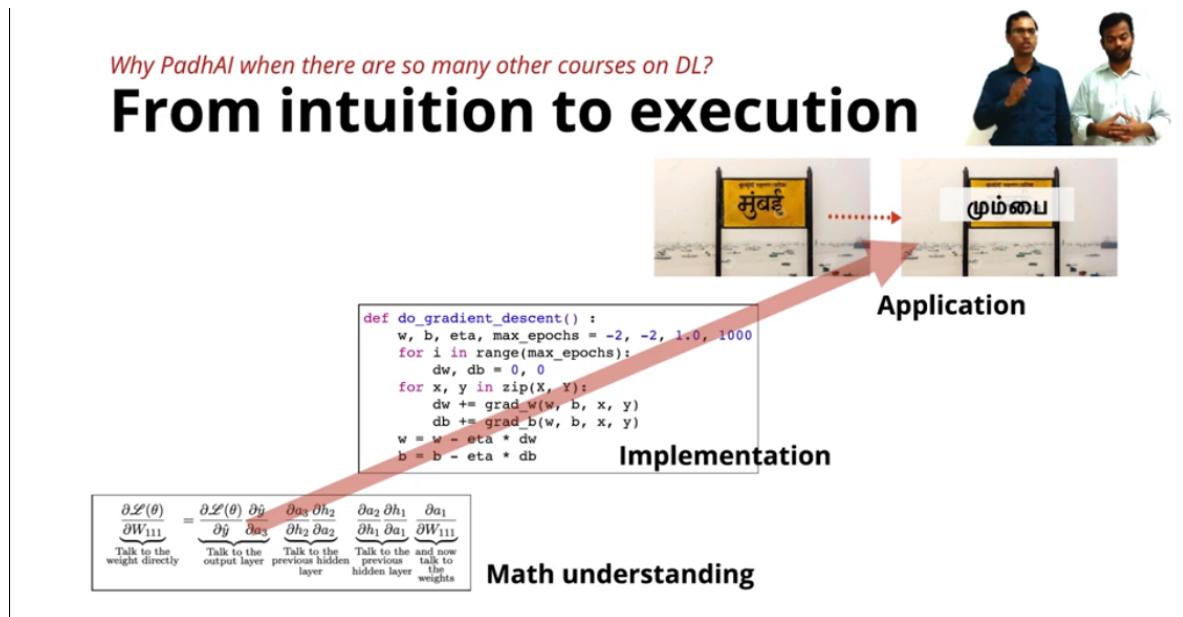


Hands-on deep learning course with focus on solving problems in India

- Mitesh Khapra
- Pratyush Kumar
- IIT Madras and One Fourth labs
- Intuition -----> execution
- Math understanding ---> Implementation ---> Application

From Intuition to execution



- 1) Math understanding and theory
- 2) Implementation in coding
- 3) Application : Sign board detection

Final capstone: Syllabus

Input : An image that contains a sign board

What to do : Develop a deep learning model

Output : Automatically translate this in to a language of choice

4 modules --> 4 months

How to engage throughout the course?



Learn-compete-learn

Module 1		<i>Does this image have text?</i> Perceptron, Shallow neural networks, loss functions, gradient descent, NumPy basics		kaggle
Module 2		<i>Which character is this?</i> Deep neural networks, TensorFlow, Stochastic gradient descent, Hyperparameter tuning, Optimisation algorithms, Bias-variance trade-off		kaggle
Module 3		<i>Which regions have text? How to segment into chars?</i> CNN, Regularisation, Batch Normalisation, Object detection networks, Region proposal networks, Segmentation network, Visualising CNNs		kaggle
Module 4		<i>Which combination of characters is a likely word?</i> Sequence Modelling, Word2Vec, RNN, LSTM, GRU, Encoder-decoder models, Attention mechanisms		kaggle
Final Capstone		<i>How to deploy this in the real world?</i> TensorFlow Lite, Android		DL garage

Tools of the trade

--> Google colab

--> Kaggle

--> Github

--> Medium

Month 1: Given an image --> Binary classification --> Whether the image contains text or not?

