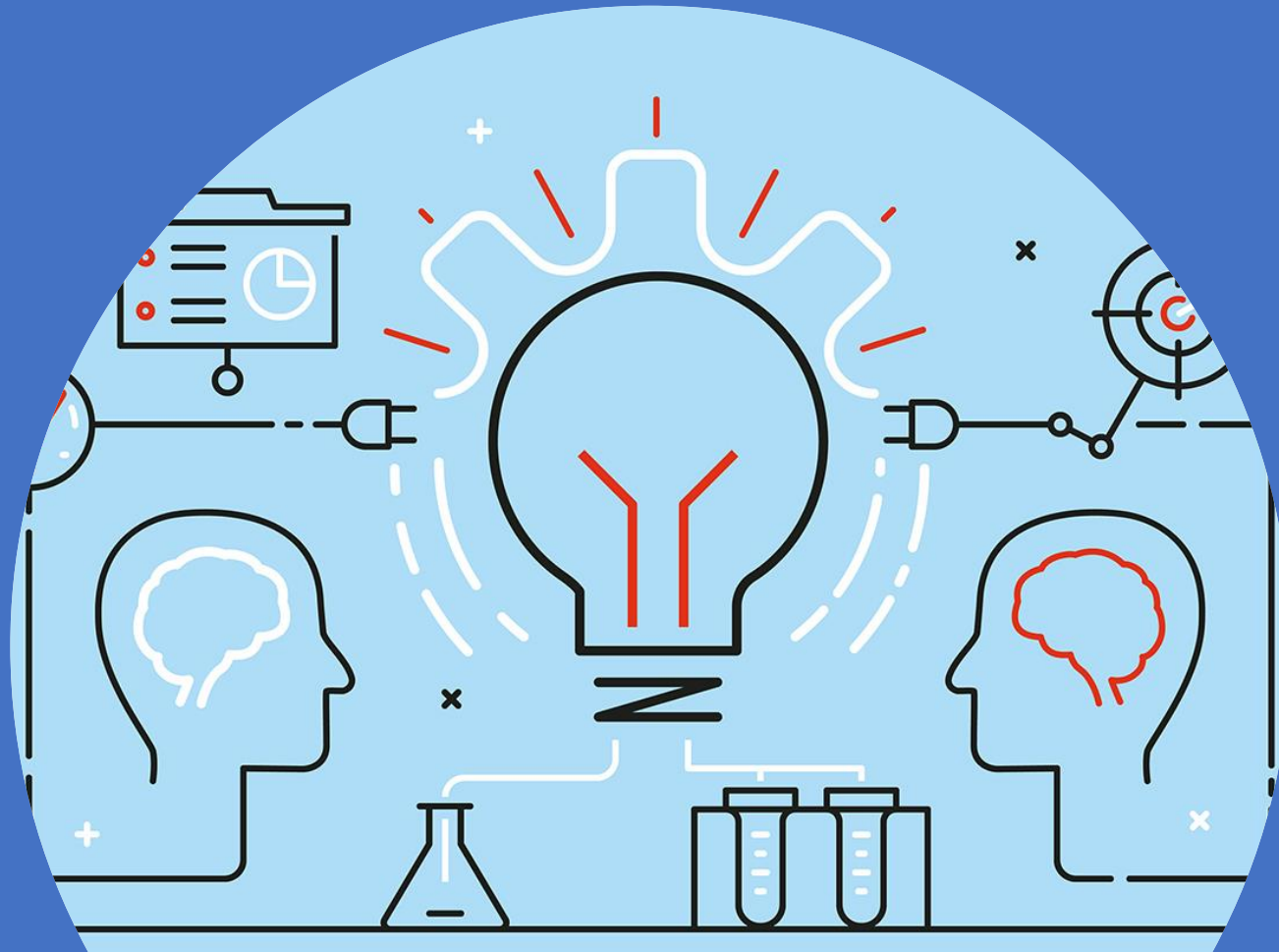statinfer

ML-MACHINE LEARNING
DL-DEEP LEARNING
AI – ARTIFICIAL INTELLIGENCE
DS – DATA SCIENCE

VENKATA REDDY KONASANI

PART-1
What is Machine Learning?

Activity

Close your eyes and think about two terms – Machine Learning and Artificial Intelligence.

- What is the first thing that comes to your mind when you hear these terms
  - Machine Learning
  - Artificial Intelligence

WHAT ARE YOU THINKING ABOUT – ROBOTS?

OR ANY AI BASED MOVIE?

HAVE YOU THOUGHT ABOUT

- Mathematics
- Statistics
- Datasets
- Data Analysis
- Optimization
- Algorithms
- Data Mining
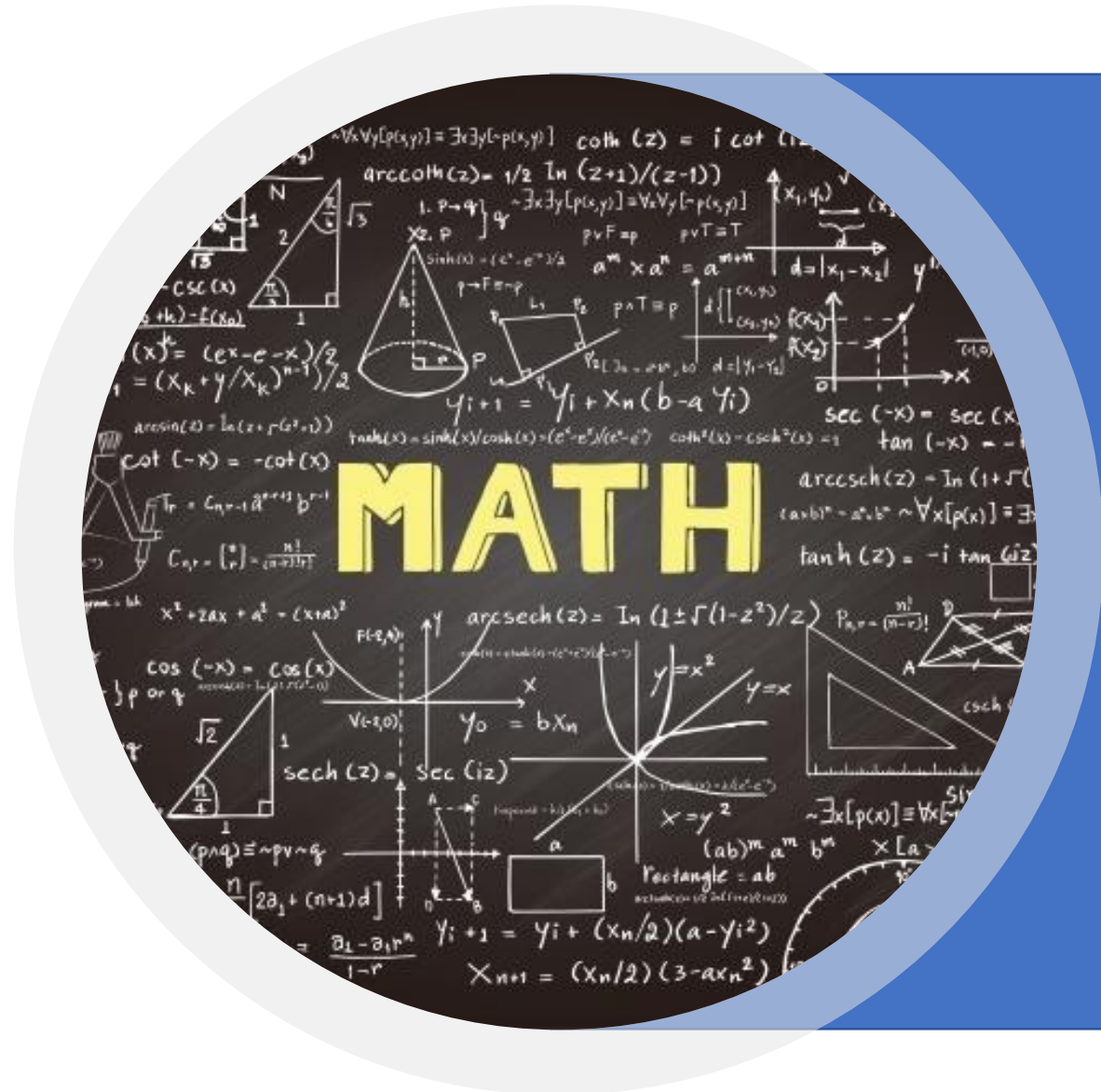
OK GOOGLE ...WHAT IS MACHINE LEARNING?

Machine learning (ML) is the scientific study of algorithms and **statistical models** that **computer systems** use to effectively **perform a specific task** without using explicit instructions, relying on models and inference instead.

statinfer

WHAT REALLY IS MACHINE LEARNING ?- WIKIPEDIA

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on models and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, kno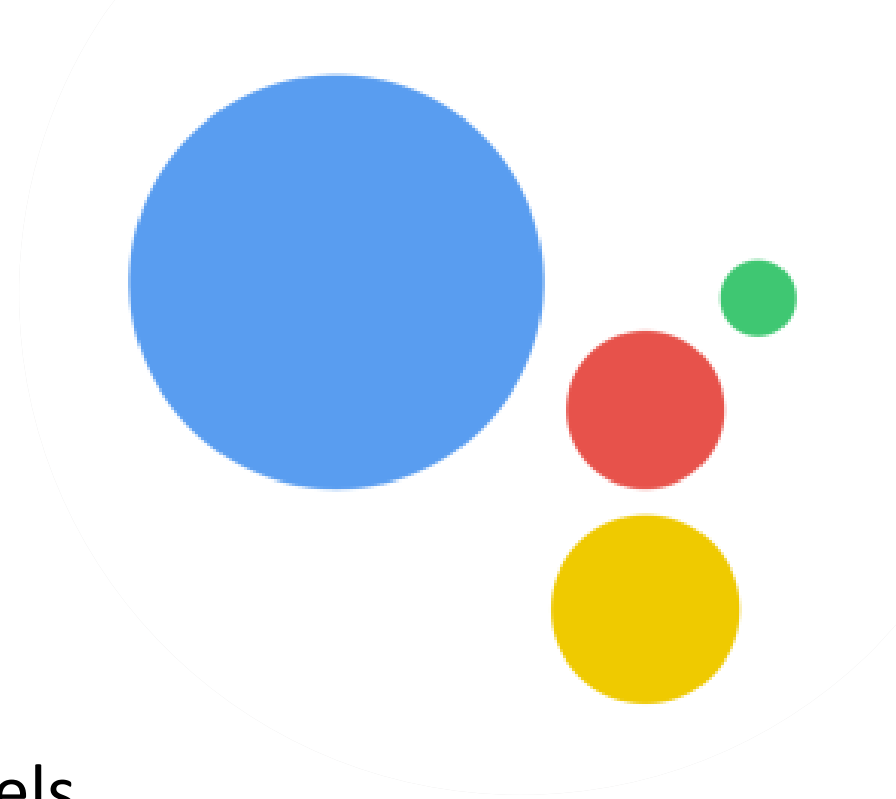wn as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.[1][2]:2 Machine learning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning.[3][4] In its application across business problems, machine learning is also referred to as predictive analytics.

