Artificial Intelligence (Machine Learning)

Dr ABDULLAH

BS(IT) 5th

Assistant Professor, Department of Computer Science & Information Technology, FMCS, The University of Agriculture Peshawar

DEEP LEARNING

CONTENTS

- Course Pre-requisites
- Basics of Machine Learning (ML)
- Popular ML Approaches
- Limitations of Traditional ML
- Deep Learning (DL) and its Benefits
- Brief History of DL
- Applications
- Milestones
- Challenges
- Current Research Trends

COURSE PRE-REQUISITES

- Calculus, Linear Algebra
  - Derivatives
  - Matrix operations
- Probability and Statistics
- Advanced programming
- Basic Machine Learning (Neural Network)

COURSE OVERVIEW

- Objectives
  - Understanding the theoretical concepts of ML
  - Get hands-on practical experience
- What will you learn?
  - How to build, train, & deploy ML models ?
  - ML for Computer Vision & other problems
  - Deep Sequence Learning
  - Unsupervised and Bayesian Deep Learning
  - Deep Reinforcement Learning
  - Generative Adversarial Networks

Introduction of Machine Learning

- What is machine learning
- Difference between machine learning, deep learning & data science
- Application of machine learning
- Types of machine learning
- Difference between Supervised, Unsupervised & Reinforcement learning
- Life cycle of machine learning
- Bias – varience trade-off
- Overfitting & underfitting

Machine Learning

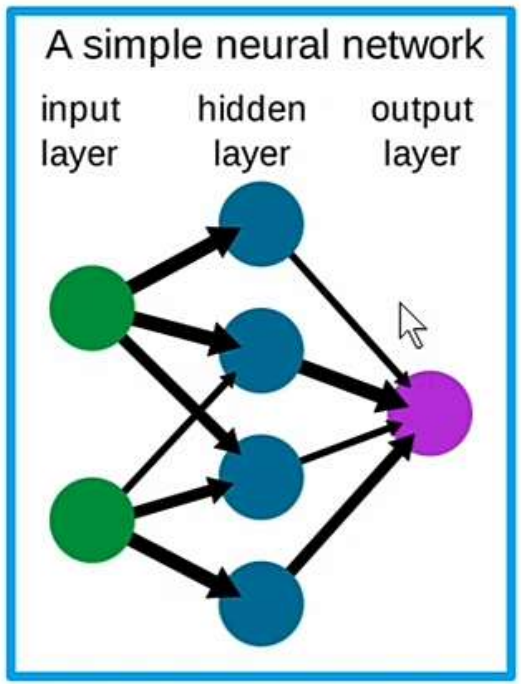
- Machine Learning is a subset of AI.
- Machine Learning is a set of algorithms that train on a data set to make predictions or take actions in order to optimize some systems.

Deep Learning

- Deep Learning is a next evolution of machine learning.
- The term Deep Learning was introduced to the machine learning community in 1986
- Deep Learning is a subset of Machine Learning Where learning method is based on data representation or feature learning.
- “Deep” refers to 1 or more hidden layers in this case.

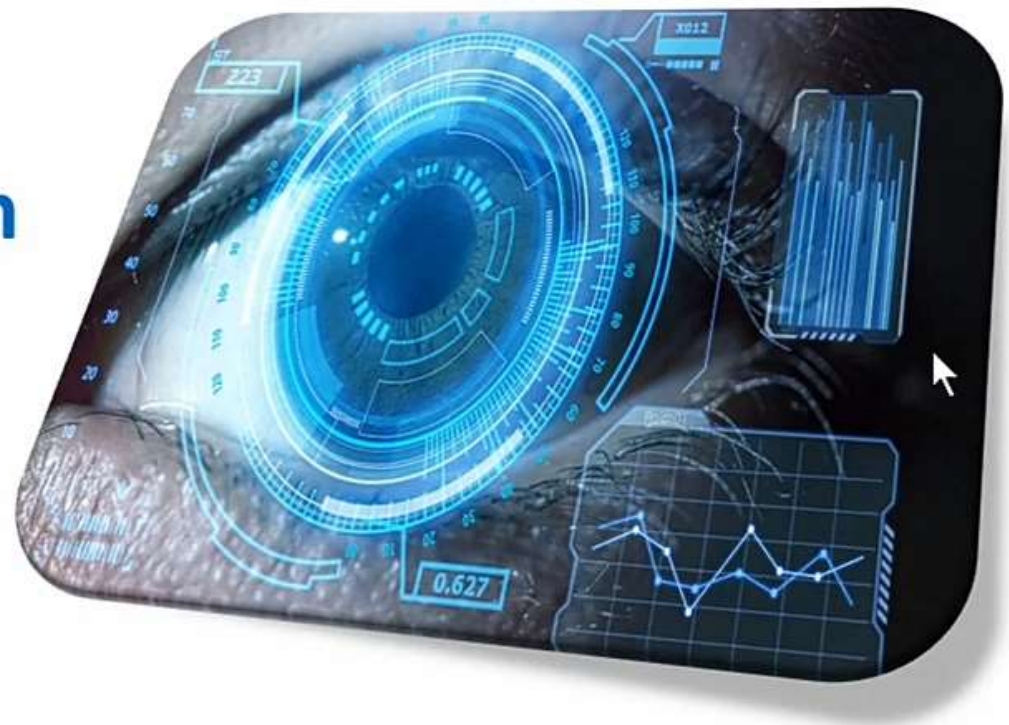
Deep Learning

- **Artificial Neural Network(ANN)**
- **Convolutional Neural Network(CNN)**
- **Recurrent Neural Network(RNN)**
- **Transfer Learning**
- **AutoEncoders**



Computer Vision

- **Techniques:**
  - **Image Classification**
  - **Object Detection**
  - **Semantic Segmentation**
  - **Object Tracking**
  - **Instance Segmentation**



Computer Vision Algorithms / Model

- CNN,
- AlexNet, ResNet, Inception, VGG16
- R-CNN, Fast R-CNN, Faster R-CNN, Masked R-CNN, R-FCN
- YOLOv1, YOLOv2, YOLOv3,YOLOv4,YOLOv5
- TFOD, SSD
- GANs, GNNs, Visual Search
- DeepSORT, Detectron2
- And much more……

Data Science

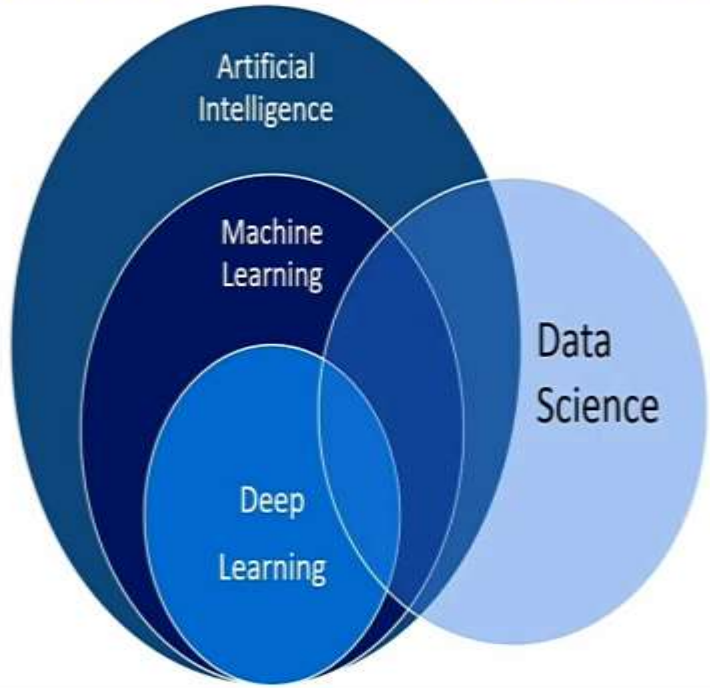
Data Science is the **Art and Science of drawing actionable insights from the data.**

Applications:

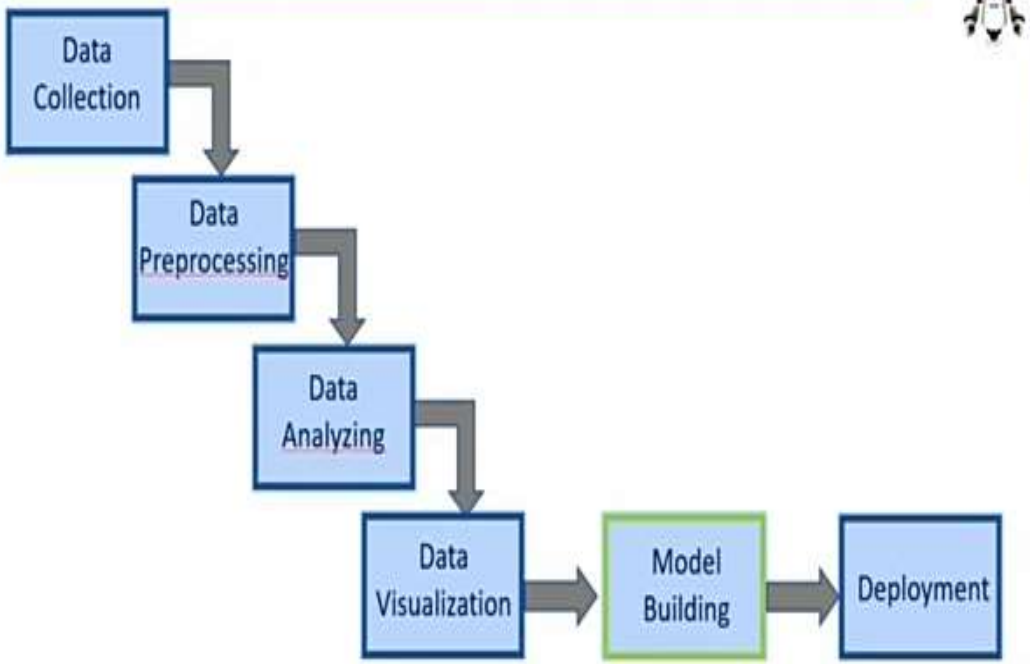
Retail, Bank, E-Commerce, Healthcare and Telecom etc.

Data Science

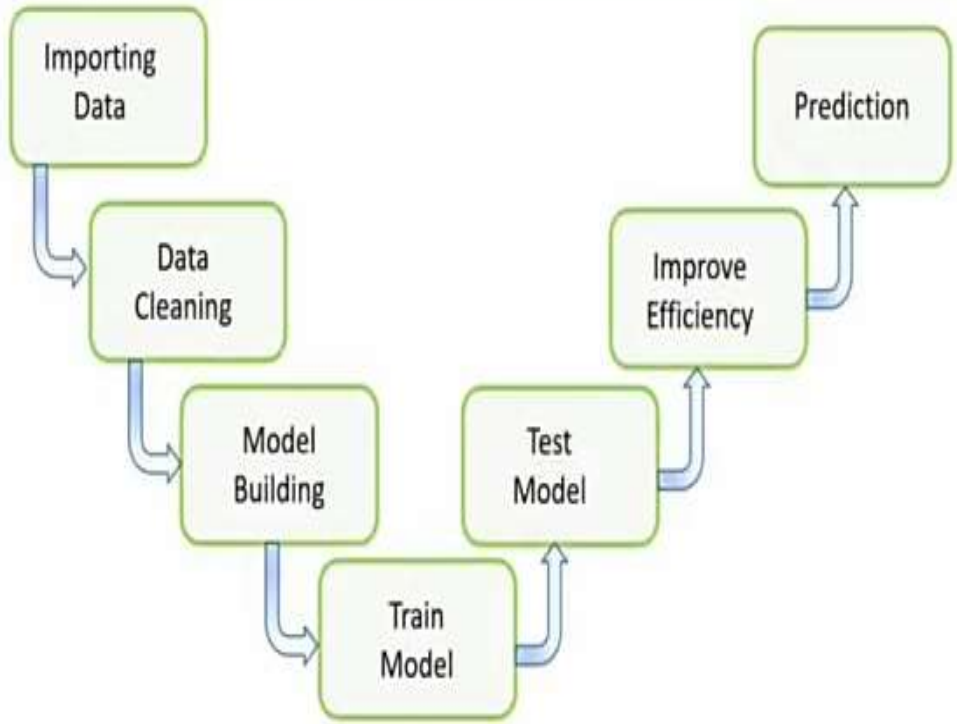
Data Science has an intersection with artificial intelligence but is not a subset of artificial intelligence.



