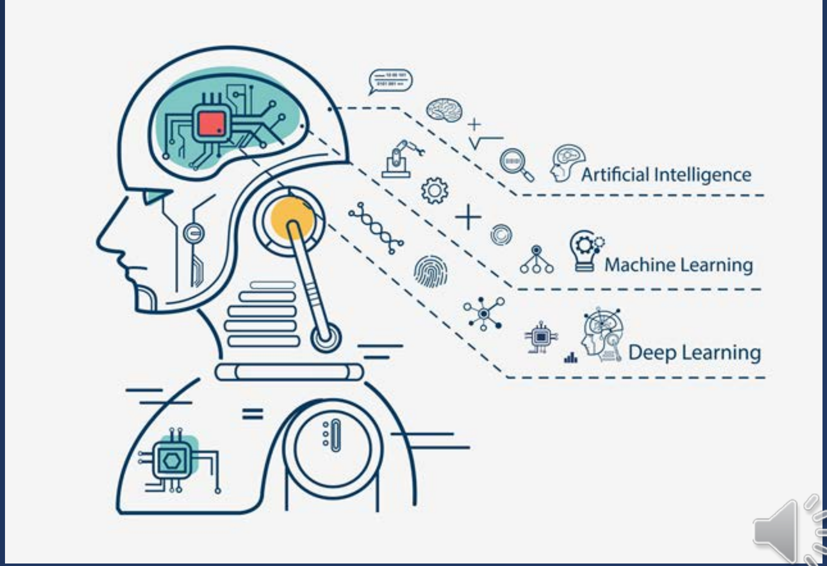
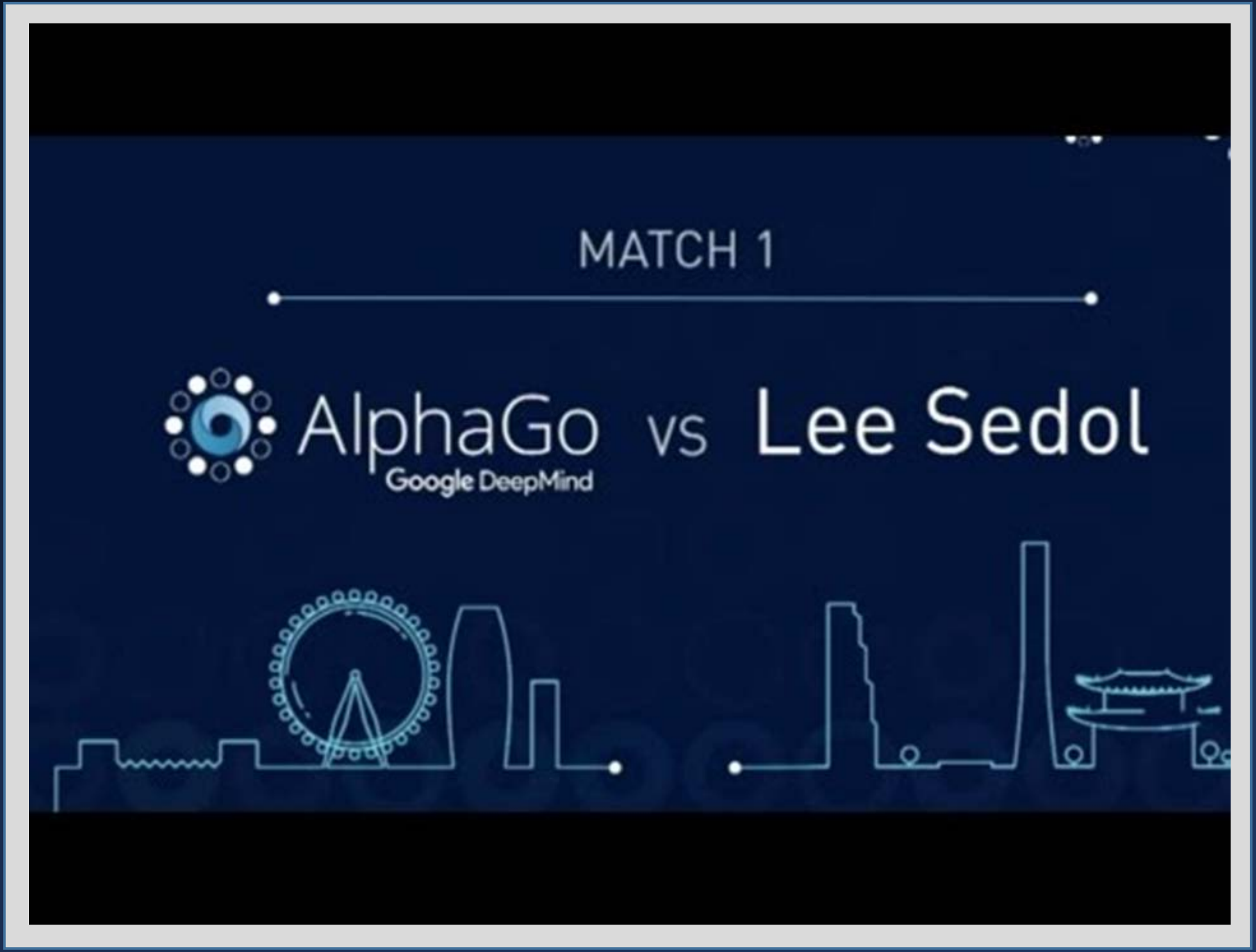
Big Data (MHI222956/MHI225101)