Jan-29	Jan-31	Theory	Jargon Busting
Feb-1	Feb-1	Gyaan	This video
Feb-2	Feb-3	Hands-on	Python Basics 1, 2
Feb-4	Feb-6	Theory	Expert Systems, 6 jars
Feb-7	Feb-8	Hands-on	Python Linear Algebra
Feb-9	Feb-10	Theory	MP Neuron, Perceptron
Feb-11	Feb-13	Hands-on	Python MP Neuron, Perceptron, Test/Train
Feb-14	Feb-15	Gyaan	Kaggle, Github, Reproducability, Test/Train, Google Facet
Feb-16	Feb-17	Contest	1.1 Mobile phone like/dislike predictor
Feb-18	Feb-20	Theory	Sigmoid Neuron, Gradient Descent
Feb-21	Feb-22	Hands-on	Python Sigmoid, Gradient Descent
Feb-23	Feb-24	Theory	Probability Theory
Feb-25	Feb-27	Hands-on	Engineering best practices: setting learning rate, initial weights, data augmentation, data normalization, class imbalance
Feb-28	Mar-1	Contest	1.2 Binary Text/NoText Classification
Mar-2	Mar-3	Contest	1.3 Binary Text/NoText Classification Advanced

Month 2: Given an image --> Recognize characters in English and Hindi

Mar-4	Mar-6	Theory	Derivatives, Partial Derivatives, Gradients	Representation Power of functions, Feedforward Neural Networks, Multiclass classification
Mar-7	Mar-8	Hands-on	Contest 1.2 analysis	Contest 1.3 analysis
Mar-9	Mar-10	Theory	Backpropagation (scalar), mini-batch training	Live Session
Mar-11	Mar-13	Theory	Backpropagation (vectorized)	
Mar-14	Mar-15	Hands-on	Numpy Feedforward Neural Networks, Backpropagation	Numpy Backpropagation (vectorized)
Mar-16	Mar-17	Contest	2.1 Classify Language of the text	
Mar-18	Mar-20	Theory	Optimization Algorithms (Momentum, Nesterov, AdaGrad, RMSProp, Adam)	Bias correction in Adam
Mar-21	Mar-22	Hands-on	Numpy Optimization Algorithms	
Mar-23	Mar-24	Theory	Activation Functions, Initialization Methods, Hyperparameter Tuning, Bias Variance Tradeoff, Regularization (L2)	
Mar-25	Mar-27	Hands-on	MLFlow, Experimenting with Deep models	
Mar-28	Mar-29	Contest	2.2 All characters of English	
Mar-30	Mar-31	Contest	2.3 All glyphs in Hindi	Live Session

Month 3: Object detection and CNN classifier for a single character in an image

Apr-25	Apr-26	Theory	Sequence Modeling, Recurrent Neural Networks, BPTT	Vanishing Gradient Problem
Apr-27	Apr-28	Hands-on	Contest 3.1, 3.2 analysis	
Apr-29	May-1	Theory	LSTMs, GRUs	How LSTMs solve the Vanishing Gradient problem
May-2	May-3	Hands-on	PyTorch RNN, LSTM, GRUs	
May-4	May-5	Theory	Encoder decoder models, attention mechanism	
May-6	May-8	Hands-on	PyTorch Attention models	
May-9	May-10	Contest	4.1 English to Hindi transliteration with provided data	
May-11	May-12	Contest	4.2 Scrape dataset for Hindi to Tamil transliteration	

Month 4: NLP

May-13	May-15	Capstone	Problem Setting	
May-16	May-17	Hands-on	Contest 4.1, 4.2 analysis	
May-18	May-19	Theory	PixelLink, CRNN	History of Deep Learning, Current trends
May-20	May-22	Hands-on	Overview of Tensorflow, Keras	
May-23	May-24	Hands-on	Tensorflow PixelLink, CRNN	
May-25	May-26	Hands-on	Tensorflow Lite	Live Session
May-27	May-29	Capstone	Problem Solving	
May-30	May-31	Gyaan	Announcement of winners of garage, discussion of next steps	

Final capstone: Advanced

May-13	May-15	Capstone	Problem Setting	
May-16	May-17	Hands-on	Contest 4.1, 4.2 analysis	
May-18	May-19	Theory	PixelLink, CRNN	History of Deep Learning, Current trends
May-20	May-22	Hands-on	Overview of Tensorflow, Keras	
May-23	May-24	Hands-on	Tensorflow PixelLink, CRNN	
May-25	May-26	Hands-on	Tensorflow Lite	Live Session
May-27	May-29	Capstone	Problem Solving	
May-30	May-31	Gyaan	Announcement of winners of garage, discussion of next steps	

Jargon Bursting

A lot of confusing buzz words... @IITM

The Jargon Bubble

Robert Bosch Center for Data Science and Artificial Intelligence

Introduction to Pattern Recognition and Machine Learning

Deep Learning for Natural Language Processing

Image Processing and Computer Vision Lab

Speech Technology

Deep Reinforcement Learning

Let's look at some definitions...