IN SIMPLE TERMS ..

- Using historical data to make future predictions

- Building models on historical data to predictions

- Taking training data, building models on the training data using the models to make the future predictions

- Making the machine learn the patterns in the data

statinfer

DATA IS IN DIFFERENT FORMS

Numerical data

Image data (pixel intensities)

Video data (frames per second)

Sound data (waves)

Text data (tweets, comments, feedback)

APPLICATIONS OF NUMERICAL DATA

CREDIT RISK MODELS

Identifying risky customers before offering a loan

MARKETING ANALYTICS

Do you receive any marketing calls? Have you ever received any marketing call for Audi car?

RETAIL SALES ANALYTICS

Have you ever wondered, why only you are getting promotional offers on cloths and accessories where as I am getting offers on apartments?

FRAUD ANALYTICS

How does a bank decide the potential fraud transactions from millions of credit card swipes?

statinfer

11

APPLICATIONS OF MACHINE LEARNING – IMAGES AND VIDEO DATA

- Face recognition – Using image as input data

- Object recognition – Pixels is the input data

- Digit recognition – Using text as image

- Self Driving Cars – Using video data as input

IMAGE DATA IS ALSO NUMERICAL DATA

Image data
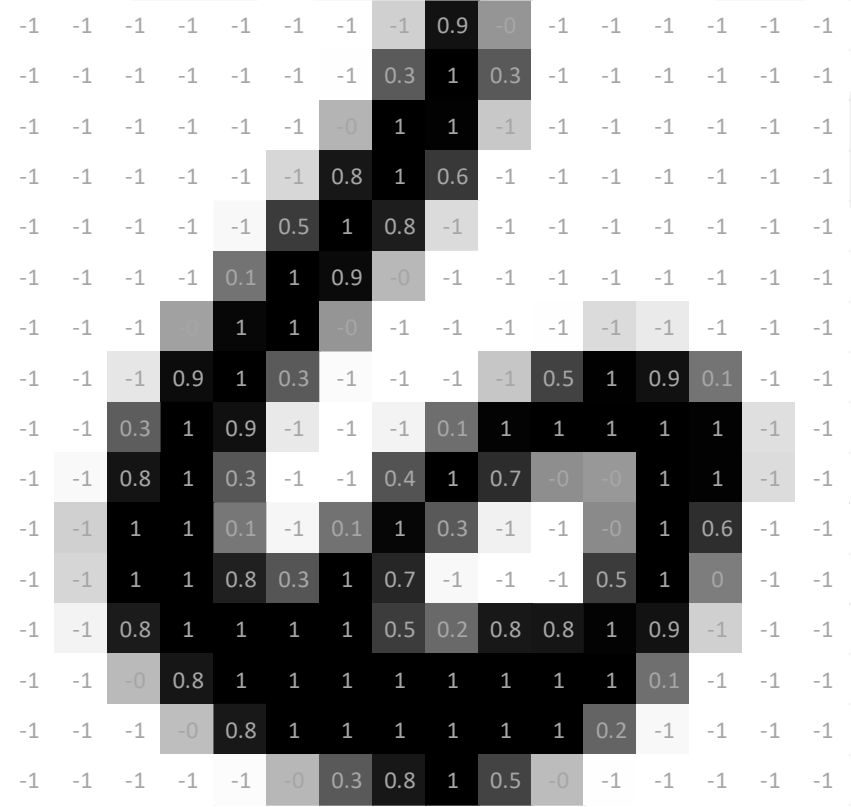
| Human Vision | Computer Vision |

APPLICATIONS ON TEXT DATA

- Sentiment Analysis
- Extraction of key topics in the data
- Document Classification

PART-2
What is Deep Learning?
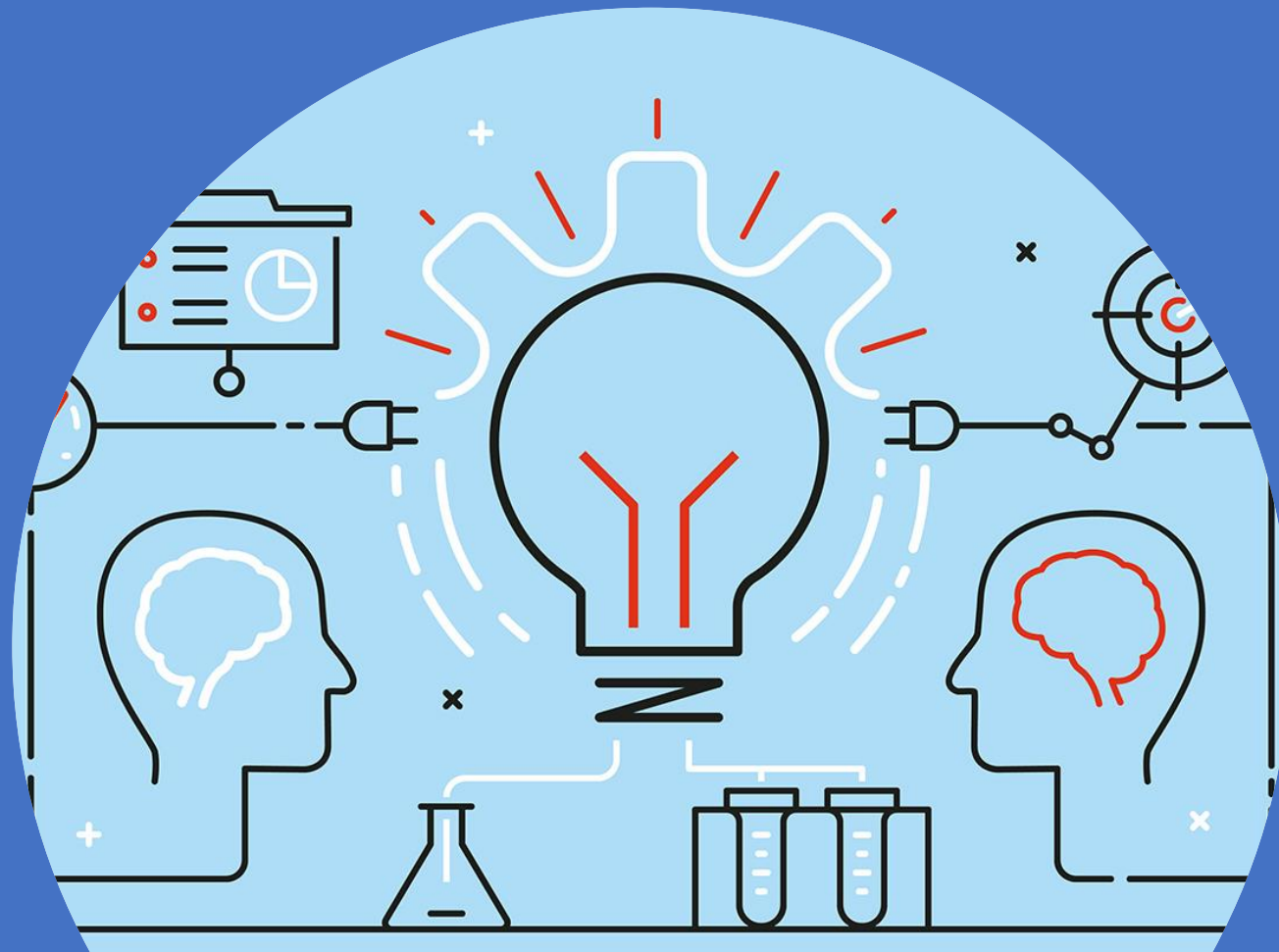
ANN

- ANN- Artificial Neural Network
- ANN is one of the technique in Machine Learning
- ANN has input layer , hidden layer and output layer
- For a really complex and non liner datasets we need several hidden layers
- ANN with multiple hidden layers is known as deep neural network

statinfer

DEEP LEARNING

- ANN with a single layer is known as shallow network

- ANN with multiple hidden layers is known as deep neural network

- Not just multiple hidden layers sometimes the type of hidden layer is also different.