8.2 Deep Learning & K-means

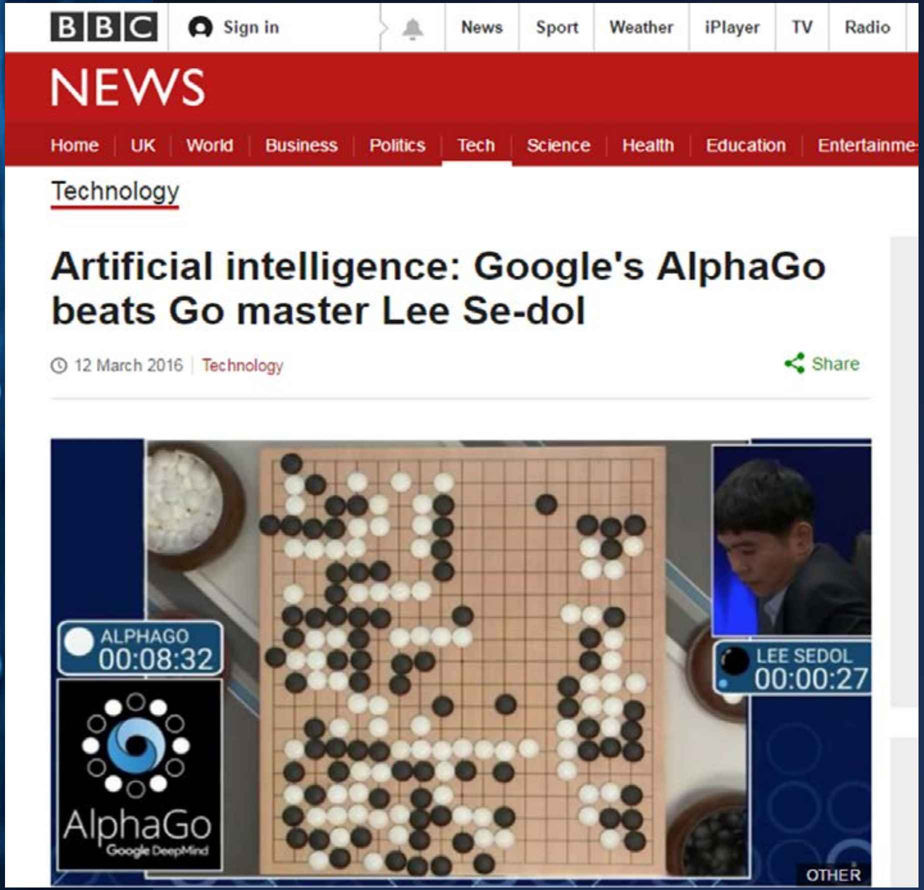
Deep Learning



8 March 2016, a landmark battle…



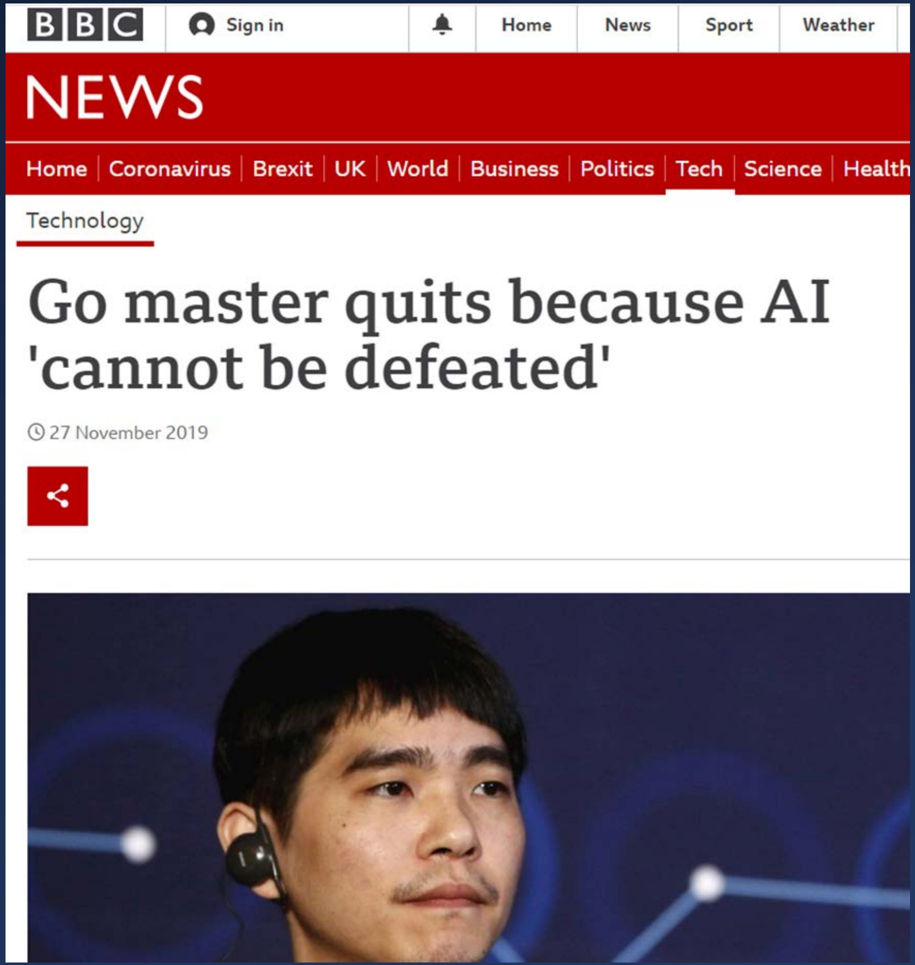
“Even if I become the number one, there is an entity that cannot be defeated”



Technology

Artificial intelligence: Google's AlphaGo beats Go master Lee Se-dol

12 March 2016 | Technology



Technology

Go master quits because AI 'cannot be defeated'

27 November 2019

- “Who” is AlphaGo?

  a computer program developed by Google DeepMind in London

- What is Google DeepMind?
  - an artificial intelligence division within Google that was created after Google bought University College London spinout, DeepMind, for a reported £400 million in January 2014.

- How does it work?
  - … The technology behind DeepMind is complex to say the least but that didn’t stop Suleyman (co-founder of Deepmind Technology) from trying to convey some of the fundamental deep learning principles that underpin it …

- Then…

What is Deep Learning?

Suleyman’s explain

- These are hierarchical based networks initially conceived back in the 80s but recently resuscitated by a bunch of really smart guys from Toronto and New York.

- The basic intuition is that at one end we take the raw pixel data or the raw sensory stream data of things we would like to classify or recognise.