The Jargon Bubble

Artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals.

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" with data, and without being explicitly programmed.

*Does AI not involve
statistical techniques ?*



Let's look at some definitions...

The Jargon Bubble

Artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals.

Data science is an **interdisciplinary** field that uses **scientific methods, processes, algorithms** and systems to extract **knowledge** and insights from data in various forms, **both structured and unstructured**, similar to data mining.

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" with data, and without being explicitly programmed.

*Is AI/ML not
interdisciplinary ?*



Let's look at some definitions...

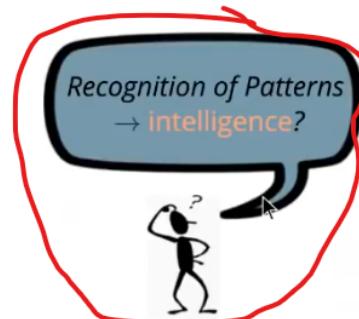
The Jargon Bubble

Artificial intelligence (AI), sometimes called machine intelligence, is **intelligence demonstrated by machines**, in contrast to the natural intelligence displayed by humans and other animals.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to **extract knowledge and insights from data** in various forms, both structured and unstructured, similar to data mining.

Pattern recognition is the automated recognition of patterns and regularities in data.

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" with **data**, and without being explicitly programmed.



Let's look at some definitions...

The Jargon Bubble

Artificial intelligence (AI), sometimes called machine intelligence, is **intelligence demonstrated by machines**, in contrast to the natural intelligence displayed by humans and other animals.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to **extract knowledge and insights from data** in various forms, both structured and unstructured, similar to data mining.

Pattern recognition is the automated recognition of patterns and regularities in data.

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" with **data**, and without being explicitly programmed.

Computer vision is an **interdisciplinary** field that deals with how computers can be made to gain **high-level understanding** from digital **images or videos** and **automate tasks that the human visual system** can do.

Let's look at some definitions...

The Jargon Bubble

Artificial intelligence (AI), sometimes called machine intelligence, is **intelligence demonstrated by machines**, in contrast to the natural intelligence displayed by humans and other animals.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to **extract knowledge and insights from data** in various forms, both structured and unstructured, similar to data mining.

Pattern recognition is the automated recognition of patterns and regularities in data.

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" with data, and without being explicitly programmed.

Deep Learning is a new area of **Machine Learning research**, which has been introduced with the objective of moving Machine Learning closer to one of its original goals: **Artificial Intelligence**.

Let's look at some definitions... (still confusing)

The Jargon Bubble

Artificial intelligence (AI),

demonstrated by machines

Data science

uses scientific methods,
processes, algorithms
and systems to extract
insights from data

Pattern recognition
recognition of patterns
and regularities in
data.

Machine learning

Is Automatic Recognition
of digits AI or ML or DS
or PR or CV or DL? Why?



What is AI? A framework to think and understand

Still Confusing....

What is AI?

Let's try an alternative approach to resolve this confusion....

We know that humans are intelligent beings so let's start from there...



What is AI

1) Abilities

- **computer Vision** : see, identify things : perception
- **Speech** : Hear and listen
- **Natural language processing**: Read and write documents
- **Decision making and planning**

2) Tasks

- **Computer vision:**
 - Make sense of patterns and distinguish or classify images
 - Horse or Giraffe
 - Pattern recognition
 - hand written digits recognition
 - Object detection or recognition or boundary detection
 - Flight
 - Pose estimation
 - Hand gesture recognition
- **Speech**
 - Speech synthesis or speech generation
 - Read out a text
 - Speech recognition
 - Understand audio input
 - Speaker recognition
 - Identify who is speaking
- **Natural language processing**
 - Document classification
 - Read document and classify them
 - Spam classification