Data Science Flow Chart

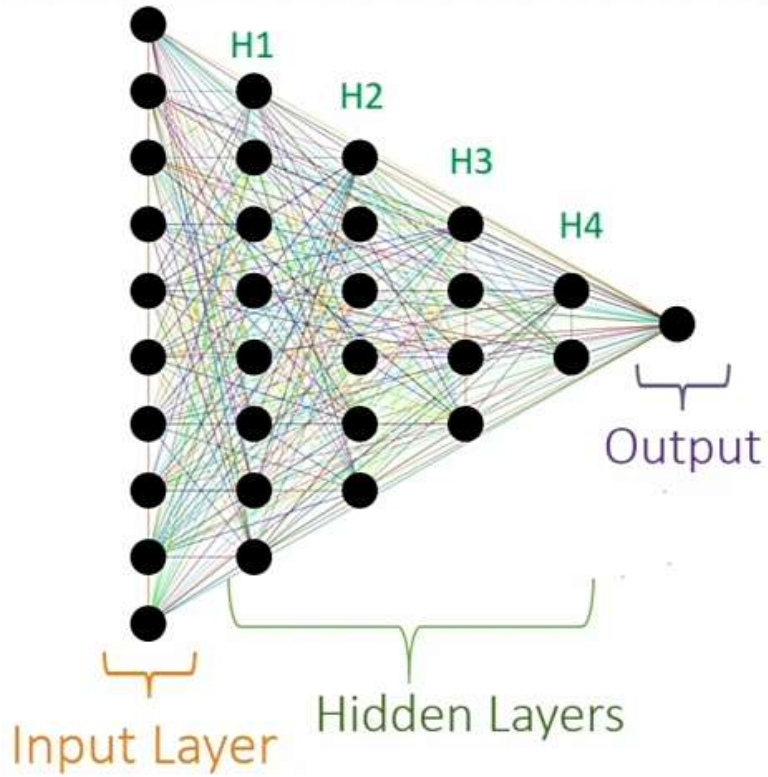


Model Building



Deep Learning

- In Deep Learning data goes through multiple numbers of non-linear transformation obtain an output.

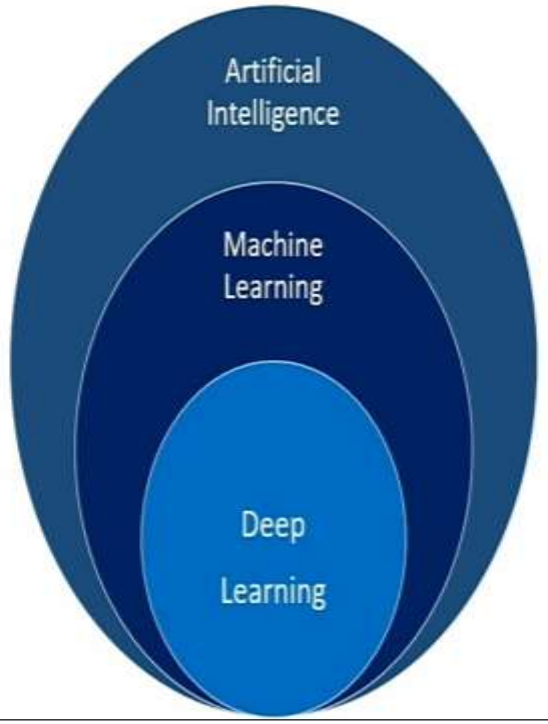


AI vs ML vs DL

Artificial Intelligence – “Human intelligence exhibited by machines”

Machine Learning – “ An approach to achieve Artificial Intelligence”

Deep Learning – “A technique for implementing machine learning”



Formal Definition of Machine Learning

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

-Tom M. Mitchell

How Does Machine Learning Works ?

Machine Learning algorithm is trained using a training data set to create a model. When new input data(test data) is introduced to the ML algorithm, it makes a prediction on the basis of the model.

Traditional Programming vs Machine Learning

Artificial Intelligence

- AI(Artificial Intelligence) is a subfield of Computer Science.
- AI term was coined by John McCarthy in year 1956s.
- Artificial Intelligence, term was meant to describe the goal that machines will be able to have humans like intelligence in future.

- Intelligence – "the ability to learn or understand or deal with new or trying situations."

- Artificial Intelligence can be split between two branches:
  1. Applied AI(weak AI)
  2. Generalized AI (strong AI)

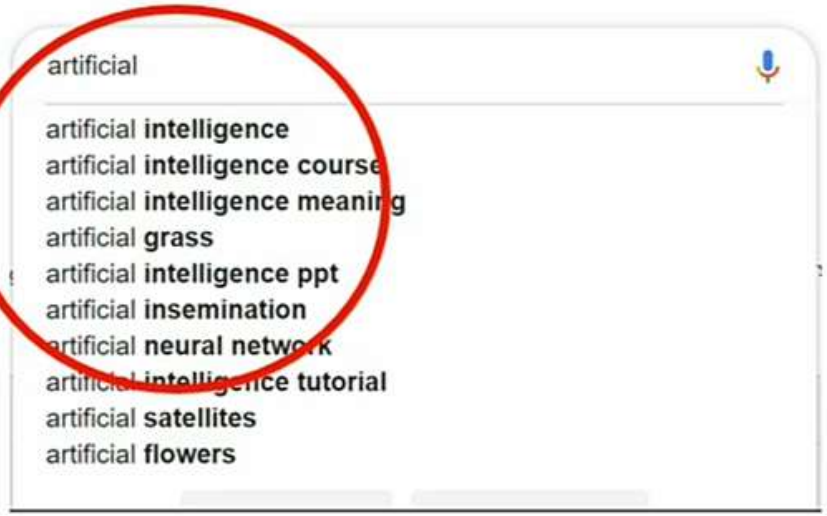
1.Applied AI(weak AI)- perform some specific tasks.

Alexa

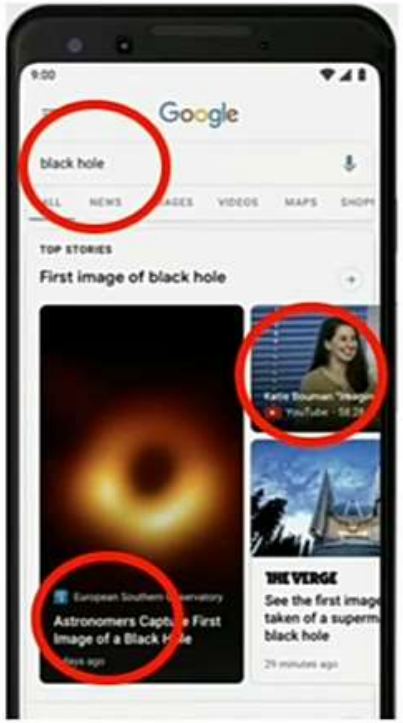
Google Assistant

2.Generalized AI(strong AI)- acts like humans.

What is Machine Learning



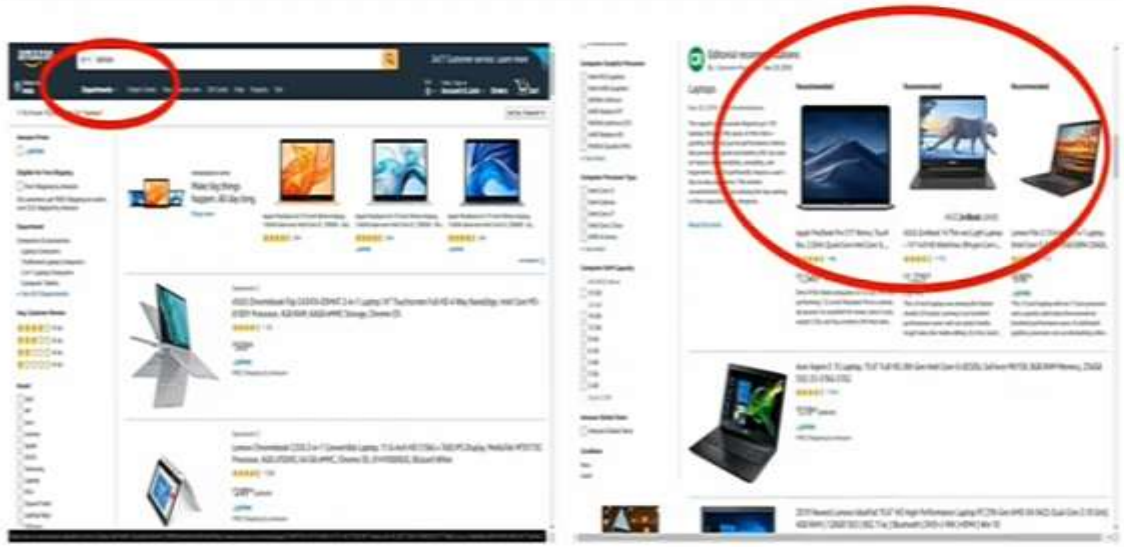
Google Recommendation



Youtube video timing - 4:47

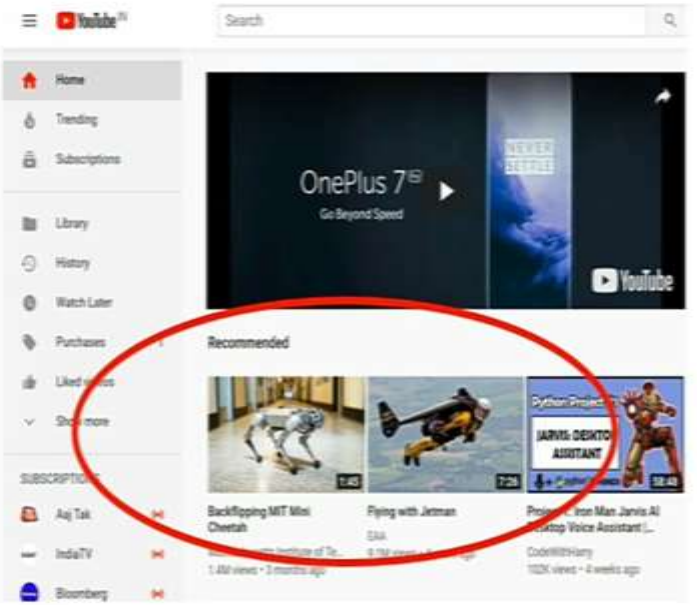
https://youtu.be/lyRPyRKHO8M

Recommendation System



Amazon Recommendation

Recommendation System



Youtube Recommendation

BASICS OF MACHINE LEARNING (ML)

- "A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T** as measured by **P**, improves with experience **E**."

------- Machine Learning, Tom Mitchell, 1997

Example 1: image classification

Task: determine if the image is indoor or outdoor
Performance measure: probability of misclassification

Indoor

Experience/Data:
images with labels

outdoor

BASICS OF MACHINE LEARNING (ML)

- Example 1: **image classification**
- A few terminologies
  - **Training data:** the images given for learning
  - **Test data:** the images to be classified
  - **Binary classification:** classify into two classes
  - **Multi-class Classification:** ?

Example 1: image classification

Task: determine if the image is indoor or outdoor
Performance measure: probability of misclassification

Indoor

Experience/Data:
images with labels

outdoor

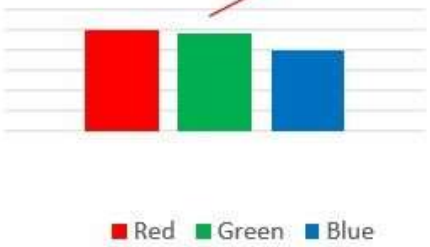
BASICS OF MACHINE LEARNING (ML)

- Example 2: **clustering images**
- A few terminologies
  - Unlabeled data vs labeled data
  - Supervised learning vs unsupervised learning

BASICS OF MACHINE LEARNING (ML)

Math formulation

Extract features



Feature vector: x_i

Label: y_i

Indoor → 0

BASICS OF MACHINE LEARNING (ML)

Math formulation

- Given training data $\{(x_i, y_i): 1 \leq i \leq n\}$
- Find $y = f(x)$ using training data
- s.t. f correct on test data

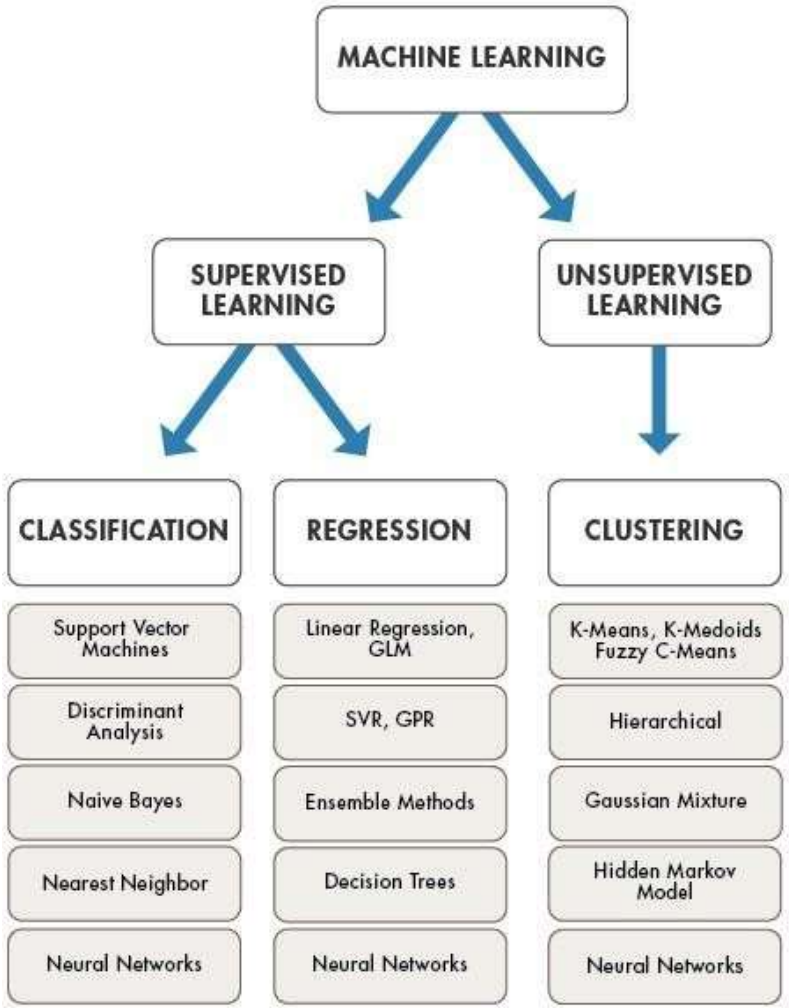
What kind of functions?