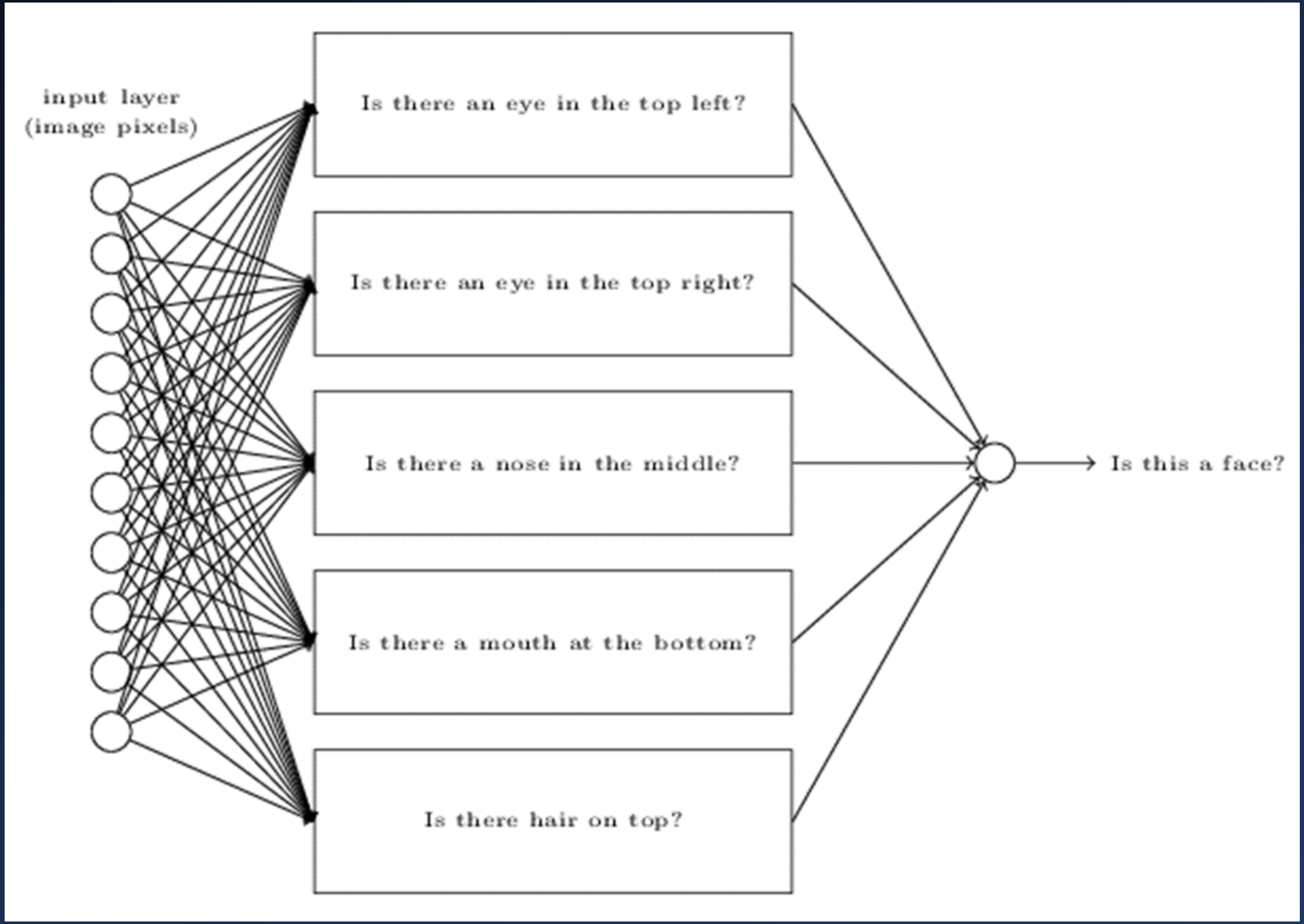
- This seems to be a very effective way of learning to find structure in very large data sets. Right at the very output we’re able to impose on the network some requirement to produce some set of labels or classifications that we recognise and find useful as humans.