- This concept of solving problems with multiple hidden layers is known as **deep learning**

statinfer

DEEP VS SHALLOW NETWORKS

A neural network with single hidden layer is called a shallow network
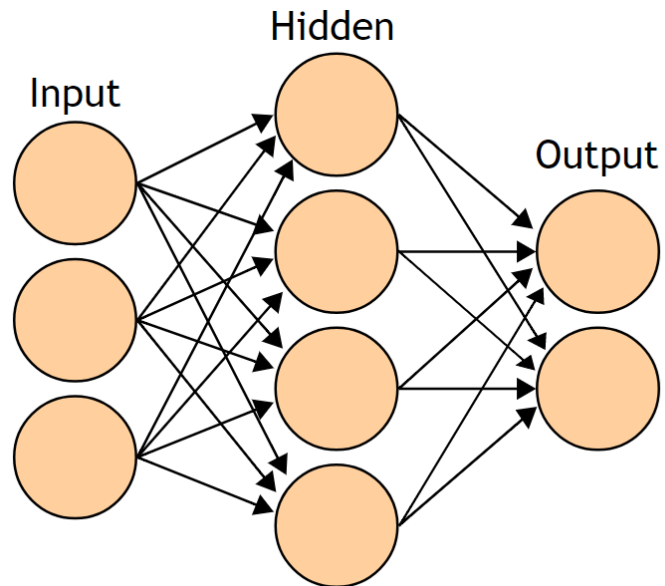
A neural network with more than one hidden layer is called deep neural network

shallow network
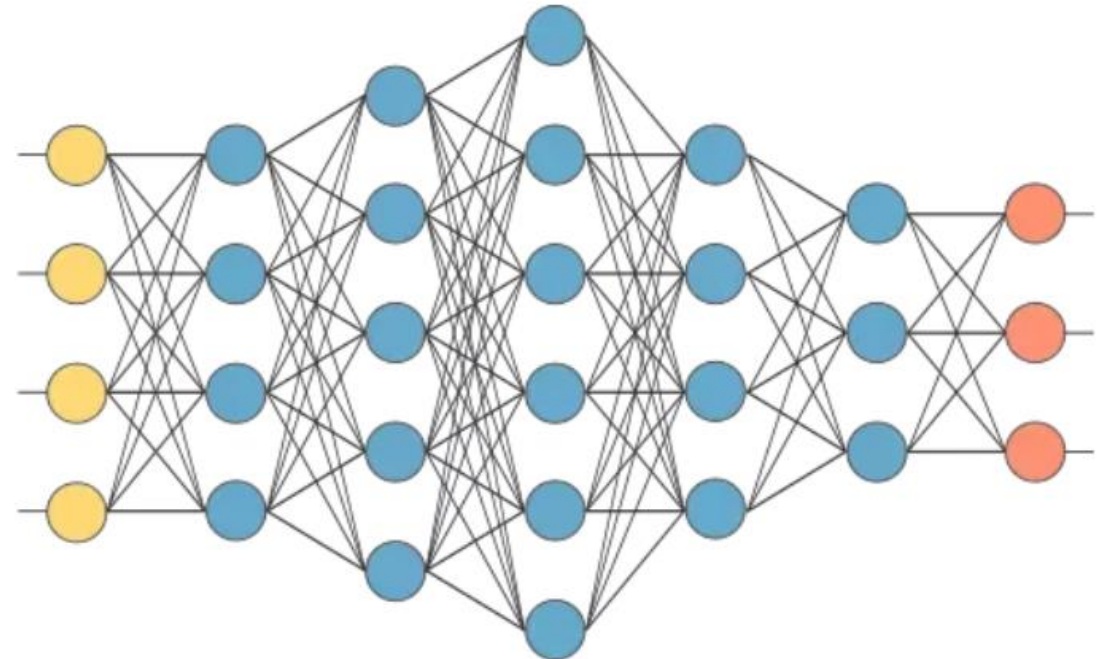
Deep network

DEEP VS SHALLOW NETWORKS

A single layer might not have the flexibility to capture all the non linear patterns in the data
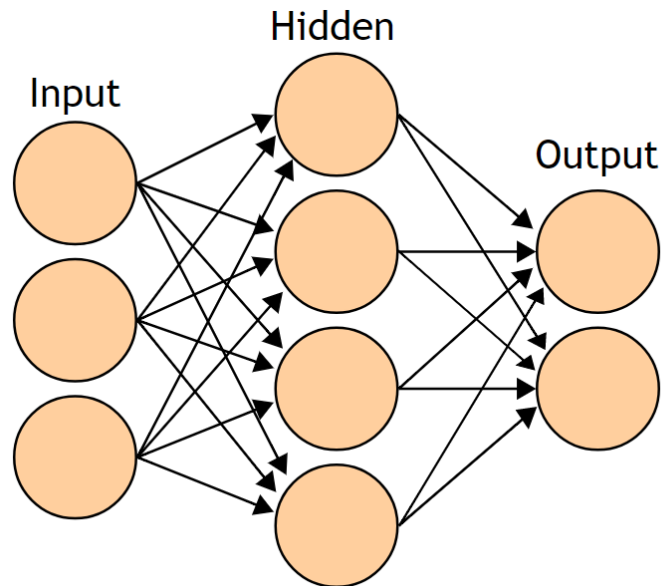
A deep network first learns the primitive features followed by high level features. This helps in building efficient models

Hidden

Input

Output

shallow network
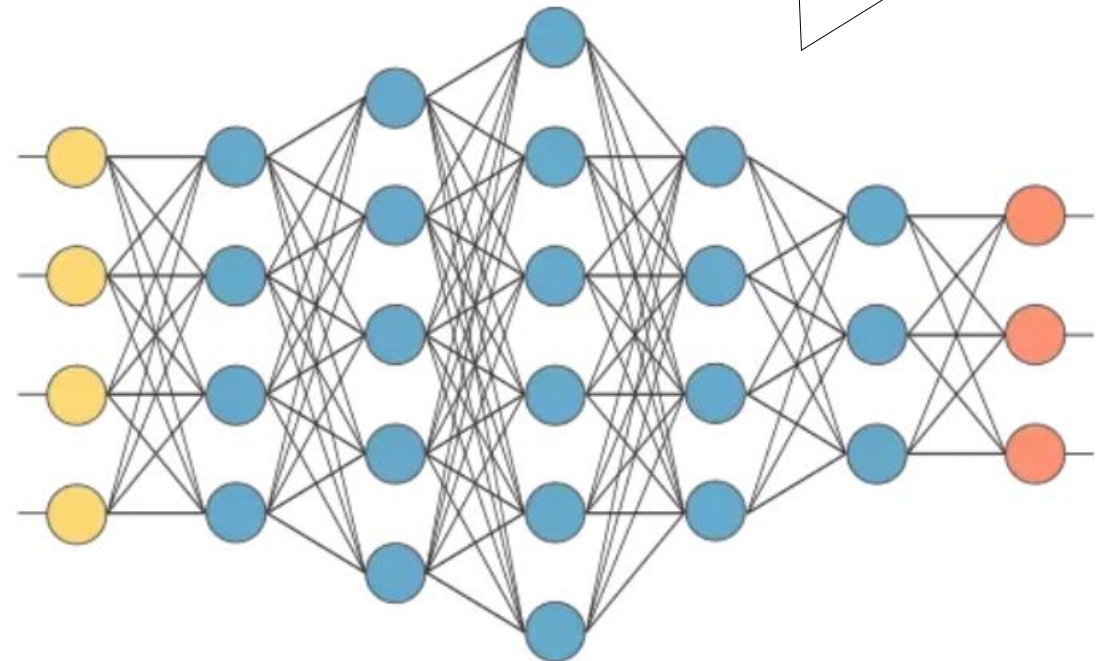
Deep network

DEEP NEURAL NETWORKS

- Lot of experiments have shown that a deep network with less parameters performs better than a shallow network
- For example deep network with hidden nodes `[10,10,10,10]` might perform better than shallow network with `[80]` hidden nodes
- Deep neural networks are amazingly powerful.
- With sufficient number of hidden layers and nodes, we can fit a model to any type of data
- They have the power to capture any amount of non linearity

statinfer

DEEP LEARNING IS A SUBSET OF MACHINE LEARNING

Machine Learning

Deep Learning
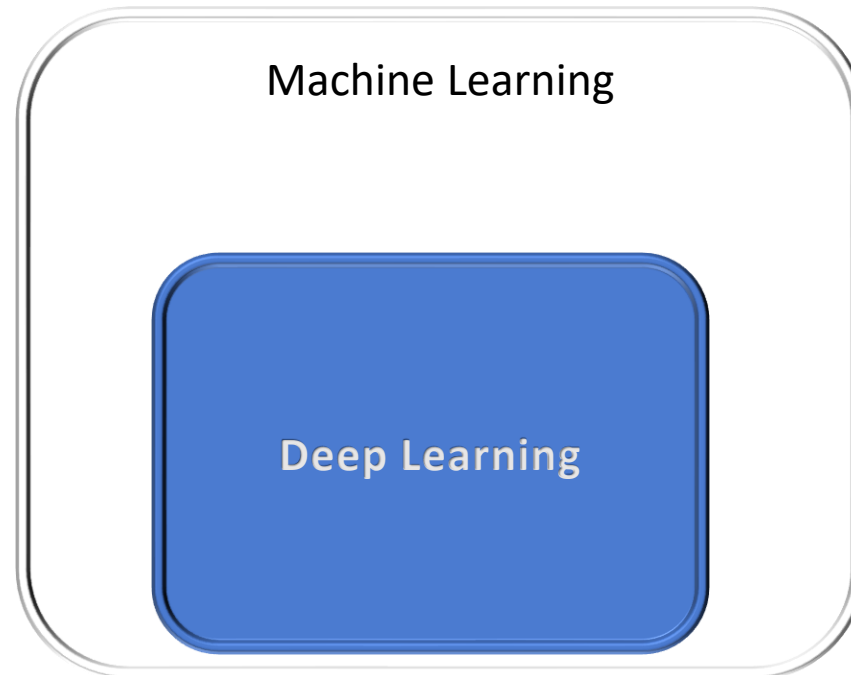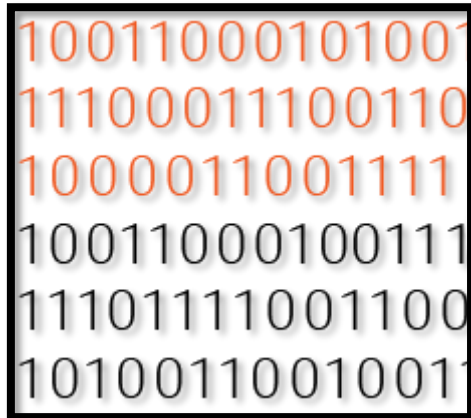
PART-3
What is Artificial Intelligence?