BASICS OF MACHINE LEARNING (ML)

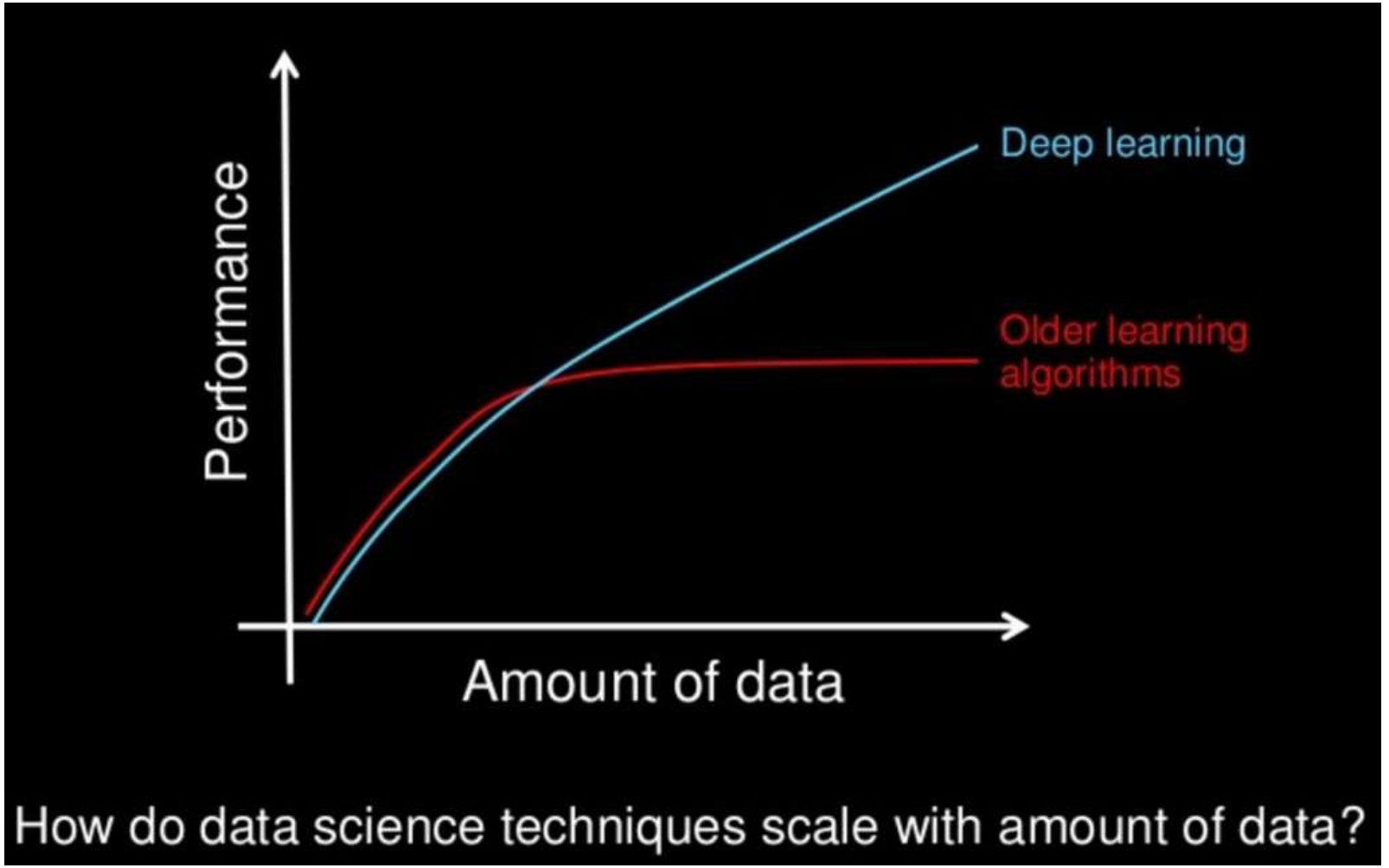
Basic steps in ML

1. Collect data and extract features
2. Build model
3. Optimization: minimize the error

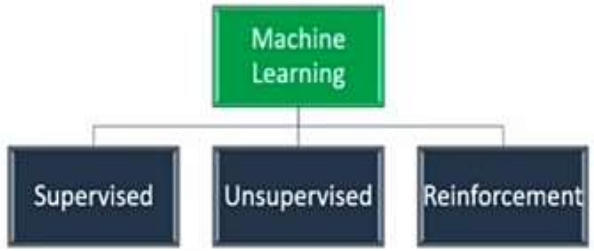
POPULAR ML APPROACHES



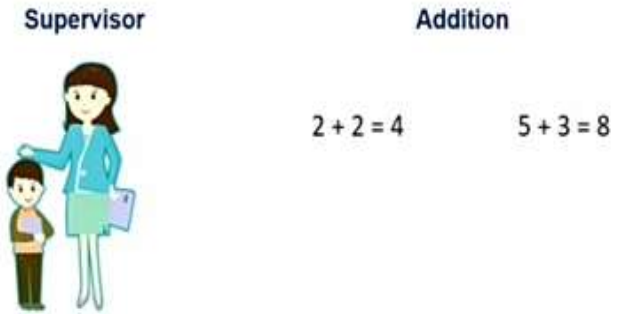
LIMITATIONS OF TRADITIONAL ML



Types of Machine Learning in Artificial Intelligence

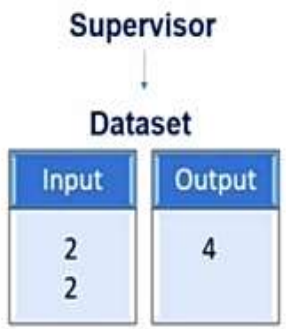


Supervised Learning

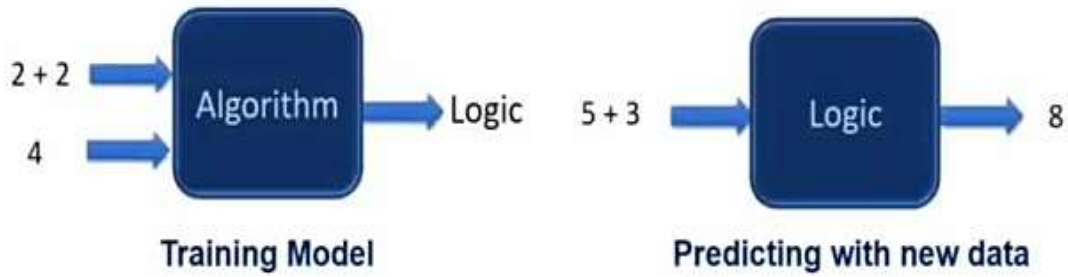


Supervised Learning

Supervised learning is when the model is getting trained on a labelled dataset.



Supervised Learning



Supervised Learning

1. Regression
   -Continuous Value

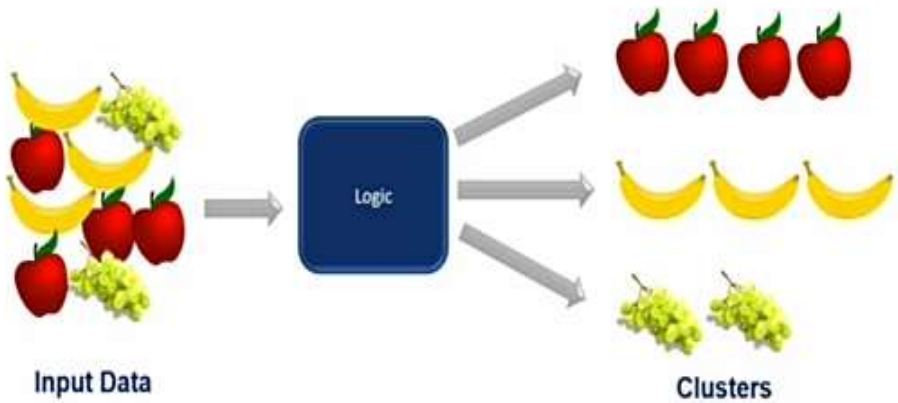
2. Classification
   -Categorical Value

Supervised Learning

1. Regression
   - -Linear Regression
   - -Multiple Linear Regression
   - -Polynomial Regressoin
   - -Support Vector Regression
   - -Decision Tree Regression
   - -Random Forest Regression

2. Classification
   - -Logistic Regression
   - -K-Nearest Neighbors (KNN)
   - -Support Vector Machine (SVM)
   - -Naïve Bayes
   - -Decision Tree Classification
   - -Random Forest Classification

Unsupervised Learning



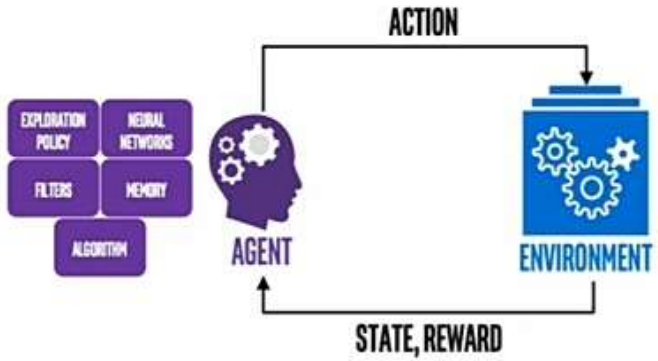
Unsupervised Learning

1. Clustering
   - -K-Means Clustering
   - -Hierarchical Clustering

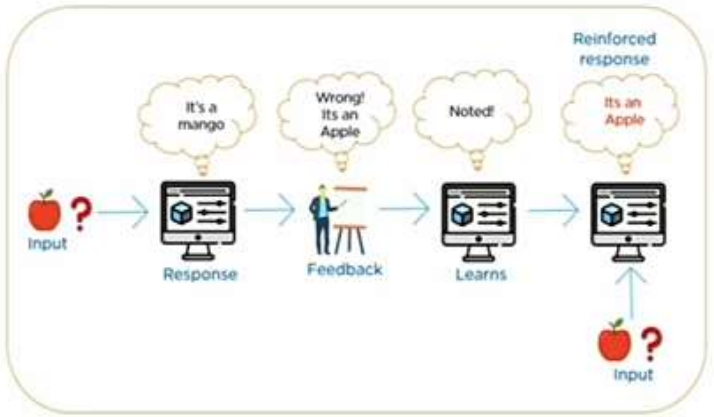
2. Association
   - -Apriori
   - -Eclat

Reinforcement Learning

Reinforcement learning is a type of machine learning where an agent learns to behave in a environment by performing actions and seeing the results.

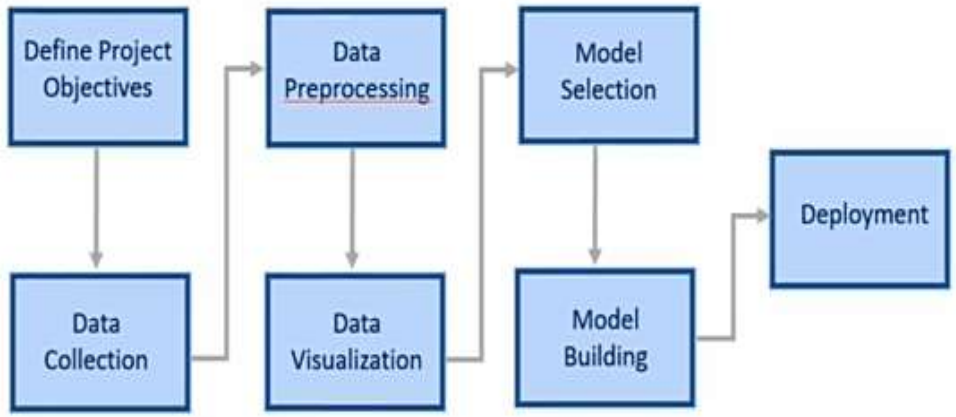


Reinforcement Learning



Life Cycle of Machine Learning

Machine Learning Life Cycle



Define Project Objectives

- Specify Business Problem
- Requirements

Data Collection

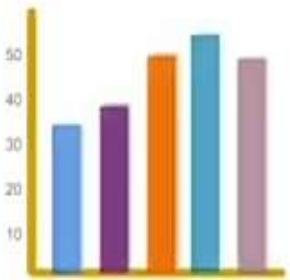
- Primary Data – Collected by researcher from first-hand source.
- Secondary Data – Collected by someone else and already been passed through the statistical process.

Data Preprocessing

1. Data Cleaning :
   - -Filling Missing Data
   - -Smoothing Noisy Data
2. Data Transformation
   - -Normalization
3. Dimensionality Reduction

Data Visualization

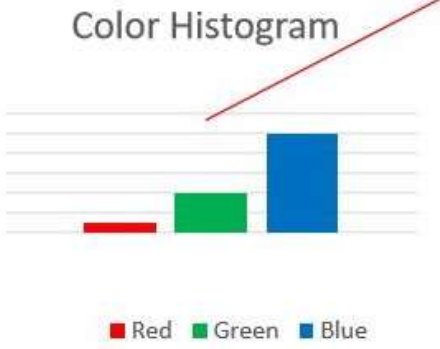
Graphical representation of data.



BASICS OF MACHINE LEARNING (ML)

Math formulation

Extract features



Feature vector: x_j

outdoor

1

Label: y_j

What is Data Preprocessing?