From Neural Networks to Deep Learning

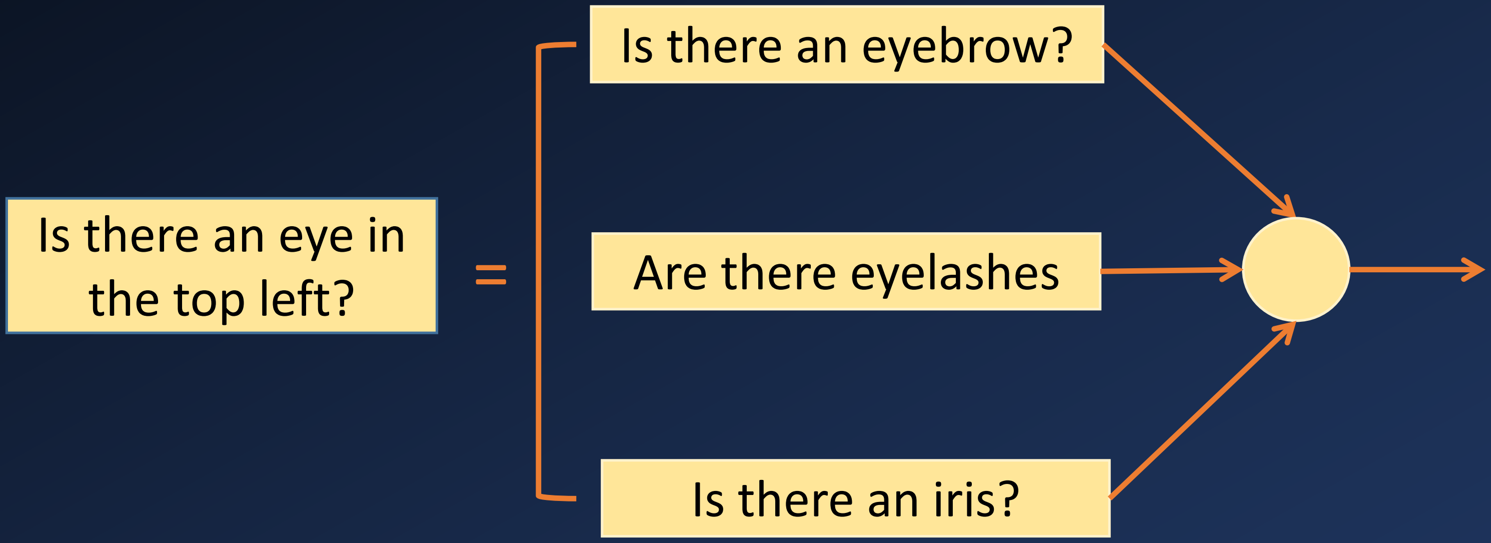
- Weights and biases in Neural Network were discovered automatically.
- No immediately explanation of how the network does what it does!
- Can we think artificial neurons as a means of weighing evidence?