MACHINE LEARNING MODELS

Training data

Build Model

MACHINE LEARNING MODELS

New data

Apply Model

Get Prediction

11100011100110
1000011001111

$$h_w(x) = a_{4,1} \qquad a_{1,1} = x_1$$
$$a_{4,1} = g\left(\sum_i w_{3,i,1}\, a_{3,i}\right)$$
$$a_{3,1} = g\left(\sum_i w_{2,i,1}\, a_{2,i}\right) \qquad h_w(x) = g\left(\sum_k w_{3,k,1}\; g\left(\sum_j w_{2,j,k}\; g\left(\sum_i w_{1,i,j}\; x_i\right)\right)\right)$$
$$a_{\ell,1} = g\left(\sum_i w_{\ell-1,i,1}\, a_{\ell-1,i}\right)$$
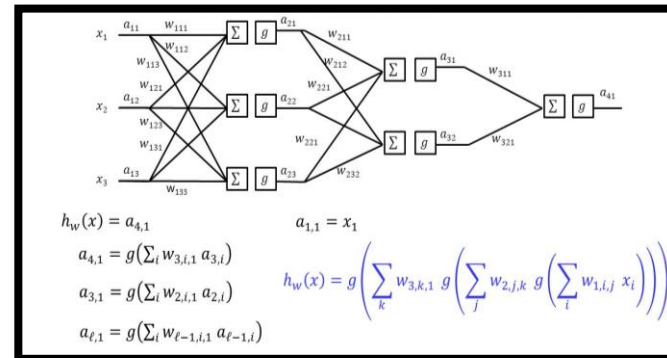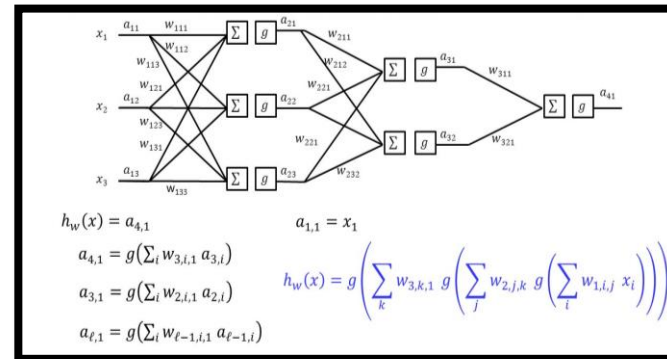
Class1

This prediction can be right or wrong

MACHINE LEARNING MODELS

New data

Apply Model

Get Prediction
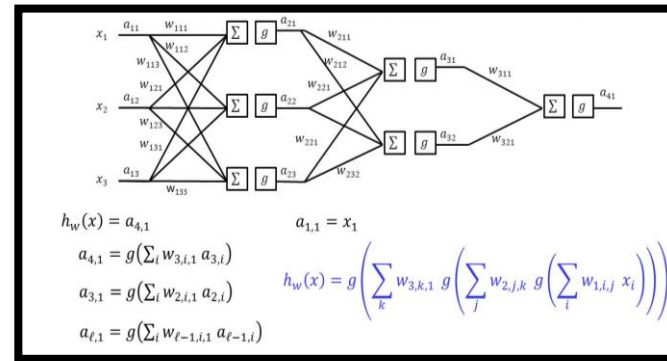
11100011100110
1000011001111
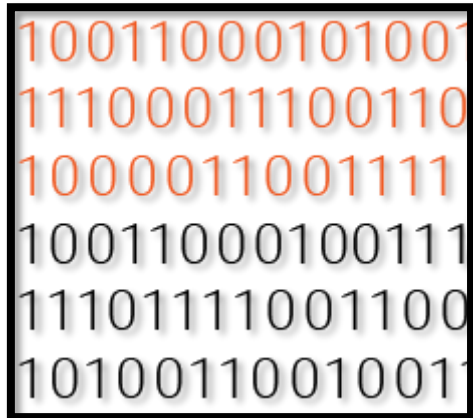
Class2

This prediction can be right or wrong

One way models

AI = MACHINE LEARNING MODELS + FEEDBACK LOOP

Training data

Model

AI = MACHINE LEARNING MODELS + FEEDBACK LOOP

New data

Apply Model

Get Prediction

111100011100110
1000011001111
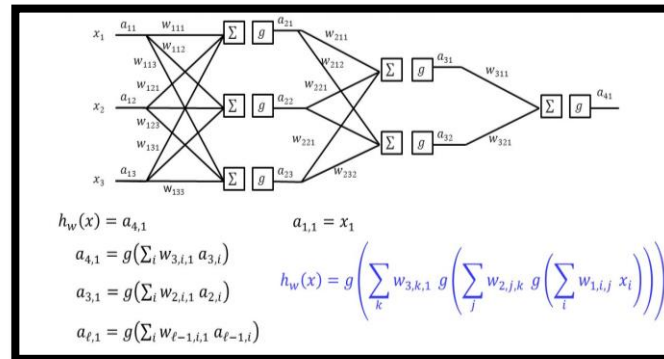
Class2
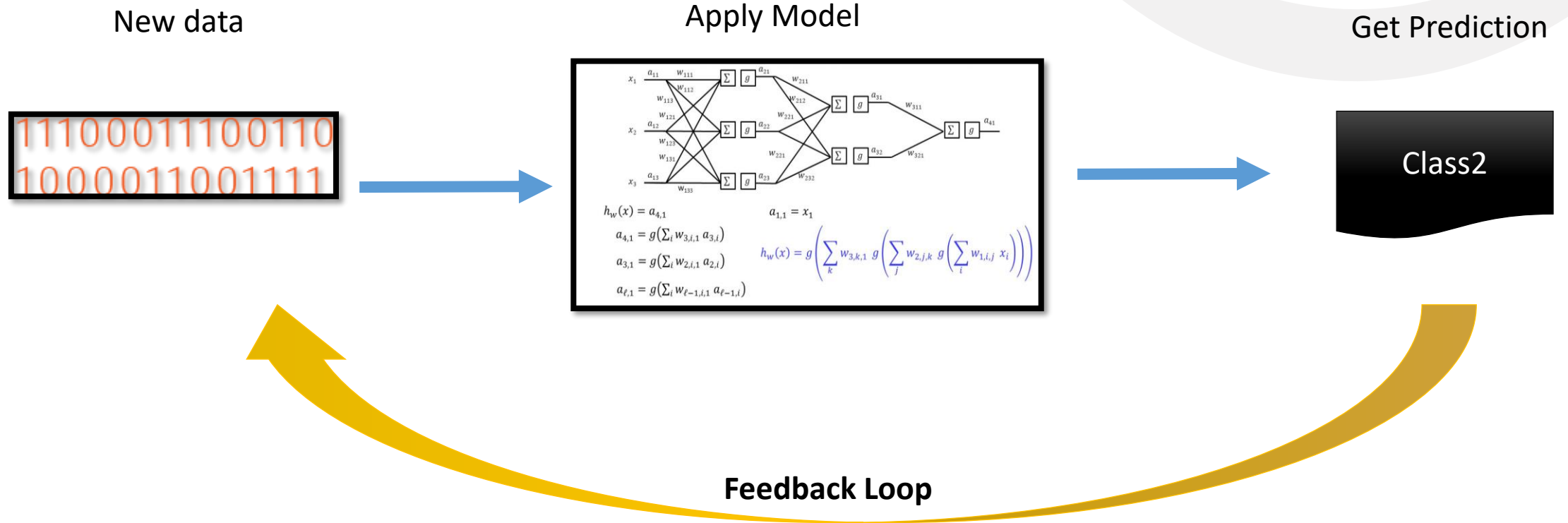
Feedback Loop

AI = MACHINE LEARNING MODELS + FEEDBACK LOOP

Update Training data based on feedback

Update the Model based on data
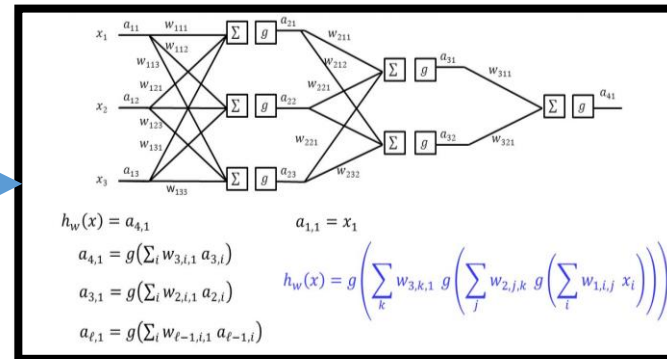
Prediction

Class2

Feedback Loop

HOW IS FEEDBACK COLLECTED

- Manual entry after going through test cases – Google maps

- Indirect feedback collection based on user actions for - User click vs not click on your YouTube ad

- Indirect feedback collection based on actions – In case of self driving car, hitting a wall is an action.

statinfer

APPLICATIONS OF AI

- Self driving cars
- SIRI / Ok-google
- Alexa /Google home
- Recommendation systems
- Image recognition
- Speech recognition
- Spam filtering

MACHINE LEARNING IS A SUBSET OF ARTIFICIAL INTELLIGENCE

Artificial Intelligence

Machine Learning

Deep Learning

statinfer
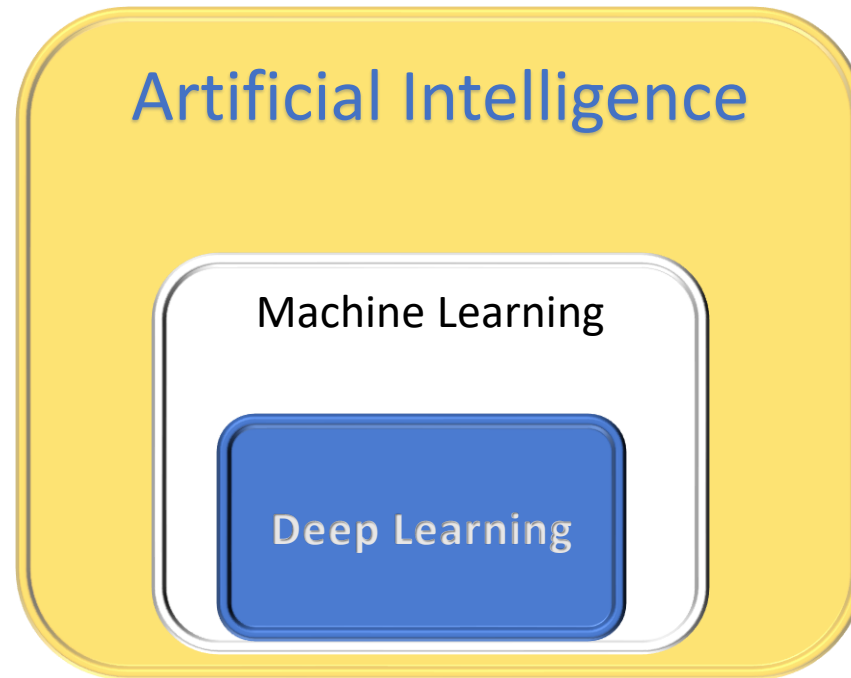
PART-4
What is Data Science?