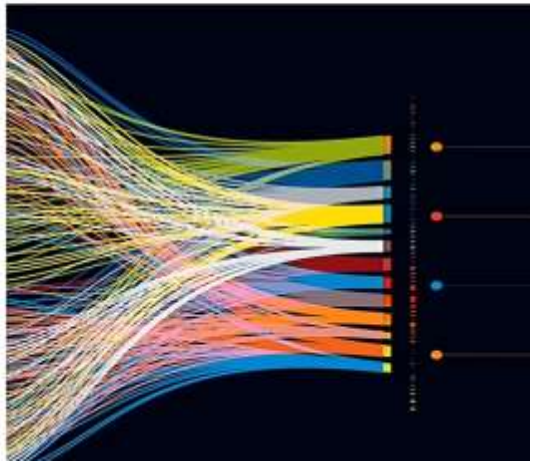
- Data
  - Text
  - Image
  - Video
  - Audio
- Data Preprocessing is a process to convert raw data into meaningful data using different techniques.

Data Preprocessing

- Major steps in Data Preprocessing
  - Data Cleaning
  - Data Integration
  - Data Reduction
  - Data Transformation
  - Data Discretization

Data Collection / Free Data Repositories

- **Kaggle.com**
- UCI Machine Learning Repository
- Awesome Public Datasets
- Data is Plural
- Data World
- Google Data Set Search
- Makeover Monday
- r/datasets/
- United States Government

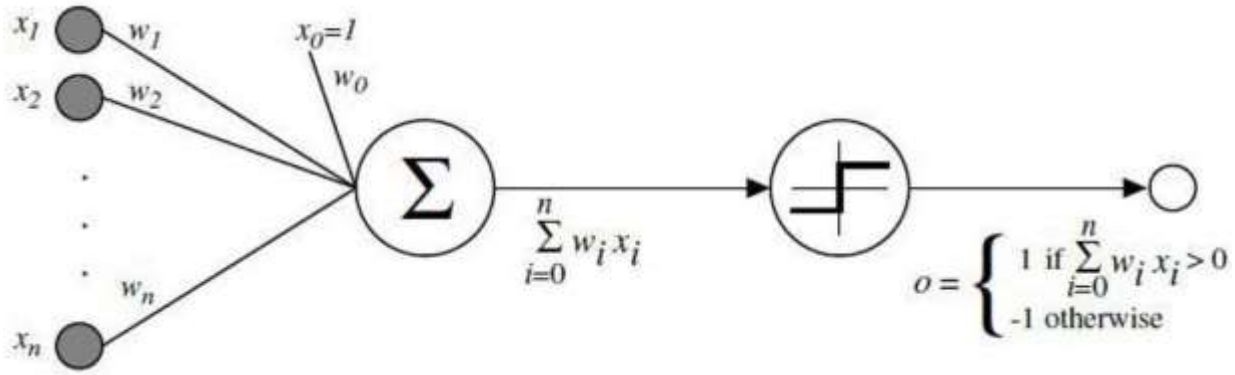


BRIEF HISTORY OF NN & DL

- First appearance (roughly)

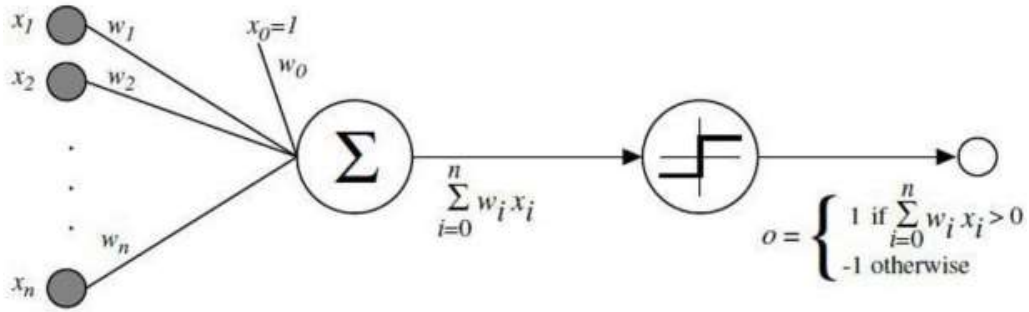
BRIEF HISTORY OF NN & DL

- **Perceptrons**
  - Rosenblatt proposed a machine for binary classifications
  - Main idea
    - One weight $w\,i$ per input $x\,i$
    - Multiply weights with respective inputs and add bias $x\,0$ =+1
    - If result larger than threshold return 1, otherwise 0



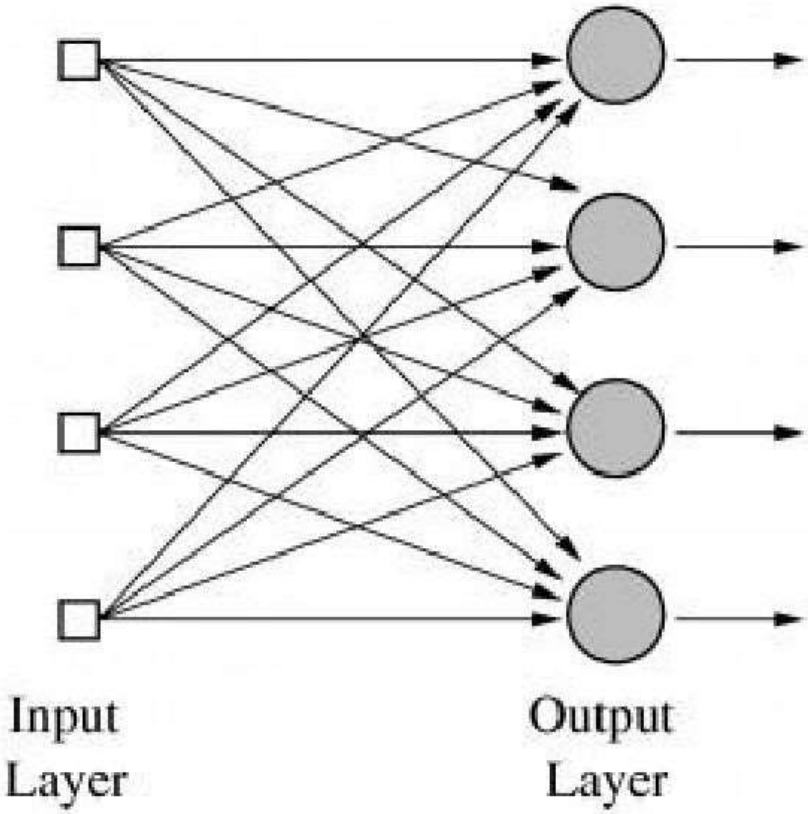
BRIEF HISTORY OF NN & DL

- Training a perceptron
- Rosenblatt's innovation was mainly the learning algorithm for perceptrons
  - Learning algorithm
    - Initialize weights randomly
    - Take one sample xi and predict yi
    - For erroneous predictions update weights
      - Repeat until no errors are made
        - If the output was $\hat{y}_i$ = 0 and yi = 1, increase weights
        - If the output was $\hat{y}_i$ = 1 and yi = 0, decrease weights



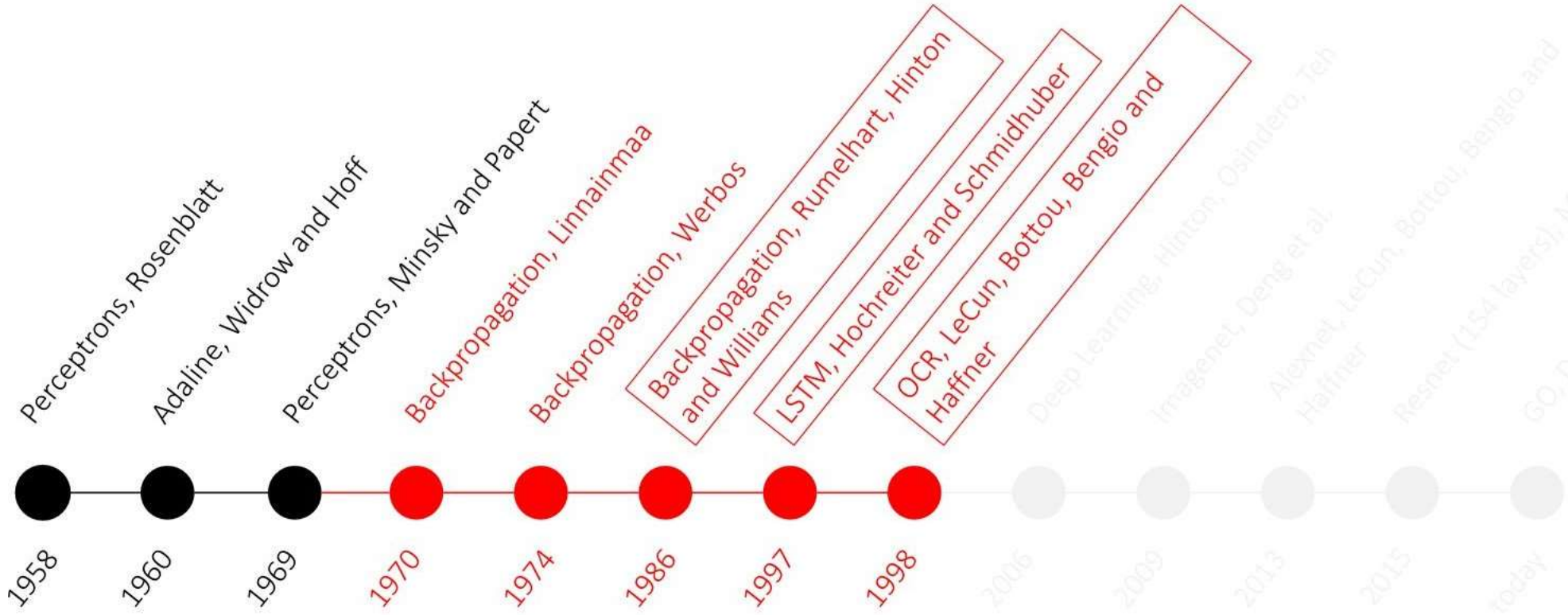
BRIEF HISTORY OF NN & DL

- **From a perceptron to a neural network**
- One perceptron = one decision
- What about multiple decisions?
  - E.g. digit classification
- Stack as many outputs as the possible outcomes into a layer
  - Neural network
- Use one layer as input to the next layer
  - Multi-layer perceptron (MLP)



BRIEF HISTORY OF NN & DL

- **The "AI winter" despite notable successes**



BRIEF HISTORY OF NN & DL

- **Backpropagation**
- Learning multi-layer Perceptrons now possible
  - More complicated functions can be solved
- Efficient algorithm
  - Process hundreds of example without a sweat
  - Allowed for complicated neural network architectures
- Backpropagation still is the backbone of neural network training today
- Digit recognition in cheques (OCR) solved before the 2000

BRIEF HISTORY OF NN & DL

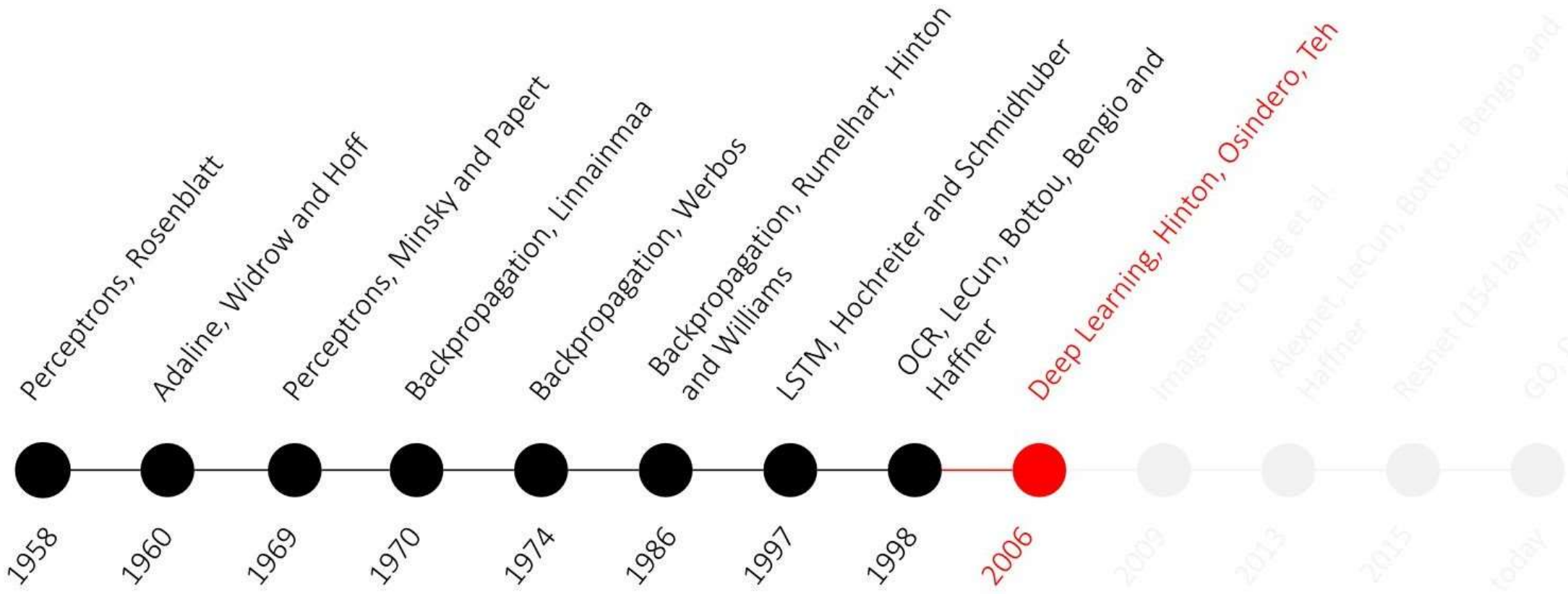
- **Recurrent networks**
- Traditional networks are “too plain”
  - Static Input □Processing □ Static Output
- What about dynamic input
  - Temporal data, Language, Sequences
- Memory is needed to “remember” state changes
  - Recurrent feedback connections
- What kind of memory
  - Long, Short?
  - Both! Long-short term memory networks (LSTM), Schmidhuber 1997