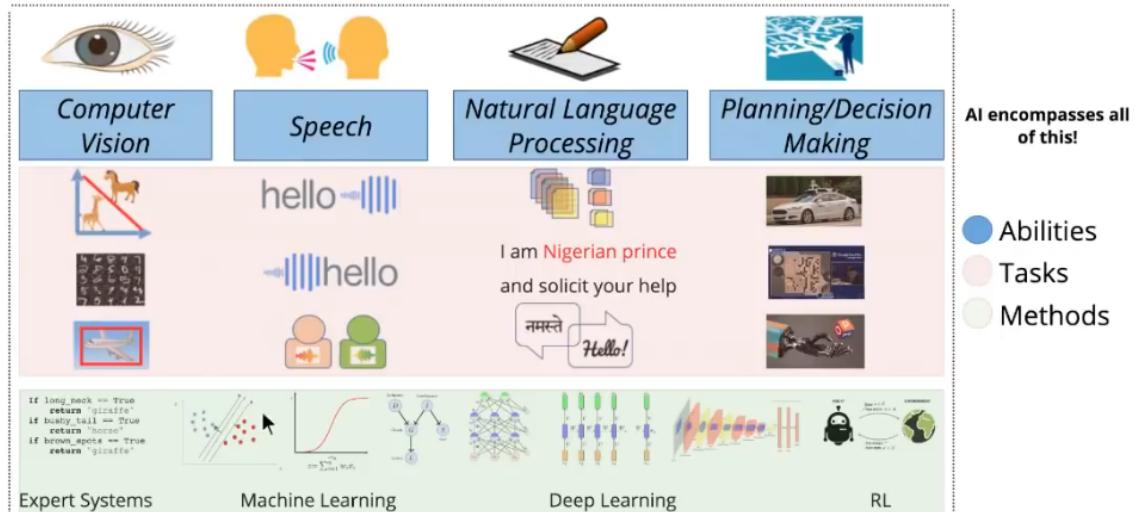
- Conversation
- Machine Translation
- Summarization
- Question answering
- **Planning and decision making**
 - Self driving cars
 - Make use other abilities like computer vision
 - Game playing
 - Robotics

3) Methods

- Expert systems:
 - Rule based
 - Knowledge based
 - Supervisor knowledge
 - No learning
 - Explicitly programmed
- Machine learning
 - Labelled data
 - Learn input and output relationship
 - Estimate the parameters
 - SVM, Logistic regression, Graphical models, decision tree, naïve Bayes
- Deep learning
 - Neural networks
 - Feed forward networks
 - Recurrent neural networks
 - Convolutional neural networks
 - Customized models
 - More complex models
 - large amounts of data
 - Enough engineering of parameters
 - Highly performant
- Reinforcement learning

Abilities, Tasks and Methods

What is AI?



Generalized AI:

Multiple abilities, Multiple tasks

Narrow AI:

Specific tasks

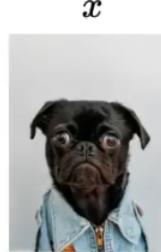
Knowledge based systems vs Machine learning

Explicitly Programmed v/s Learning Based Systems

Zooming in into ML

Explicitly Programmed

```
If long_neck == True
    return "giraffe"
if bushy_tail == True
    return "horse"
if brown_spots == True
    return "giraffe"
if tusk == True
    return "elephant"
if long_tongue == True
    return "dog"
if count(round_objects) == 4
    return "vehicle"
```



- y**
- bat
 - car
 - dog
 - cat
 - ship

Knowledge based: Long list of rules

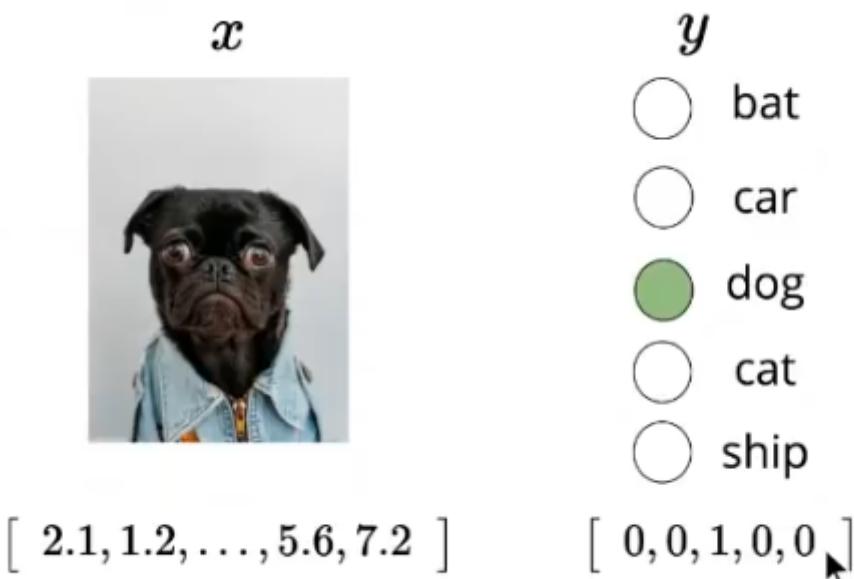
Machine learning:

each Image = 30 x 30 grey scale image

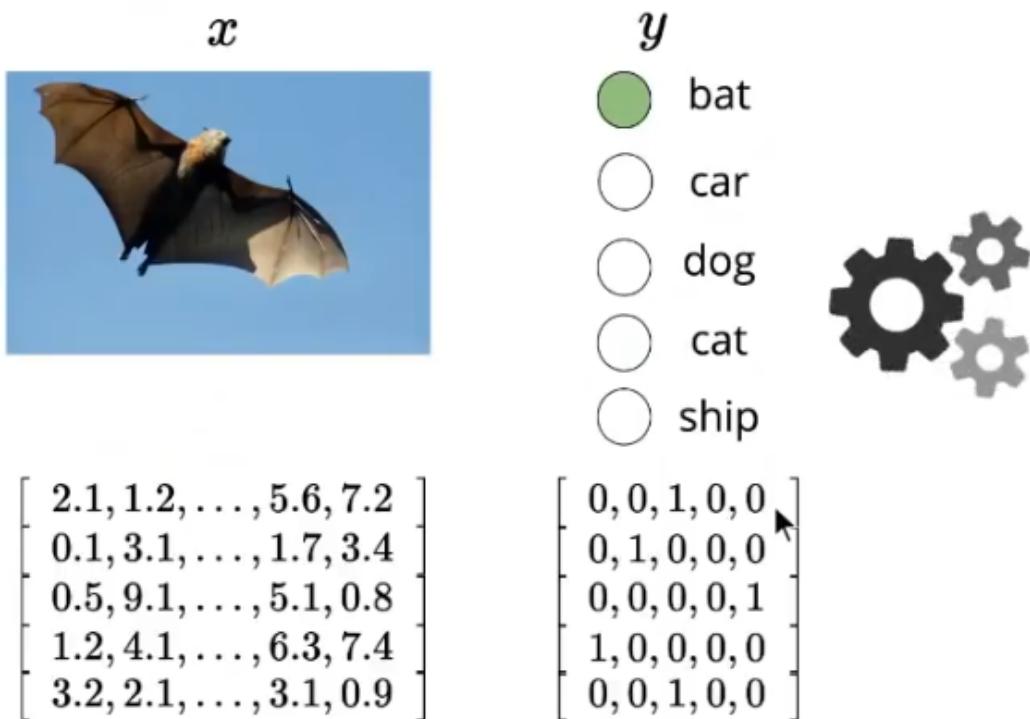
each pixel tell intensity of the pixel

It will be a value from 0-255.

A single labelled example



Now we can give **many such labelled images to system**. We will also tell the outputs.