WHAT IS DATA SCIENCE?

- Data Driven Decision making

- Making sense out of data

- Finding hidden patterns in the data

- Analysis using not just machine learning models but also using data visualizations, intelligent reports

- Most of the techniques and tools seen in data analysis in early days are now falling under data science

statinfer

DATA SCIENCE IS A FUSION OF MANY FIELDS

- Mathematics

- Statistics

- Coding

- Database management

- Data Analytics

- Predictive modelling

- Machine Learning

- Deep Learning

statinfer

DATA SCIENCE– FOUR MAJOR TYPE OF SKILLS

Database

Analytics & ML

Bigdata

Presentation
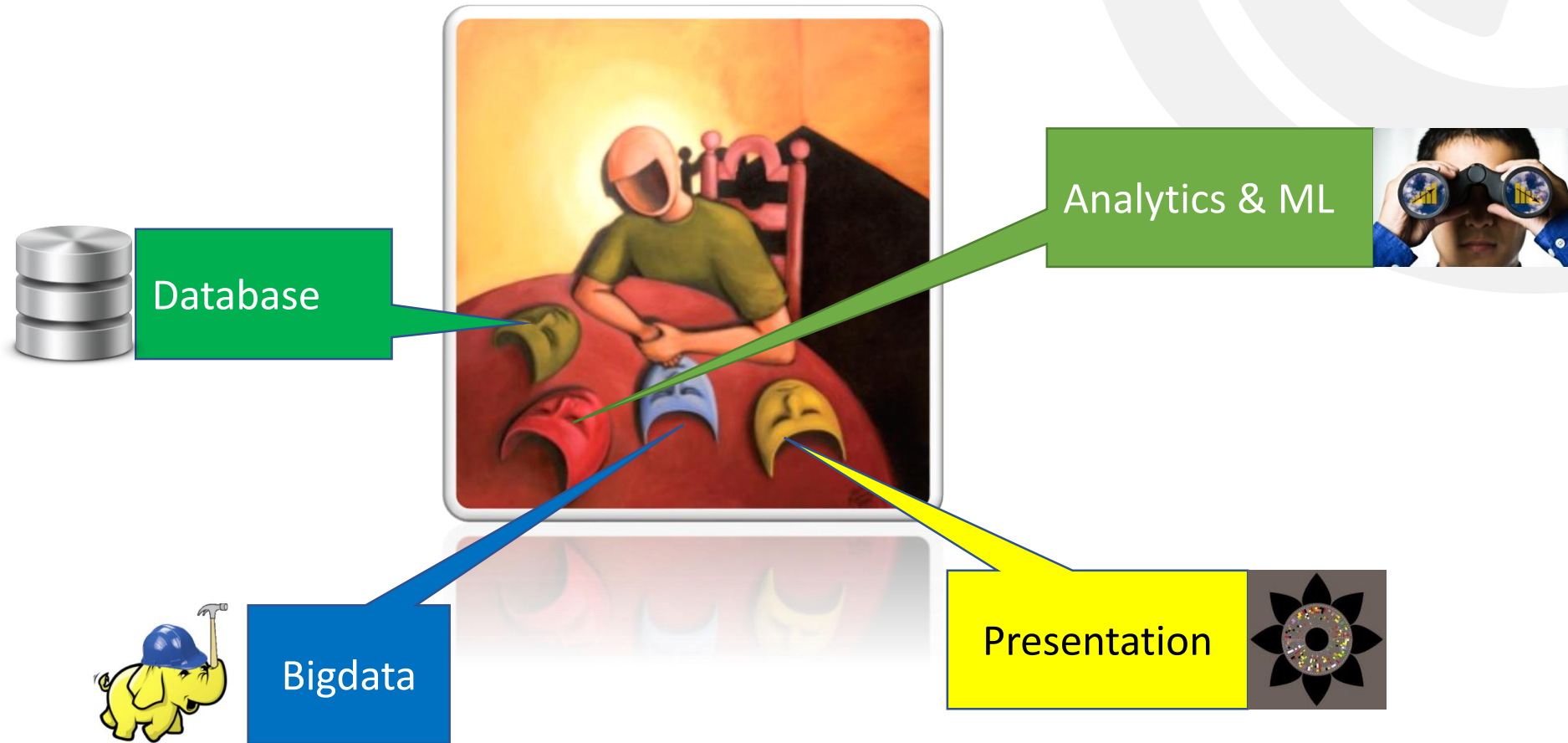
statinfer

THE TECHNIQUES YOU NEED TO KNOW

Database Knowledge

- Data base Management
- Data blending
- Querying
- Data manipulations
- ETL

Predictive Analytics & ML

- Basic descriptive statistics
- Advanced analytics
- Predictive modeling
- Machine Learning

Big Data knowledge

- Distributed Computing
- Big Data analytics
- Unstructured data analysis

Presentation Skill

- Data visualizations
- Report design
- Insights presentation

statinfer

36

MACHINE LEARNING TOOLS AND SOFTWARE'S

Database tools

SQL/MySql

OLAP cubes

Teradata

DB2/Sql Server/ Oracle/ Informix/Exadata

Analytical tools

SAS/R/SPSS/Python

Weka/MATLAB/ TensorFlow/OCR

Big Data Tools

Hadoop, Hive, Pig, Mahout, Spark, Java

Presentation Tools

Excel

Tableau, Qlikview

statinfer

DATA SCIENCE -DESIGNATIONS

Database Developer
ETL Developer
MIS & DB Developer
Data Architect
Data Engineer

Data Analyst
Statisticians
Business Analyst
Data Scientist

Bigdata Developer
Hadoop Developer
Software Engineer

MIS Analyst
Reporting Analyst
Business Analyst

statinfer

38
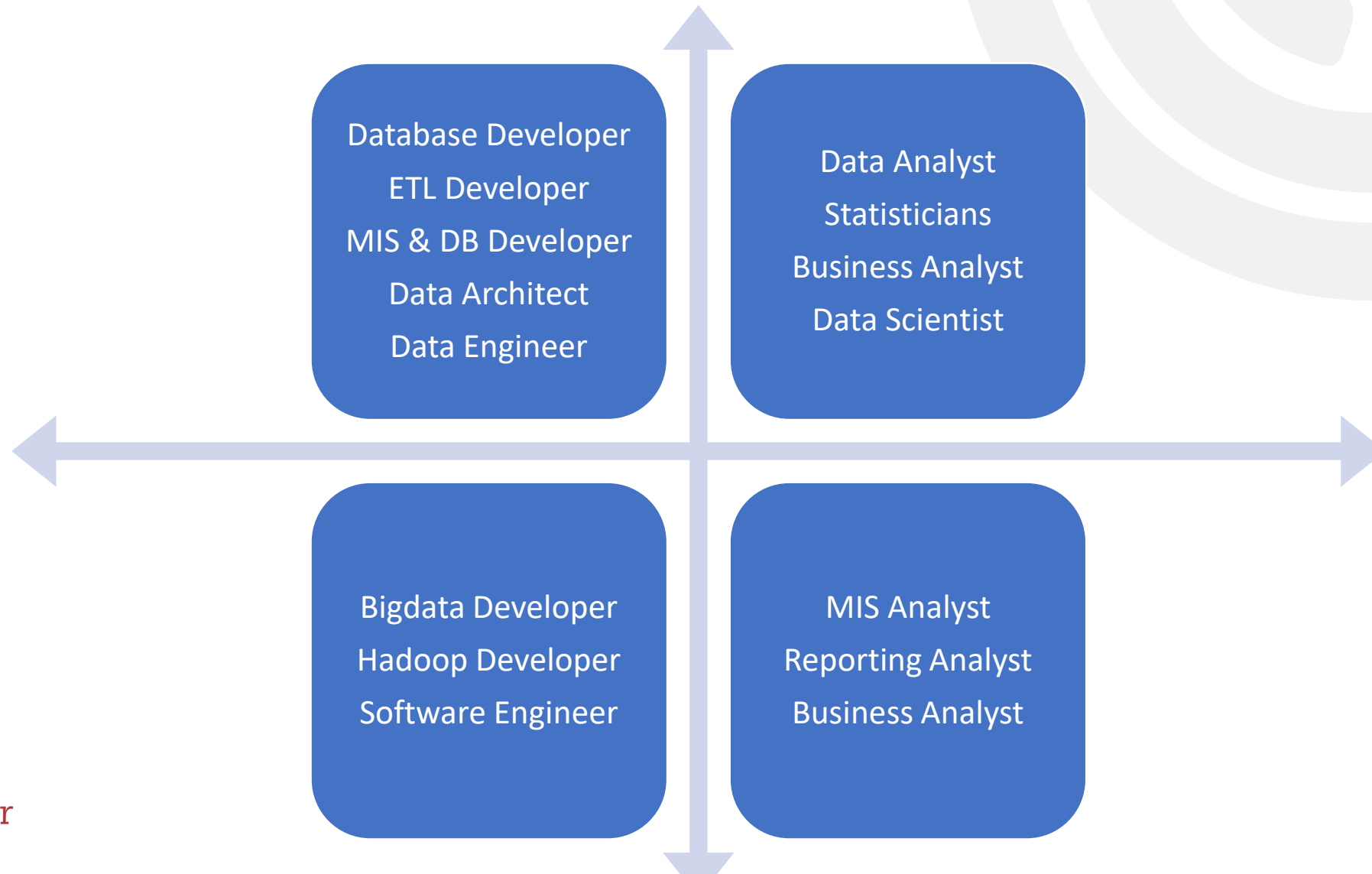
MACHINE LEARNING IS A PART OF DATA SCIENCE