BRIEF HISTORY OF NN & DL

- **The second "AI winter"**
- Until 1998 some nice algorithms and methods were proposed
  - Backpropagation
  - Recurrent Long-Short Term Memory Networks
  - OCR with Convolutional Neural Networks
- However, at the same time Kernel Machines (SVM etc.) suddenly become very popular
  - Similar accuracies in the same tasks
  - Neural networks could not improve beyond a few layers
  - Kernel Machines included much fewer heuristics & nice proofs on generalization
- As a result, once again the AI community turns away from Neural Networks

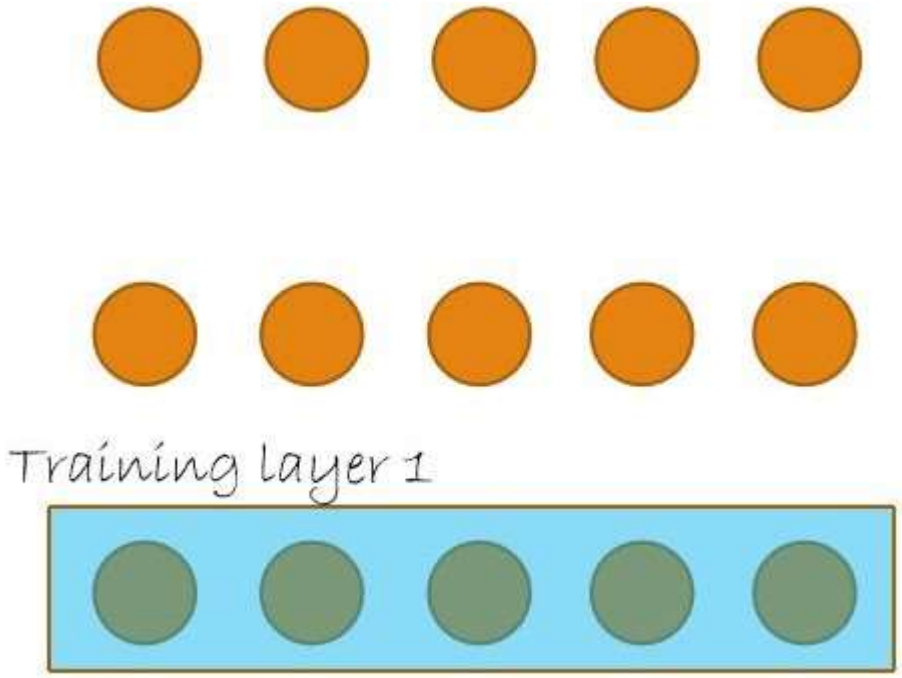
BRIEF HISTORY OF NN & DL

- **The thaw of the "AI winter"**



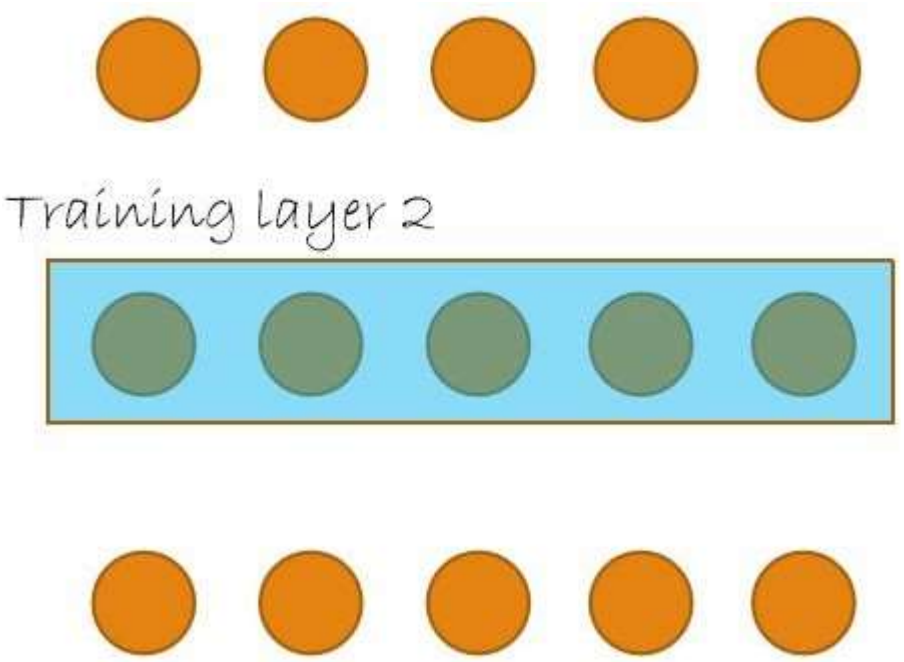
BRIEF HISTORY OF NN & DL

- **Deep Learning arrives**
- Layer-by-layer training
  - The training of each layer individually is an easier undertaking
- Training multi-layered neural networks became easier
- Per-layer trained parameters initialize further training using contrastive divergence



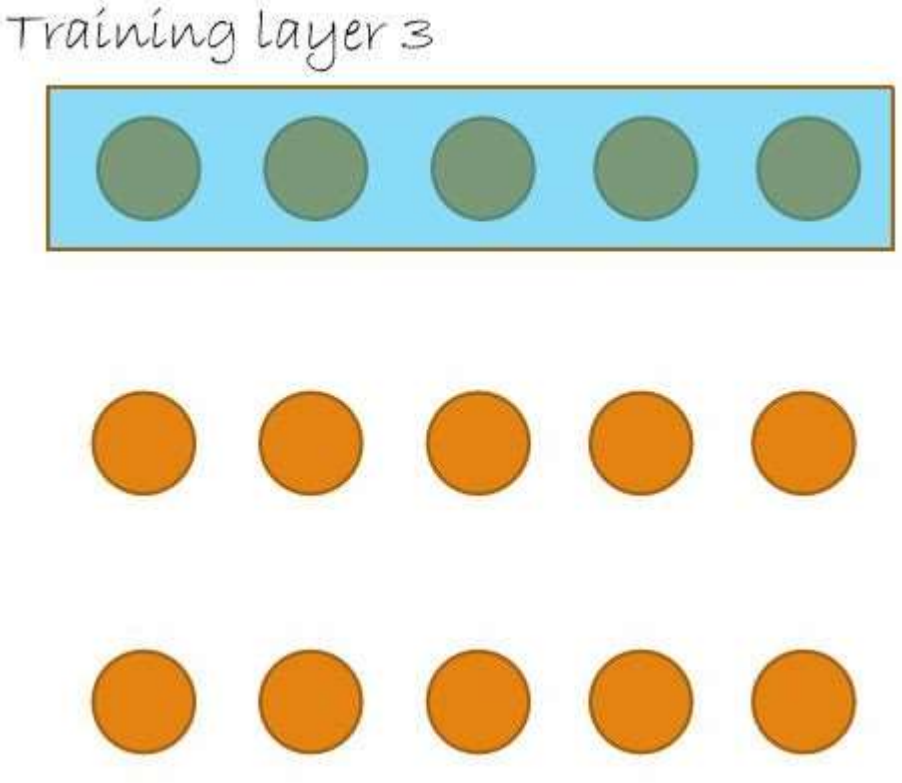
BRIEF HISTORY OF NN & DL

- **Deep Learning arrives**
- Layer-by-layer training
  - The training of each layer individually is an easier undertaking
- Training multi-layered neural networks became easier
- Per-layer trained parameters initialize further training using contrastive divergence



BRIEF HISTORY OF NN & DL

- **Deep Learning arrives**
- Layer-by-layer training
  - The training of each layer individually is an easier undertaking
- Training multi-layered neural networks became easier
- Per-layer trained parameters initialize further training using contrastive divergence

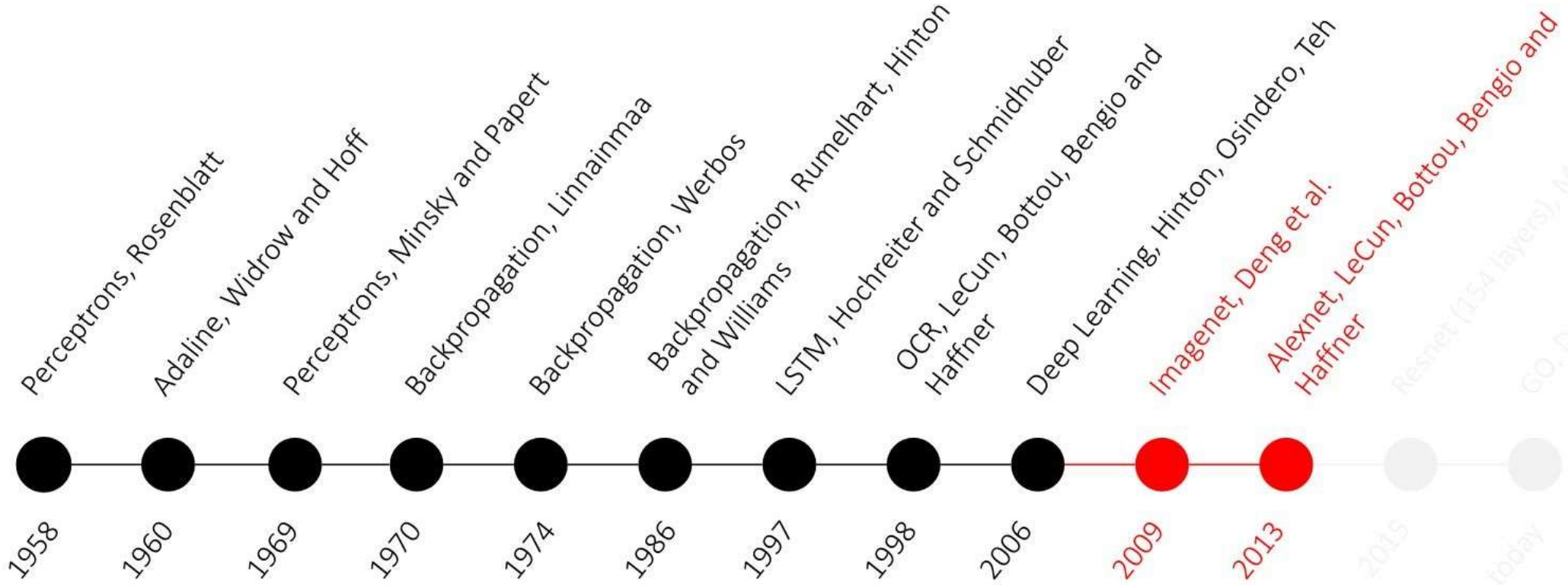


BRIEF HISTORY OF NN & DL

- **Neural Network and Deep Learning problems**
- Lack of processing power
  - No GPUs at the time
- Lack of data
  - No big, annotated datasets at the time
- Overfitting
  - Because of the above, models could not generalize all that well
- Vanishing gradient
  - While learning with NN, you need to multiply several numbers $a1 \cdot a2 \cdot \cdots a n$.
  - If all are equal to 0.1, for n = 10 the result is 0.0000000001, too small for any learning

BRIEF HISTORY OF NN & DL

- **Deep Learning Renaissance**



BRIEF HISTORY OF NN & DL

- **More data, more …**
- In 2009 the Imagenet dataset was published [Deng et al., 2009]
- Collected images for each term of Wordnet (100,000 classes)
- Tree of concepts organized hierarchically
  - "Ambulance", "Dalmatian dog", "Egyptian cat", …
  - About 16 million images annotated by humans
- Imagenet Large Scale Visual Recognition Challenge (ILSVRC)
  - 1 million images
  - 1,000 classes
  - Top-5 and top-1 error measured