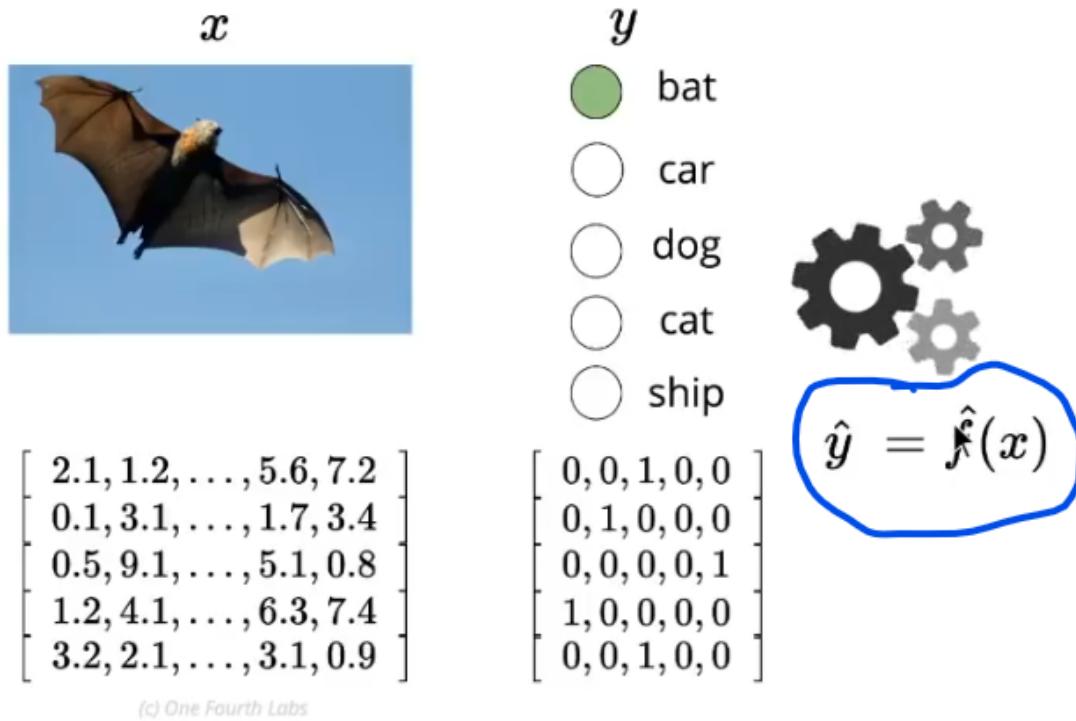


(c) One Fourth Labs

We will tell **which function best describes input and output relationship**

example: SVM, graphical model or neural network etc.,

- 1) Supervised learning: Labelled data has to be provided by humans
- 2) function or a family of functions has to be provided by humans
- 3) Machine then learns parameters of the function



(c) One Fourth Labs

4) Engineering has to be done by humans

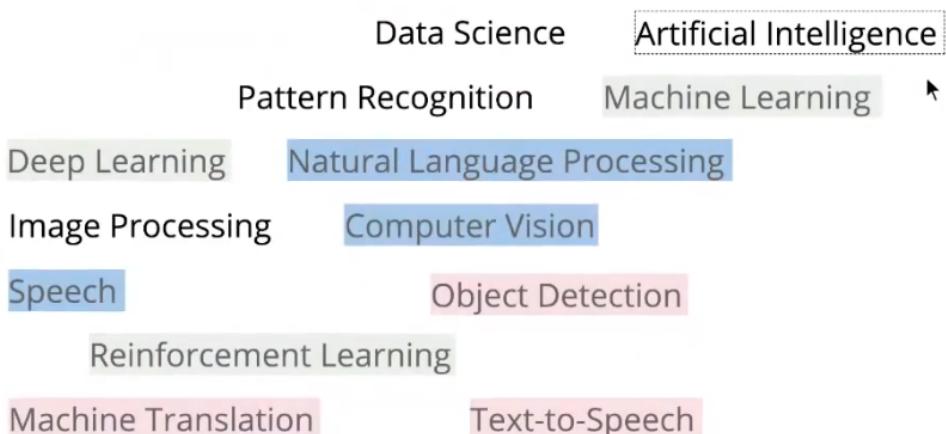
- Hyperparameter tuning, fix learning rate etc.,

Source: <https://www.semanticscience.org/>

Revisiting the buzz words

What is AI?

- Abilities
- Tasks
- Methods



Abilities

- Natural language processing
- Computer vision
- Speech

Tasks

- Object detection
- Machine translation
- Text to speech

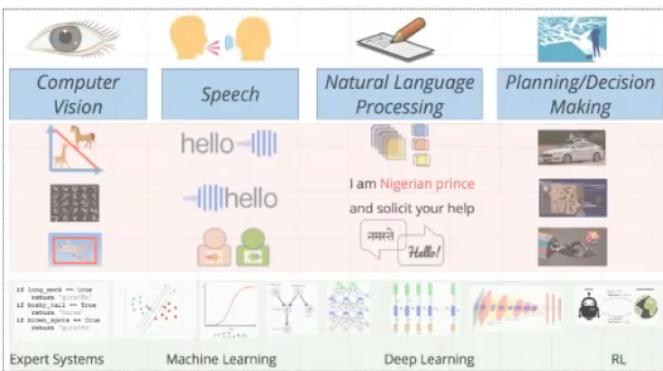
Methods

- Machine learning
- Deep Learning
- Reinforcement learning

What about Data science, Pattern recognition and image processing?