Artificial Intelligence

Machine Learning

Deep Learning
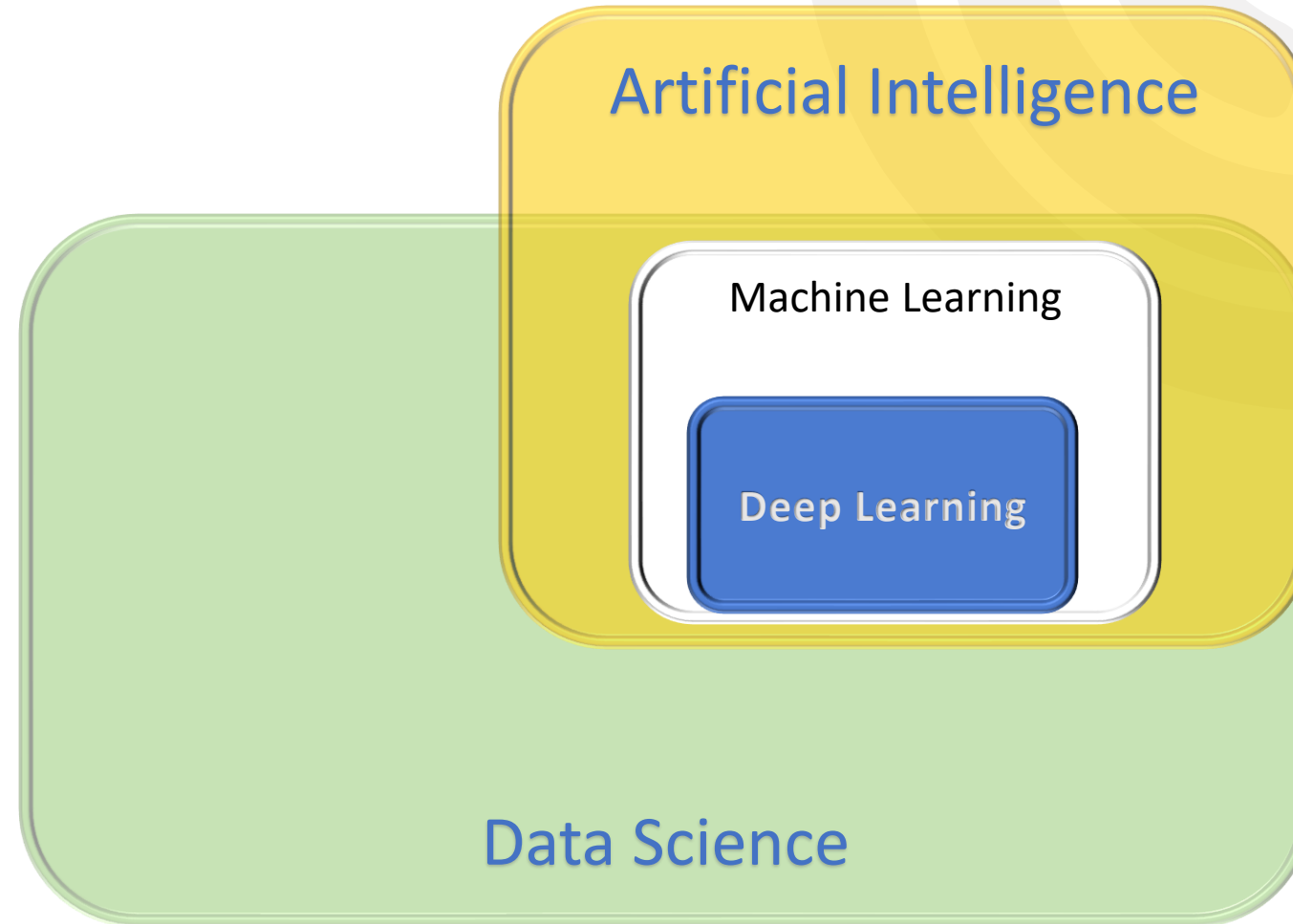
Data Science

* These are individual interpretations

statinfer

Statinfer.com

39

PART-5
The Learning Path

FAQ BY DATA SCIENCE ASPIRANTS

- I want to be data scientist what training should I take?

- I already have knowledge on few tools, what are my next steps?

- What skill should I add to my profile to make it to next level?

- I am new to data science, where can I start ?

statinfer

CATEGORIES OF PROFILES

You need training based on your skill level. Based on skill set we can divide the whole data science aspirants into four categories

1. Beginner - Completely new to Data Science and ML

2. Intermediate - MIS and Reporting Analyst

3. Advanced – Data Analyst and Predictive Modeler

4. Complete Data Scientist – ML, Hadoop, R, Python, DL, AI

statinfer

THE LEARNING PATH

Tools & Coding
R/SAS/Python/
Hadoop/Weka

Basic Statistics and Mathematics

Basic Algorithms - Regression, Classification and Segmentation

Advanced ML Algorithms -Neural Networks, SVMs, Random Forest and Boosting

Deep Learning Models

CNN, RNN and LSTM

AI Models

Deep Q Learning

Reinforced Learning

Markov Decision process
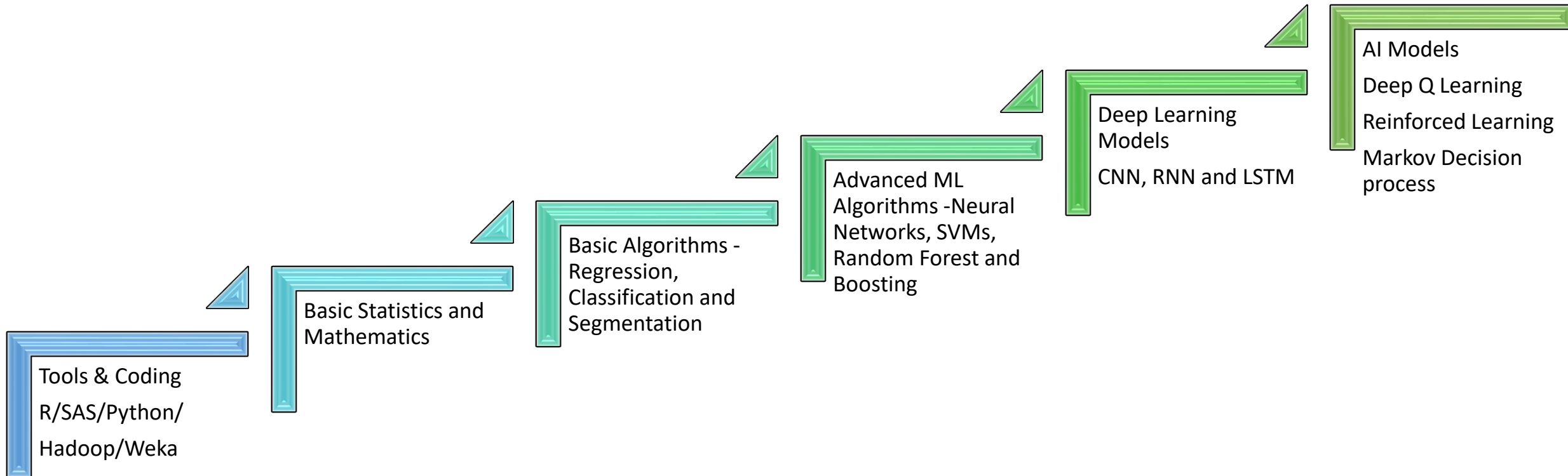
THE LEARNING PATH – OUR SUGGESTIONS

1. Do not try to learn all the steps in one sitting.
2. You need to learn, absorb and then practise before you reach the next step

Tools & Coding
R/SAS/Python/
Hadoop/Weka

Basic Statistics and Mathematics

Basic Algorithms - Regression, Classification and Segmentation

Advanced ML Algorithms -Neural Networks, SVMs, Random Forest and Boosting

Deep Learning Models

CNN, RNN and LSTM

AI Models

Deep Q Learning

Reinforced Learning

Markov Decision process

Stage-1

Stage-2

Stage-3
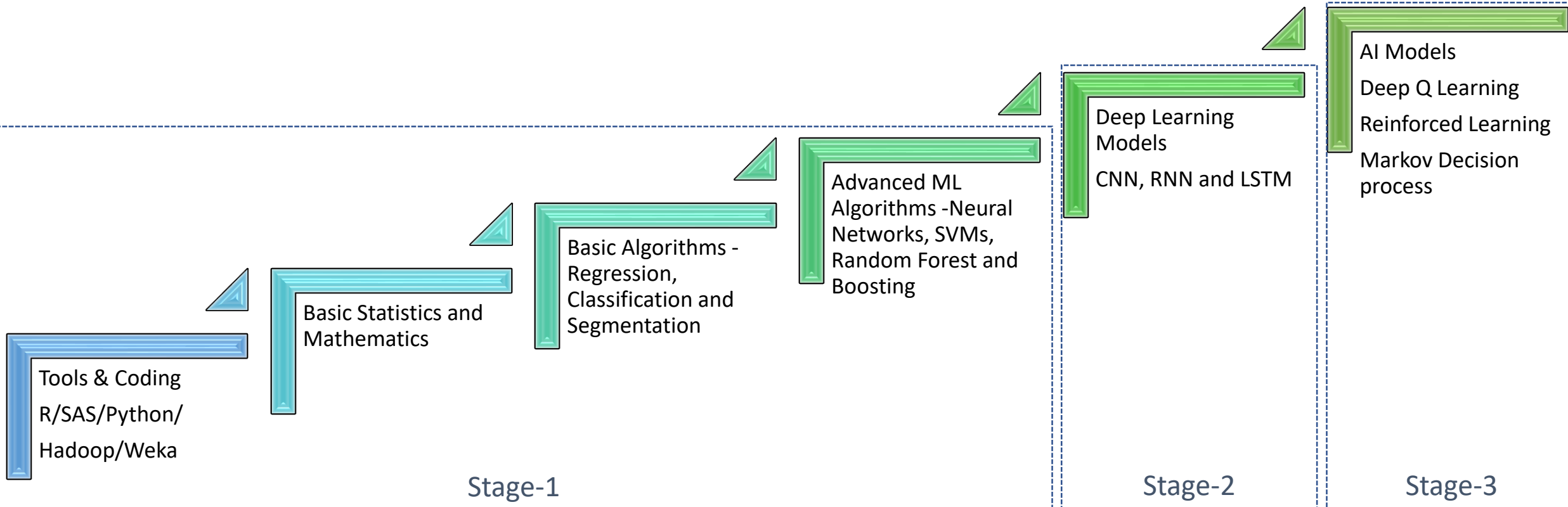
THE LEARNING PATH – OUR SUGGESTIONS