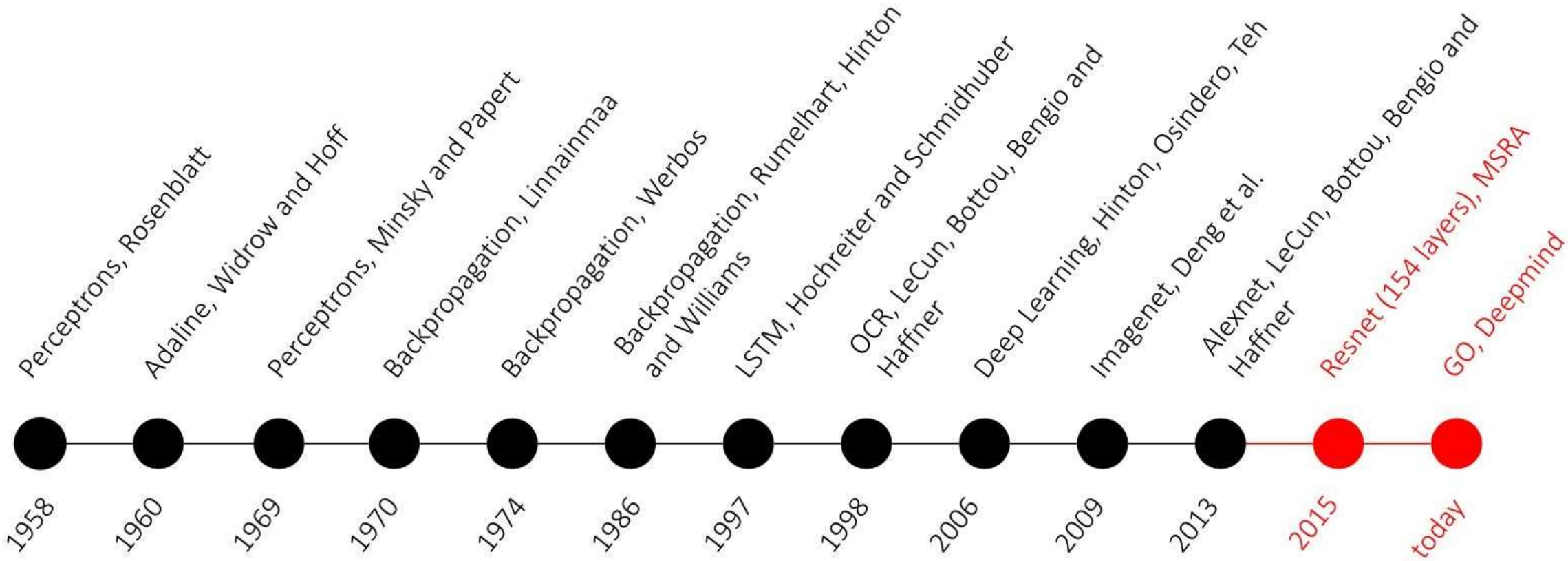
2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

BRIEF HISTORY OF NN & DL

- **Alexnet**
- In 2013 Krizhevsky, Sutskever and Hinton re-implemented [Krizhevsky 2013] a convolutional neural network [LeCun1998]
- Trained on Imagenet, Two GPUs were used for the implementation
- Further theoretical improvements
  - Rectified Linear Units (ReLU) instead of sigmoid or tanh
  - Dropout
  - Data augmentation
- In the 2013 Imagenet Workshop a legendary turmoil
  - Blasted competitors by an impressive 16% top-5 error, Second best around 26%
  - Most didn’t even think of NN as remotely competitive
- At the same time similar results in the speech recognition community
  - One of G. Hinton students collaboration with Microsoft Research, improving state-of-the-art by an impressive amount after years of incremental improvements [Hinton 2012]

BRIEF HISTORY OF NN & DL

- **Deep Learning Golden Era**



BRIEF HISTORY OF NN & DL

- **The today**
- Deep Learning is almost everywhere
  - Object classification
  - Object detection, segmentation, pose estimation
  - Image captioning, question answering
  - Machine translation
  - Speech recognition
  - Robotics
- Some strongholds
  - Action classification, action detection
  - Object retrieval
  - Object tracking

BRIEF HISTORY OF NN & DL

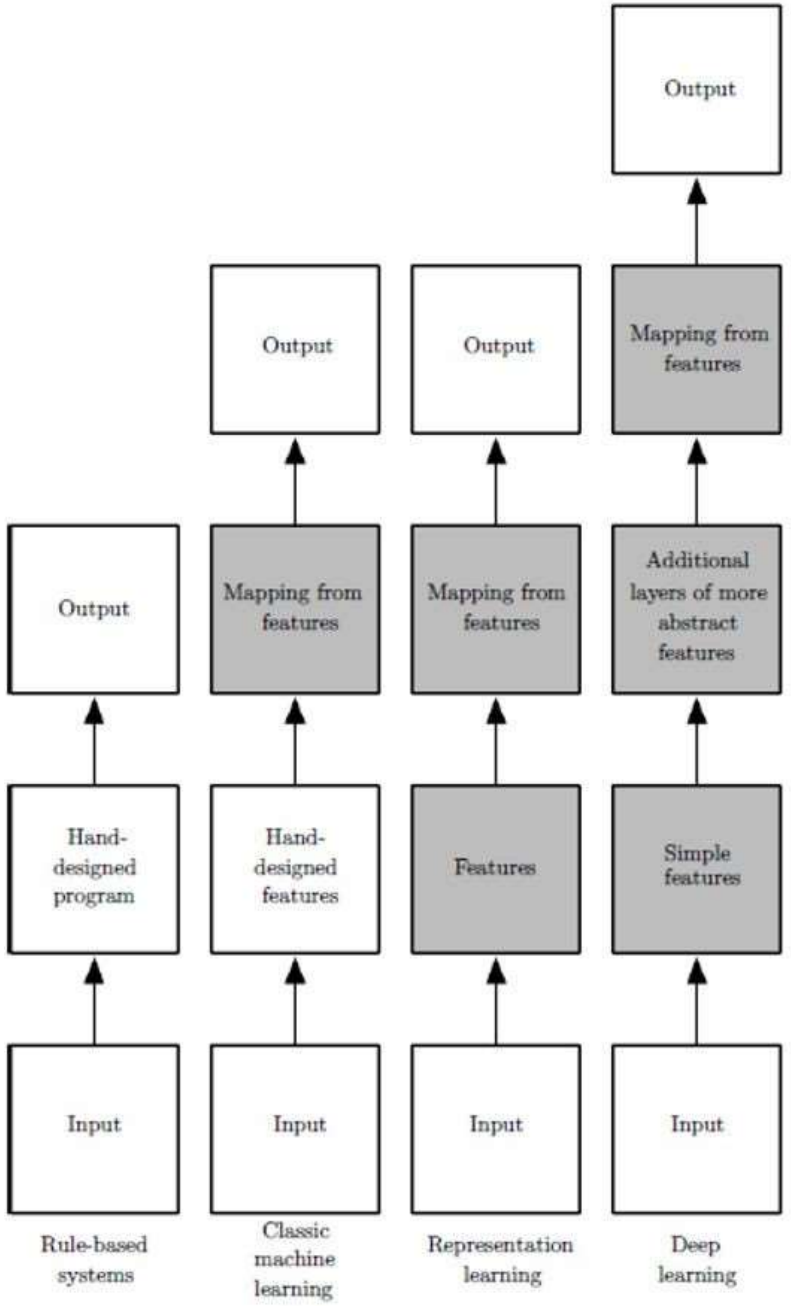
- **Now, what has changed??**
- Better hardware (GPUs)
- Bigger data (Big Data)
- Better regularization methods, such as dropout (Avoid Overfitting)
- Better optimization methods, such as Adam, batch normalization (Better means to train large networks)

DEEP LEARNING

- What is DL?
  - DL is a sub-field of ML which attempts to learn high-level abstractions in data by utilizing hierarchical architectures.
  - Using a neural network with several layers of nodes between the input and output
  - The series of layers between input & output do **feature identification** and processing in a series of stages, just as our brains seem to.
- The workflow of DL has been developed with inspiration from the human brain.
- Why hierarchical?

DEEP LEARNING

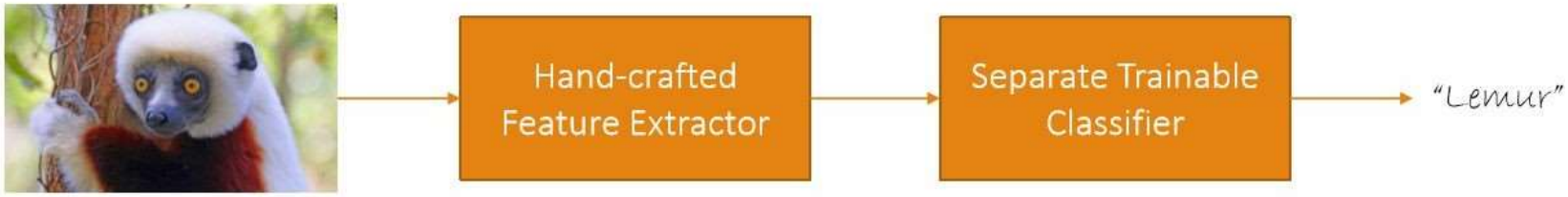
- **Shallow Learning vs Deep Learning**



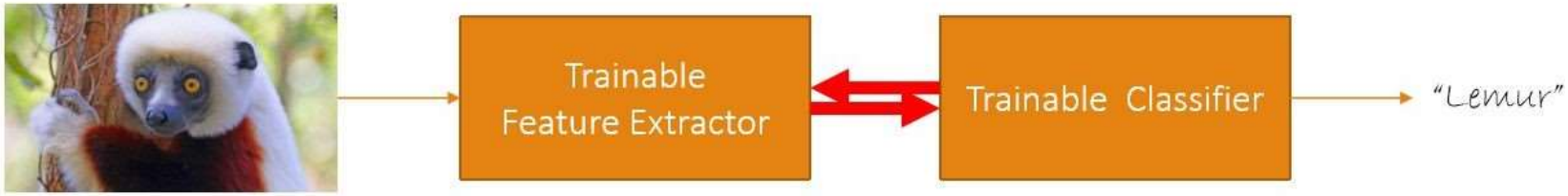
DEEP LEARNING

- **Shallow Learning vs Deep Learning**

- Traditional pattern recognition

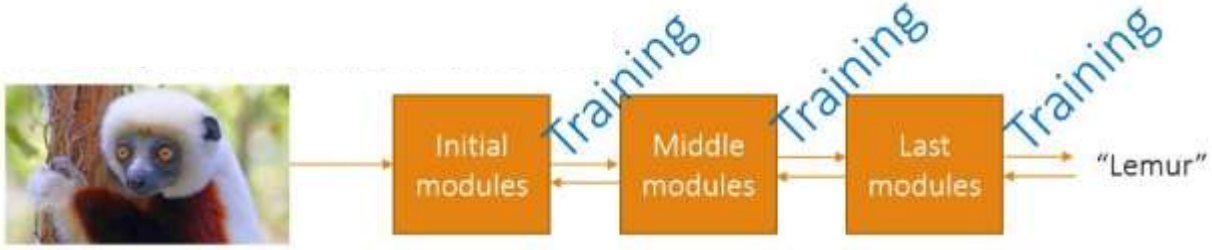


- End-to-end learning → Features are also learned from data



DEEP LEARNING

- **End-to-end learning of feature hierarchies**
- A pipeline of successive modules
- Each module's output is the input for the next module
- Modules produce features of higher and higher abstractions
  - Initial modules capture low-level features (e.g. edges or corners)
  - Middle modules capture mid-level features (e.g. circles, squares, textures)
  - Last modules capture high level, class specific features (e.g. face detector)
- Preferably, input as raw as possible
  - Pixels for computer vision, words for NLP



DEEP LEARNING

- **Why learn the features?**
- Manually designed features
  - Often take a lot of time to come up with and implement
  - Often take a lot of time to validate
  - Often they are incomplete, as one cannot know if they are optimal for the task
- Learned features
  - Are easy to adapt
  - Very compact and specific to the task at hand
  - Given a basic architecture in mind, it is relatively easy and fast to optimize
- Time spent for designing features now spent for designing architectures

DEEP LEARNING

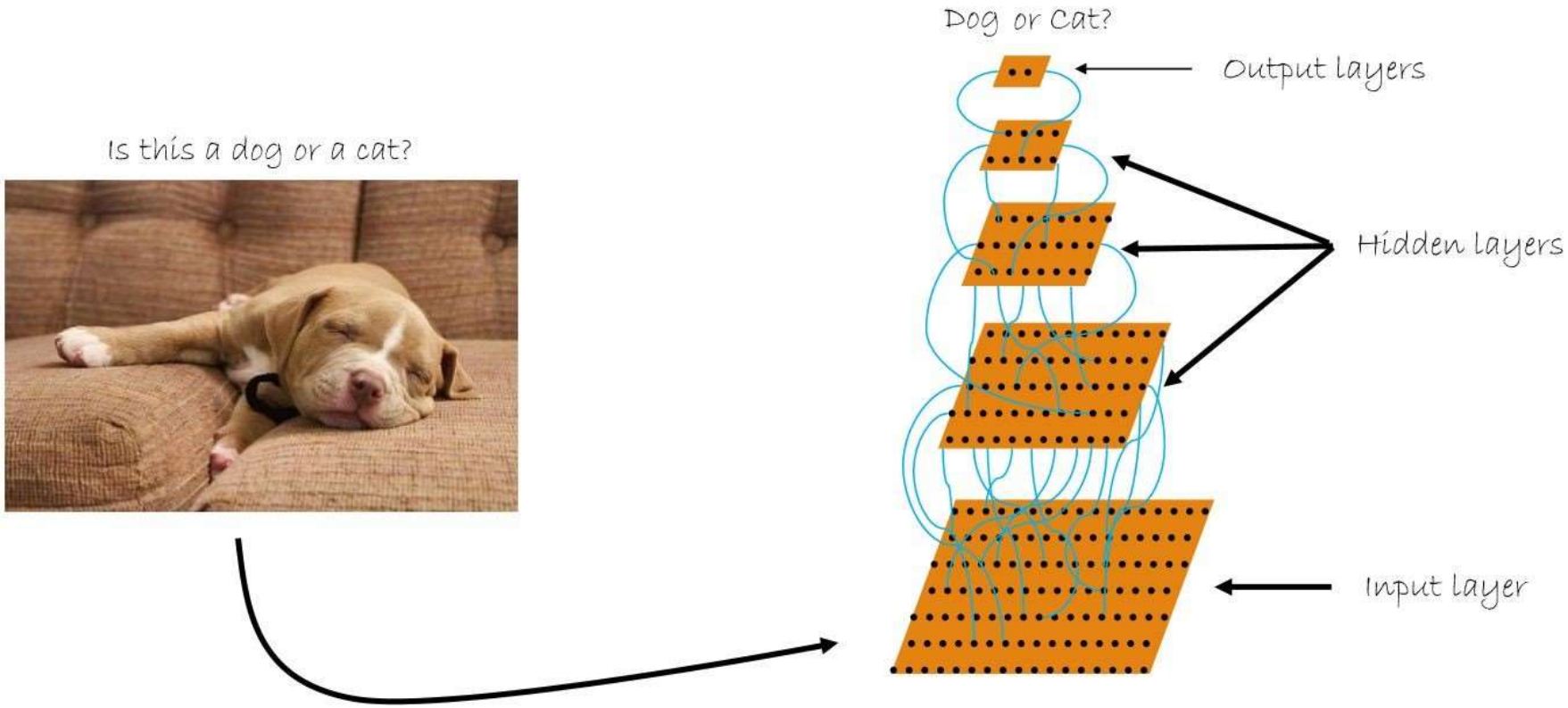
- **Types of learning**
- Supervised learning
  - (Convolutional) neural networks
- Unsupervised learning
  - Autoencoders, layer-by-layer training
- Self-supervised learning
  - A mix of supervised and unsupervised learning
- Reinforcement learning
  - Learn from noisy, delayed rewards from your environment
  - Perform actions in your environment, so as to make decisions what data to collect