Most AI tasks require pattern recognition...

How is Pattern Recognition different?

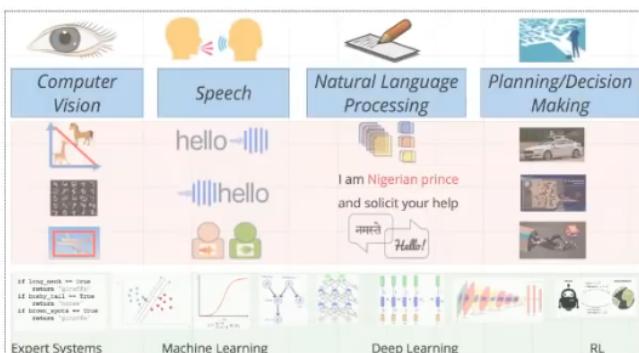


The field of **pattern recognition** is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions **such as** classifying the data into different categories.

Pattern recognition --> then --> Actions

Most AI tasks require pattern recognition...

How is Pattern Recognition different?



The field of **pattern recognition** is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions **such as classification, regression, clustering, generation.**

May span different abilities ●

May be solved by different methods ○

Is image processing different from computer vision?

A subtle difference...

Is Image Processing different from CV ?

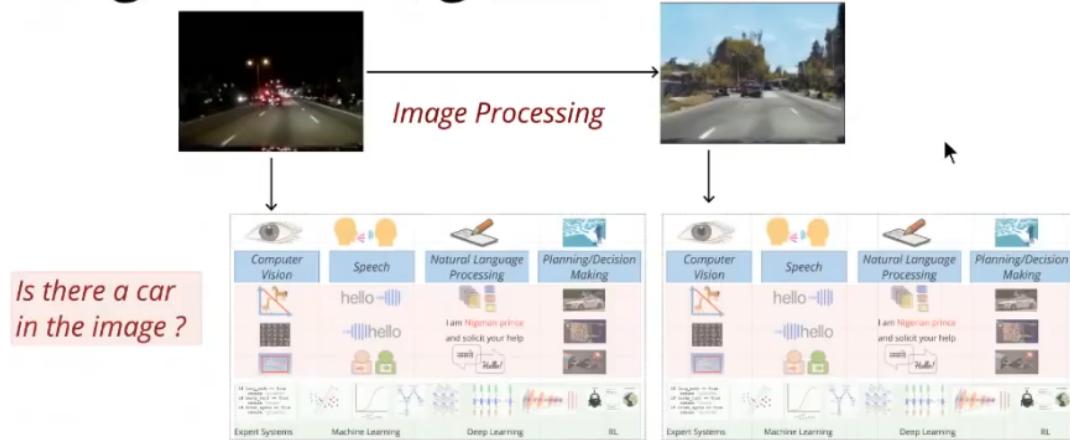


Image processing:

- Input is image and output is image
 - Input : Image
 - Output : Image (After some transformation)
- convert Night time image to day time image
 - Now the task of computer vision becomes much easier
- "Black and white image" to color image
- Low resolution image to high resolution image

Computer vision

- classification task
 - Input : Image
 - Output : Which class?
- Object detection
 - Input : Image
 - Output : Rectangular box in which the object lies
- Posture Recognition
 - Input : Image
 - Output : What is the pose person is having