1. R or Python. Both are really good. Pick any one of them
2. It also depends on your business problem
3. If you are planning to learn deep learning then go for python

AI Models

Deep Q Learning

Reinforced Learning

Markov Decision process

Deep Learning Models

CNN, RNN and LSTM

Advanced ML Algorithms -Neural Networks, SVMs, Random Forest and Boosting

Basic Algorithms - Regression, Classification and Segmentation

Basic Statistics and Mathematics

Tools & Coding

R/SAS/Python/

Hadoop/Weka

Stage-1
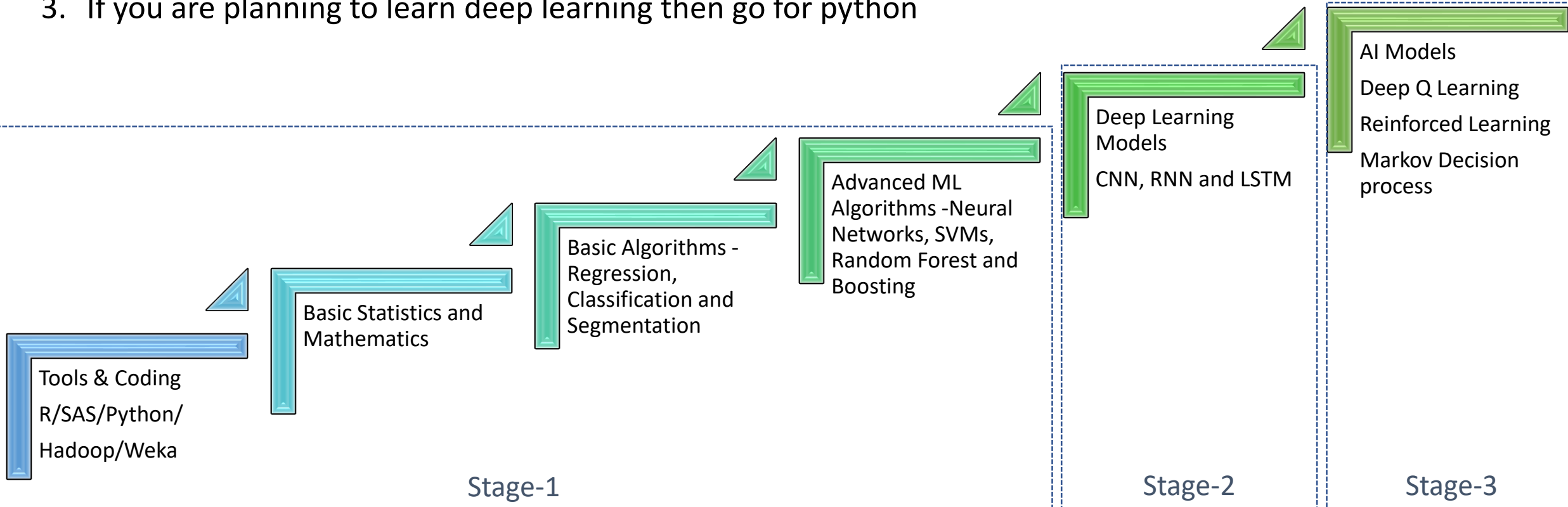
Stage-2

Stage-3

THE LEARNING PATH – OUR SUGGESTIONS

1. Do not start with stage-2 or stage-3 directly.
2. Strong fundamentals will make the learning easy in later stages.

AI Models

Deep Q Learning

Reinforced Learning

Markov Decision process

Deep Learning Models

CNN, RNN and LSTM

Advanced ML Algorithms -Neural Networks, SVMs, Random Forest and Boosting

Basic Algorithms - Regression, Classification and Segmentation

Basic Statistics and Mathematics

Tools & Coding

R/SAS/Python/

Hadoop/Weka

Stage-1   Stage-2   Stage-3
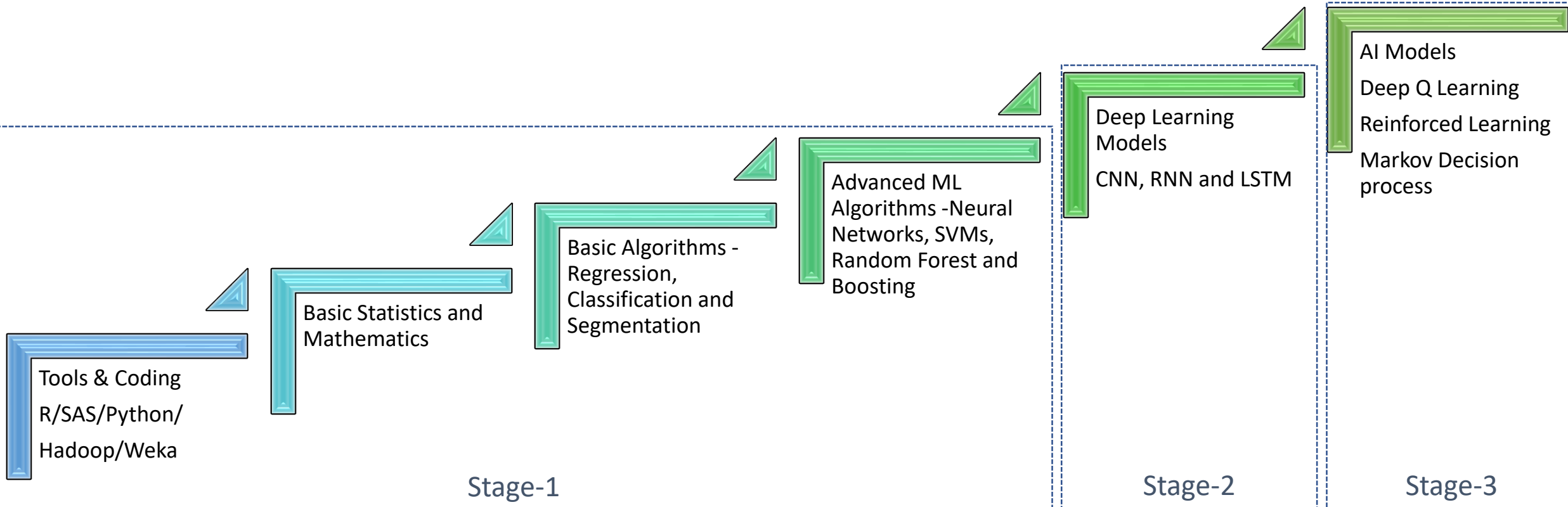
THE LEARNING PATH – OUR SUGGESTIONS

1. While learning these concepts, try to avoid academic style courses.
2. Look for the courses with lot of hands-on exercises and case studies