DEEP LEARNING

- **Deep architectures**
- Feedforward
  - (Convolutional) neural networks
- Feedback
  - Deconvolutional networks
- Bi-directional
  - Deep Boltzmann Machines, stacked autoencoders
- Sequence based
  - RNNs, LSTMs

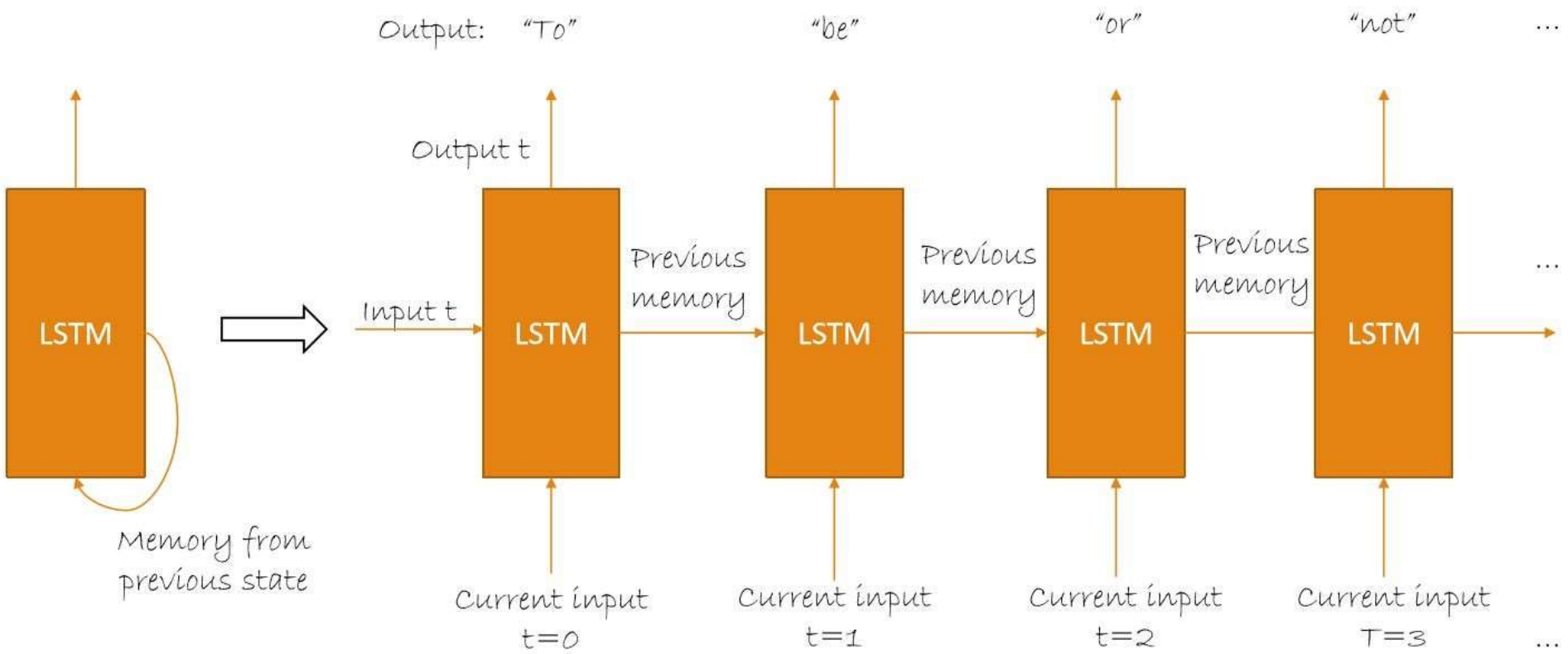
DEEP LEARNING

Convolutional networks in a nutshell



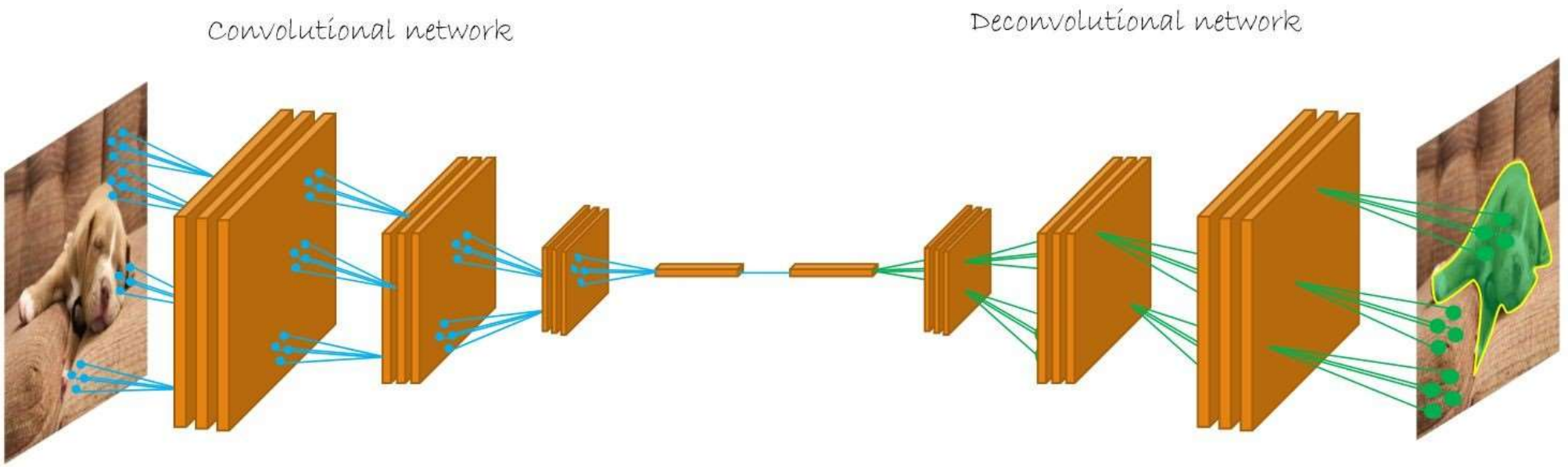
DEEP LEARNING

Recurrent networks in a nutshell



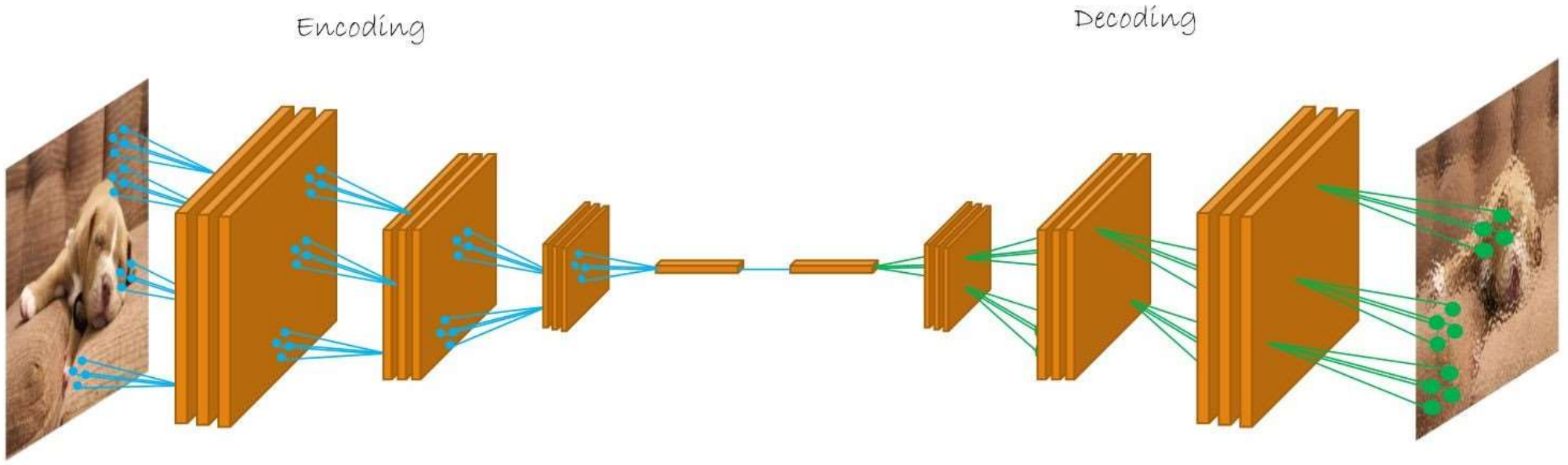
DEEP LEARNING

Deconvolutional networks



DEEP LEARNING

Autoencoders in a nutshell

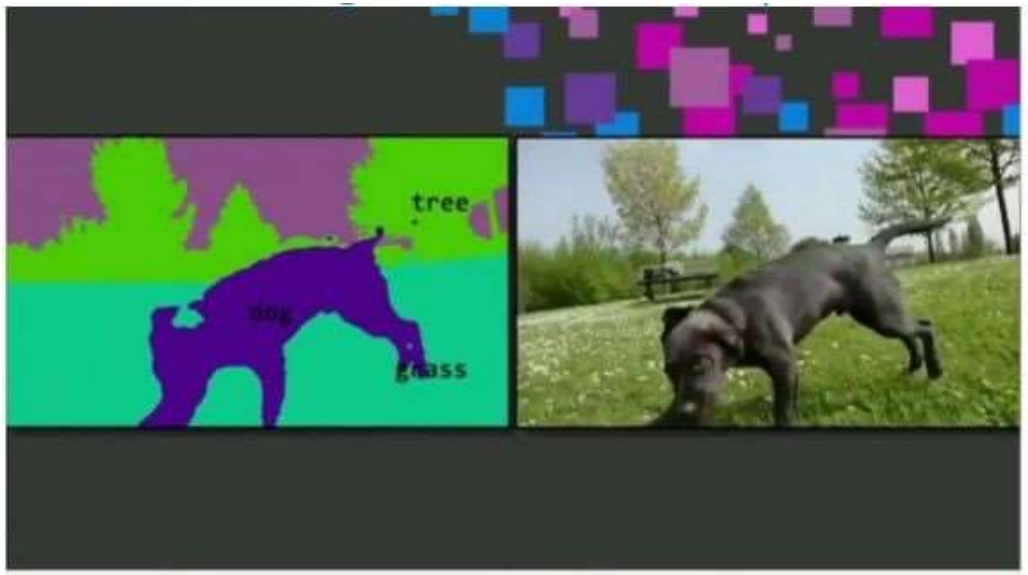


APPLICATIONS

- **Computer Vision**

Large-scale Video Classification with Convolutional Neural Networks, CVPR 2014

Object and activity recognition



Microsoft Deep Learning Semantic Image Segmentation

Object detection and segmentation

APPLICATIONS

- **Computer Vision**

Image Captioning

Visual Question Answering

APPLICATIONS

- **Robotics**

Self Driving Cars

Drones and robots

APPLICATIONS

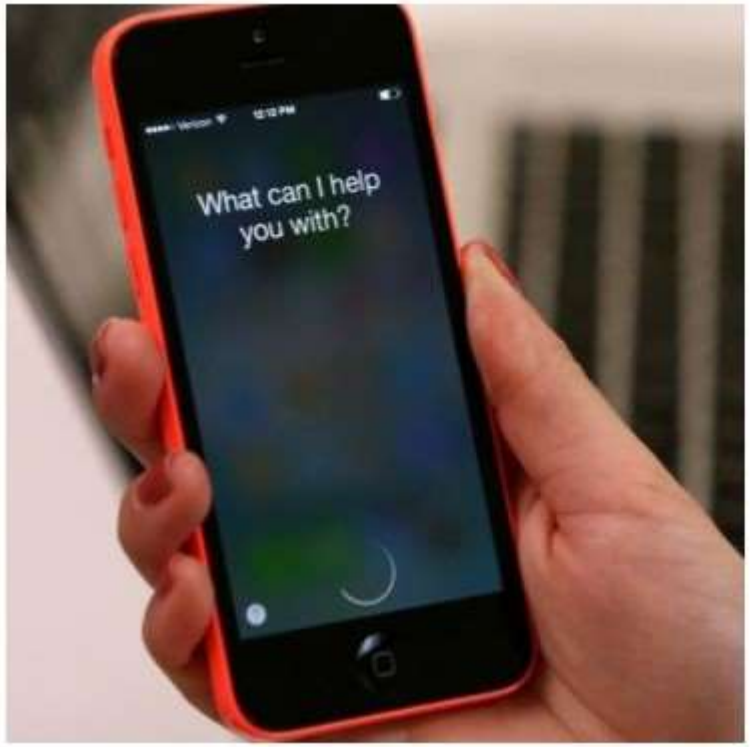
- **Game AI**