Note

- 1) Image processing can make input clear to computer vision

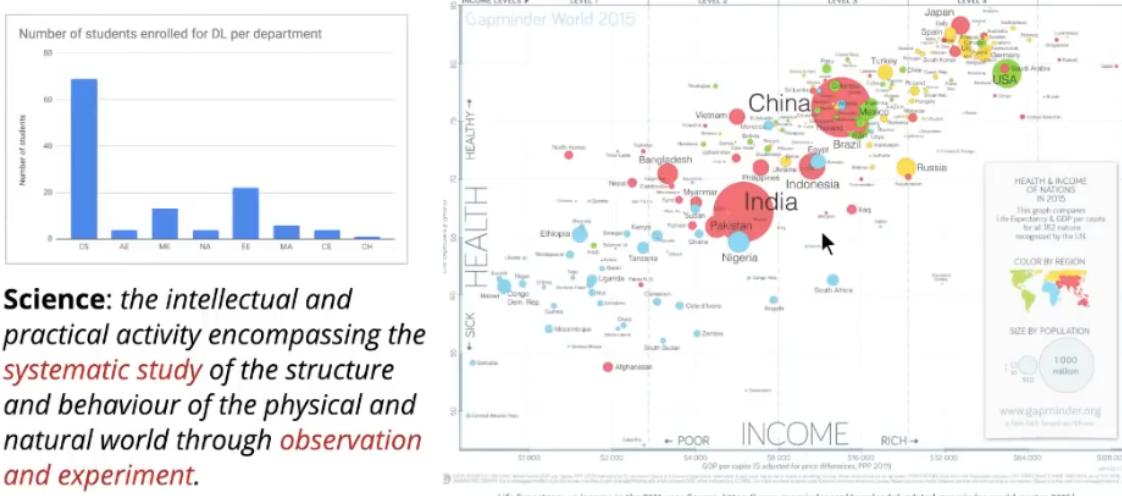
2) The same can be done with computer vision

- Give a model lot of training data ("night time images" ---> "day time image")
- We can make model learn to do the same task

What is Data science?

Is it a subset of AI, a super set or neither ?

What is Data Science?

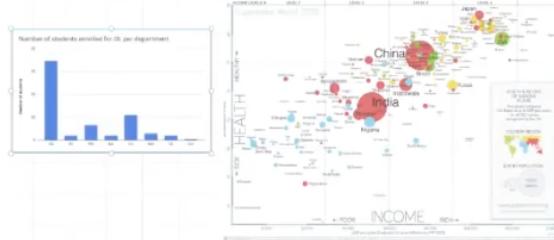


Science: the intellectual and practical activity encompassing the systematic study of the structure and behaviour of the physical and natural world through observation and experiment.

- Data visualization
- No learning
- Data science
 - Systematic study
 - Observation
 - Experiment

Google

Is it a subset of AI, a super set or neither ?
What is Data Science?



Descriptive Statistics

I would prefer the term Data Science when I am dealing with numeric data (think of database tables, employee, customer, sales, etc.)



How many students will enrol in Jan 2021 ?

When will the average life expectancy in India go upto 85 ?

Predictive Statistics

Structured data

---> prefer data science

---> Broader

Unstructured data

---> Image : computer vision

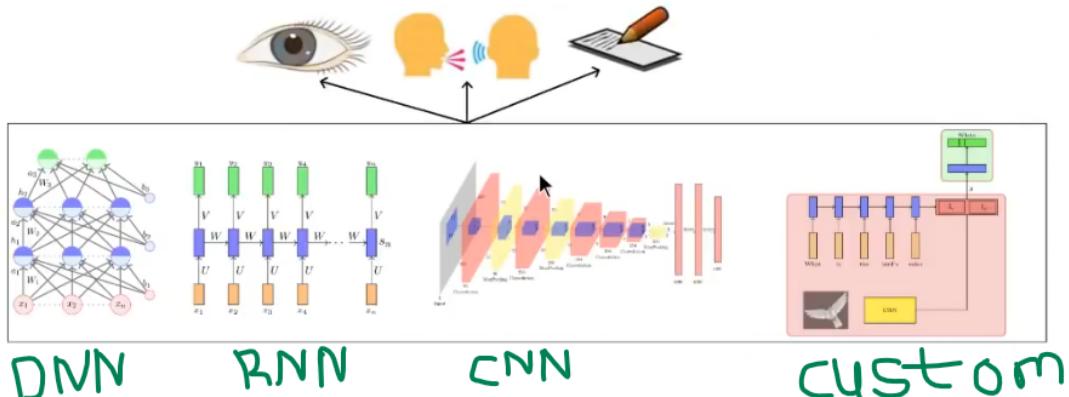
---> Audio : NLP

Final : DL landscape

Clouds

Wide applicability, Data/Compute intensive, Hard to Interpret

The DL landscape



- We can attack any task using deep learning
- Disadvantage
 - Data intensive : lot of supervised data
 - High computing power : GPUs and TPUs
 - Hard to interpret the output
 - Prone to adversarial affects / attacks
 - a system can detect panda
 - Now add some random noise to image
 - Human eye wont get confused. no difference
 - But the deep learning network will get confused
 - We don't know how the DNN is coming to a decision

Wide applicability, Data/Compute intensive, Hard to Interpret

The DL landscape