Tools & Coding
R/SAS/Python/
Hadoop/Weka

Basic Statistics and Mathematics

Basic Algorithms - Regression, Classification and Segmentation

Advanced ML Algorithms -Neural Networks, SVMs, Random Forest and Boosting

Deep Learning Models

CNN, RNN and LSTM

AI Models

Deep Q Learning

Reinforced Learning

Markov Decision process

Stage-1
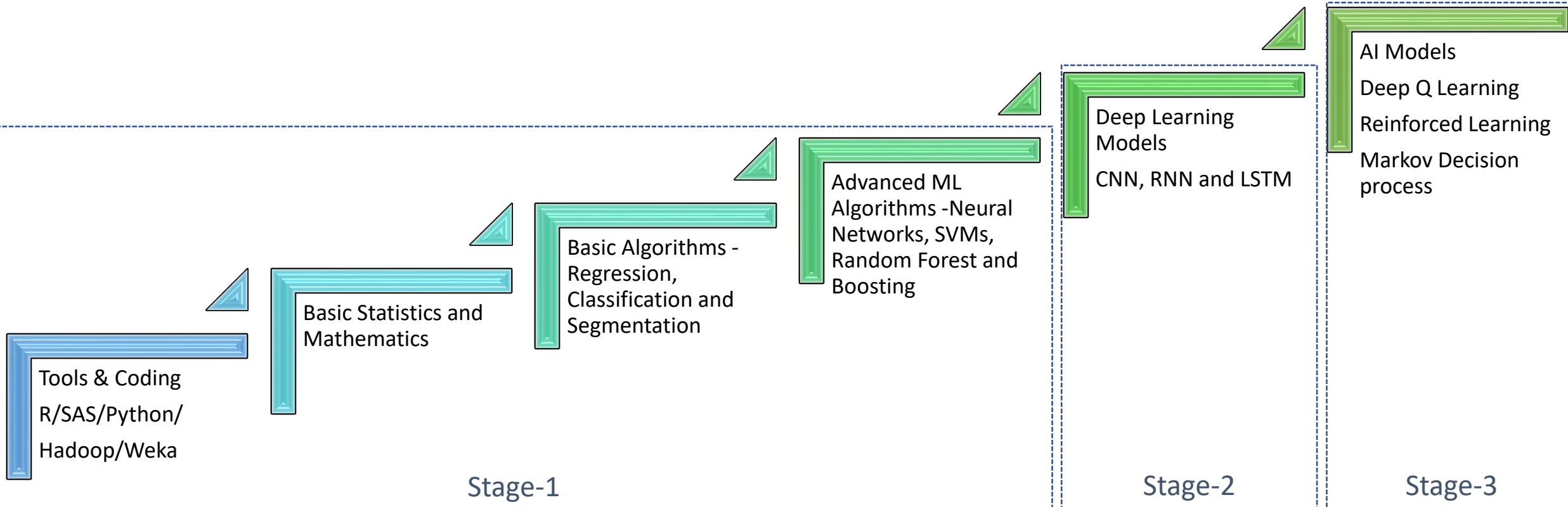
Stage-2

Stage-3

THE LEARNING PATH – OUR SUGGESTIONS

1. Do not focus on the tool, focus on the technique and algorithm
2. Learning python or R tool, will not make you a data scientist
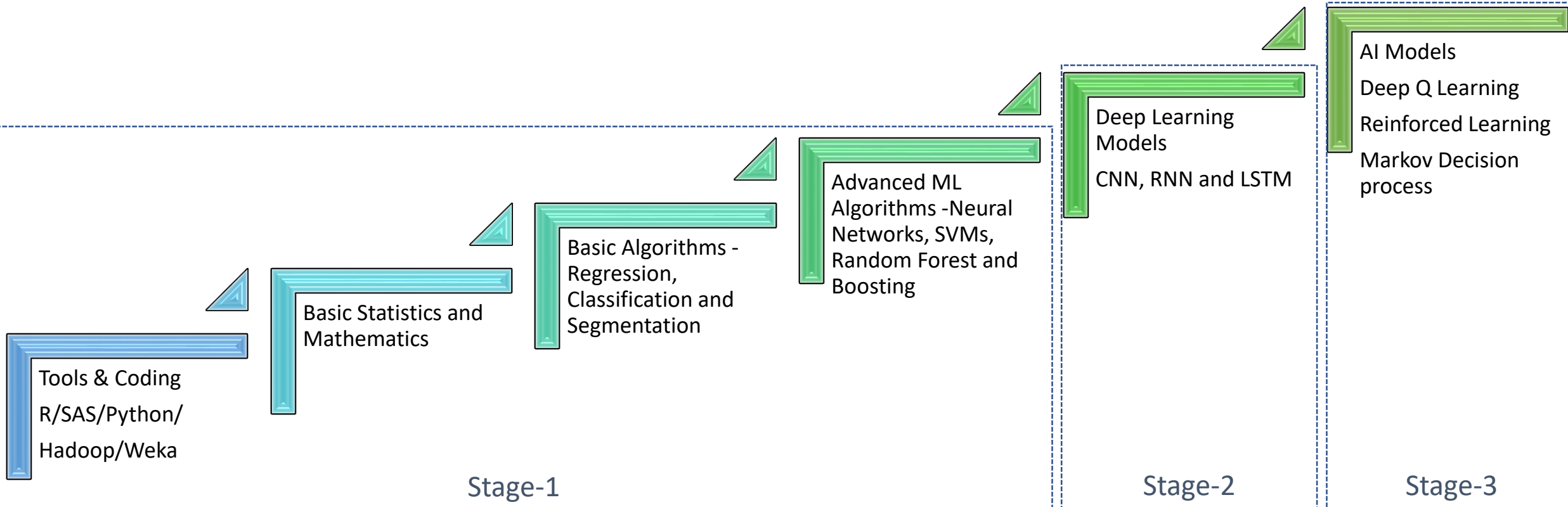
PART-6
Course Curriculum

FOCUS IS ON FIRST TWO STAGES

Tools & Coding

R/SAS/Python/

Hadoop/Weka

Basic Statistics and Mathematics

Basic Algorithms - Regression, Classification and Segmentation

Advanced ML Algorithms -Neural Networks, SVMs, Random Forest and Boosting

Deep Learning Models

CNN, RNN and LSTM

Stage-1

Stage-2

Stage-3

DURATION – 10 DAYS

Tools & Coding
R/SAS/Python/
Hadoop/Weka

Basic Statistics and Mathematics

Basic Algorithms - Regression, Classification and Segmentation

Advanced ML Algorithms -Neural Networks, SVMs, Random Forest and Boosting

Deep Learning Models

CNN, RNN and LSTM

Stage-1

Stage-2

Stage-3

TWO PHASES

PHASE-1 (5DAYS)

Python for data science

Data manipulations in python

Basic Statistics

Data validation and Cleaning

Regression

Logistic Regression

Decision Trees

Cluster Analysis

Model Selection and Cross validation

ANN – Artificial Neural networks

SVM – Support Vector Machines

Random Forest

Boosting

NLP & Text mining

TensorFlow & keras

Deep Learning Models

Convolution Neural Network

Recurrent Neural Networks

PHASE-2 (5DAYS)

statinfer

COURSE FEATURES

- 100% Hands-on Training

- 30 case studies laced in the course

- Created for Non- Statisticians

- Datasets from multiple domains, codes files and in class exercises

- Team assignments and mentoring

- Final Assessment

- E-learning material support

PART-7

Data Science and Machine Learning Myths

MYTH-1 : MATHEMATICS

Myth-1 : To be a good data scientist, you need to be exceptional at statistics, mathematics, calculous, algorithms etc.,

Not necessarily.

statinfer

MYTH-2 : PROGRAMMING

Myth-2 : To be a good data scientist, you need to have exceptional coding skills like Python, Java, C++ etc.,

Not necessarily.

MYTH-3 : COMPLICATED MODELS

Myth-3 :Data science is all about building complex predictive and machine learning models to solving business problems

Not necessarily.

MYTH-4 : MODEL BUILDING

Myth-4 :After collection of the data, most of the time is spent on model building process.

Not necessarily.

MYTH-5 : LARGE DATASETS

Myth-5:While solving machine learning problems we need to handle really large datasets or most of the datasets are really large

Not necessarily.

MYTH-6 : MACHINE LEARNING IN BUSINESS

Myth-6:Companies use really advanced deep learning and AI models for while building all their business strategies

Not necessarily.

MYTH-7 : DIVERSE ALGORITHMS

Myth-7: A data scientist will be using all the models in their day to day life
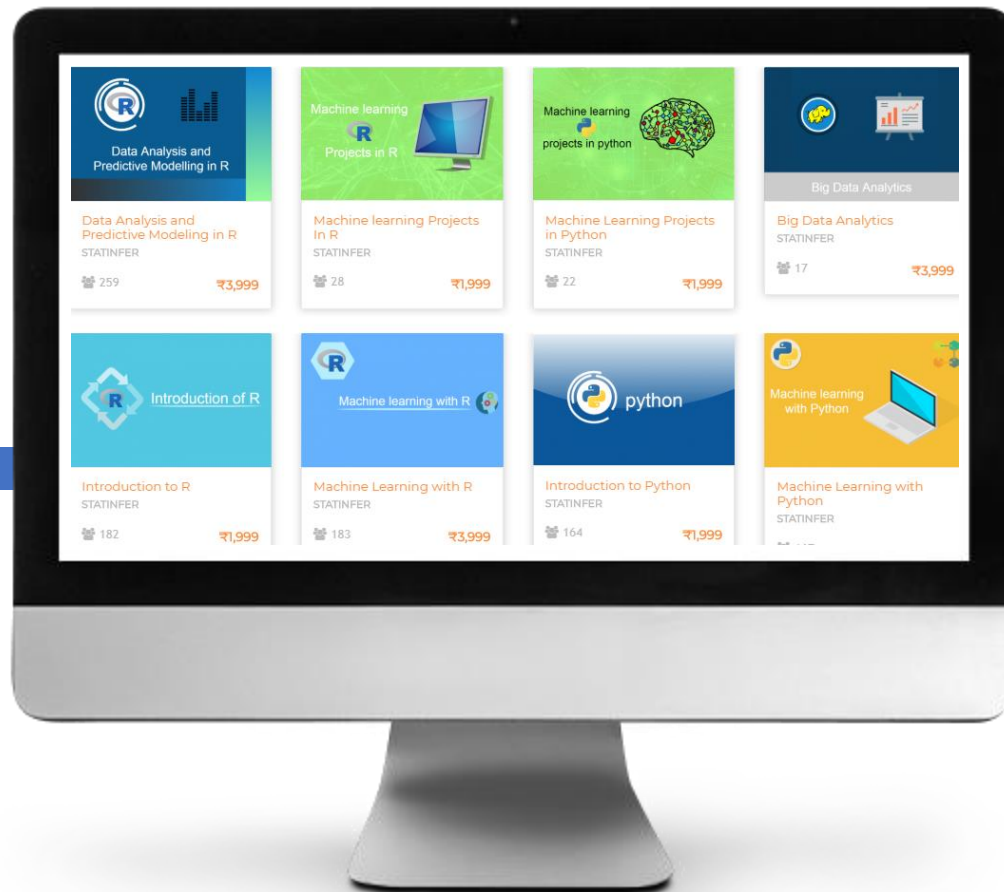
Not necessarily.

Have you tried our app?

We have data science and machine learning quiz app to make the learning easy and fun

Click here

Our e-Learning Modules

Currently all the Machine Learning and Data Science courses are available as e-learning modules

TRAINER PROFILE

VENKATA REDDY KONASANI

Trainer in
- Data Science
- Machine Learning
- Deep Learning
- Artificial Intelligence

ABOUT VENKATA REDDY KONASANI

5,000+
Training Hours

1,500+
Participants

75+
Corporate Batches

25+
Companies

- 11 Years into Data Analytics
- 5+ Years into Training
- Author of the book "Practical Business Analytics using SAS"
- Statinfer.com – Key member in the core team
- Work Experience
  - HP – Data Scientist
  - Trend wise Analytics -Data Scientist
  - HSBC – Data Analyst
  - Citi – Risk Analyst
- Masters in Applied Statistics and Informatics from IIT Bombay

statinfer

statinfer

THANK YOU

✉ VENKAT@STATINFER.COM

✳ WWW.STATINFER.COM