Google DeepMind's Deep Q-learning playing Atari Breakout

Learn to play games

APPLICATIONS

- **NLP and Speech**



Speech recognition and Machine translation

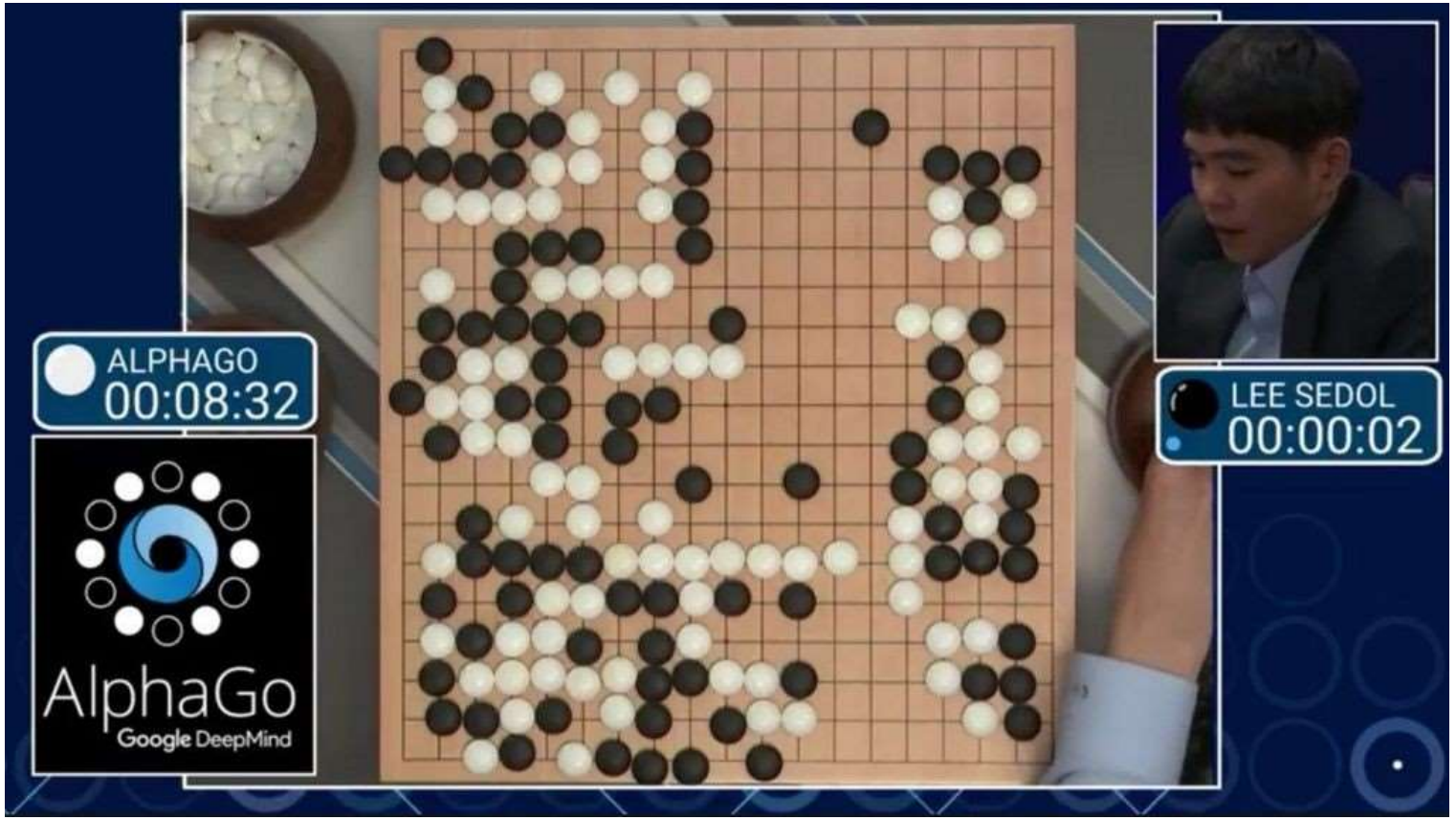
APPLICATIONS

- **Arts**

Imitating famous painters

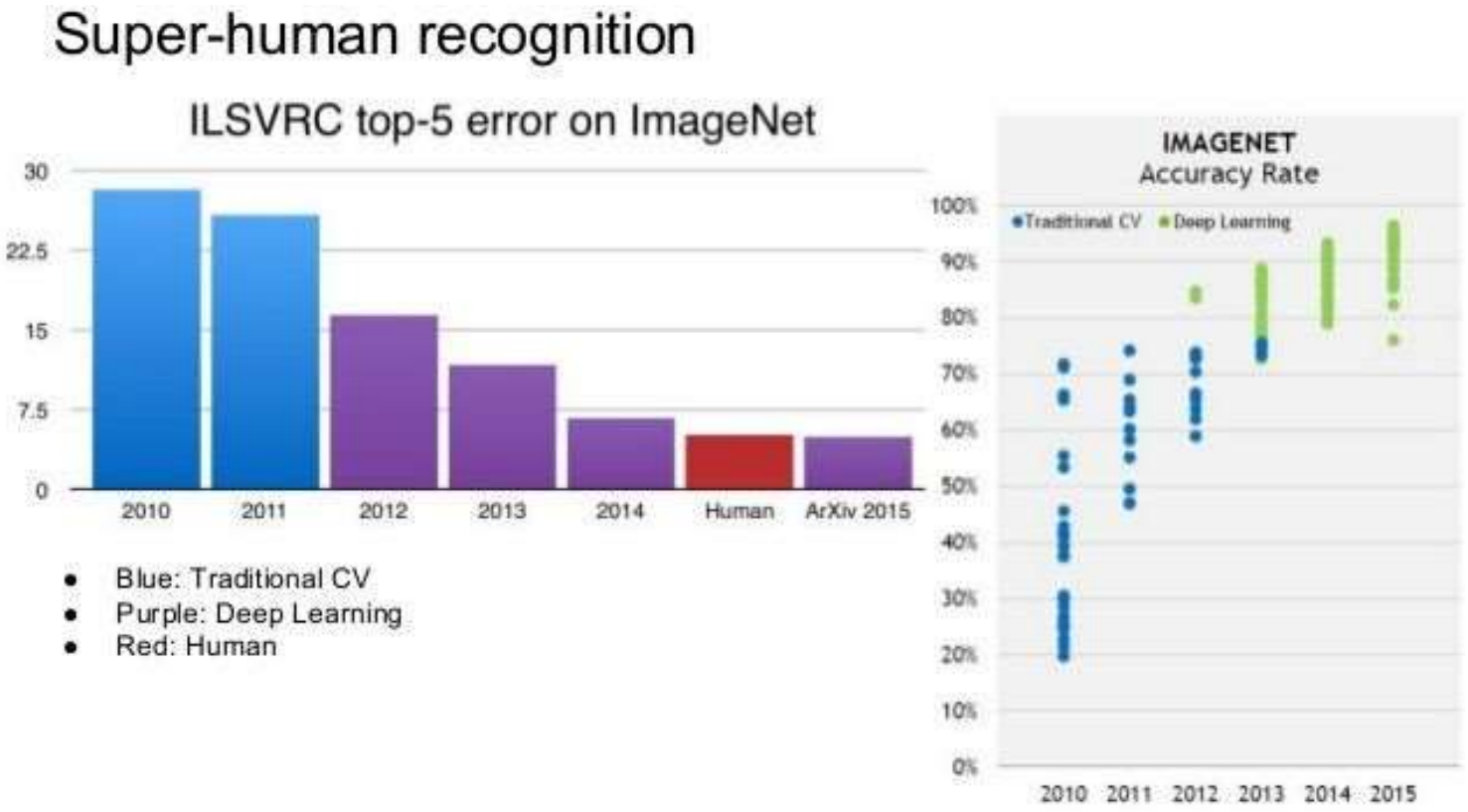
RECENT MILESTONES

- **Alpha Go** defeated the world champion in the game of Go.



RECENT MILESTONES

- Super human performance in **Image Classification**



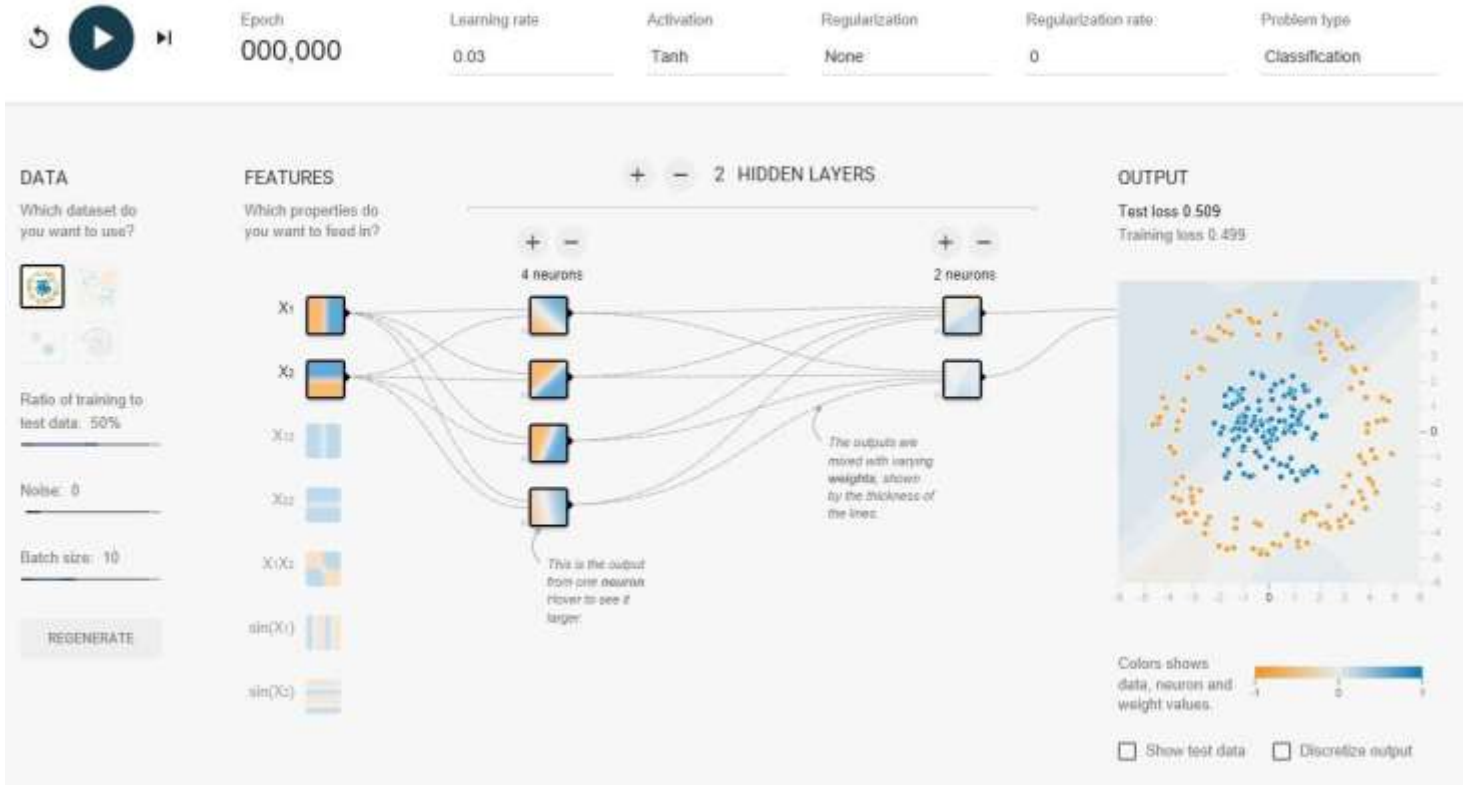
MAJOR CHALLENGES

- DL requires **huge amounts of data** which may not be always available
  - How to effectively learn from small data?
  - Transfer learning
- DL is **computationally expensive**
  - Build & train efficient-high performance models
  - Network Pruning (eliminate unnecessary connections from DNN)

ASSIGNMENTS

1. Play around in the “Neural Network Playground”

   http://playground.tensorflow.org/



ASSIGNMENTS

2. Read Paper: Deep Learning (Nature)

https://www.nature.com/articles/nature14539

THAT'S IT!

Questions ???