- Suppose we want to determine whether an image shows a human face

Decompose the problem into sub-problem


input layer
(image pixels)
Is there an eye in the top left?
Is there an eye in the top right?
Is there a nose in the middle?
Is there a mouth at the bottom?
Is there hair on top?
Is this a face?

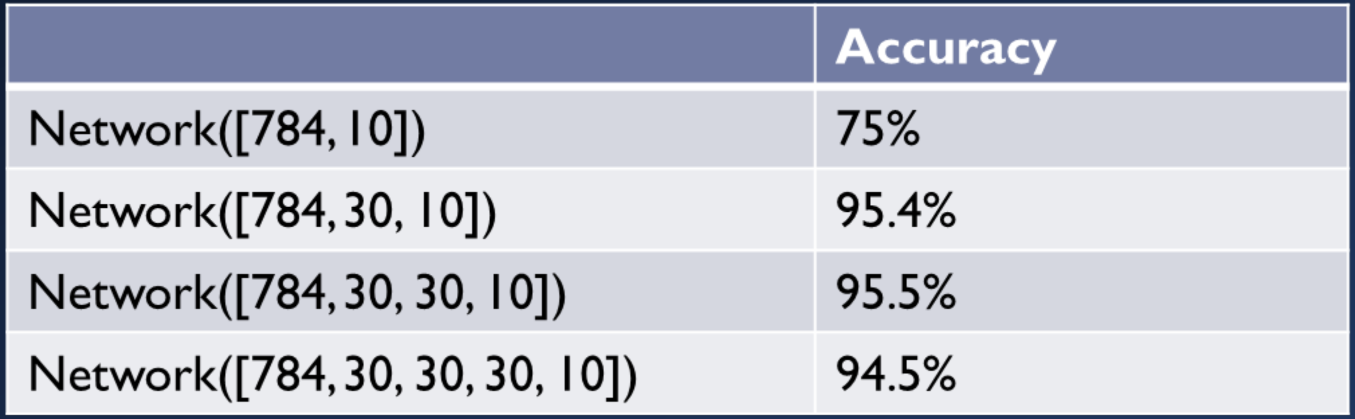
Decompose the problem into sub-problem



Break down the questions further and further through multiple layers… Ultimately, we'll be working with sub-networks that answer questions so simple they can easily be answered at the level of single pixels.

- Deep Neural Networks
  - Networks with two or more hidden layers
- How to do this recursive decomposition?
- How to train this deep neural networks?
  - Deep neural networks are hard to train!
  - Since 2006, a set of techniques has been developed that enable learning in deep neural nets.

- Training of deep neural networks?
  - The MNIST handwritten digits classification



	Accuracy
Network([784, 10])	75%
Network([784, 30, 10])	95.4%
Network([784, 30, 30, 10])	95.5%
Network([784, 30, 30, 30, 10])	94.5%

  - Different layers learning at vastly different speeds

- There's an intrinsic instability associated to learning by gradient descent in deep, many-layer neural networks.
  - Vanishing gradient problem (sigmoid neurons)
  - Exploding gradient problem

- Other obstacles
  - The way weights are initialized
  - Network architecture
  - Hyper-parameters

Deep Convolutional Networks (CNNs)

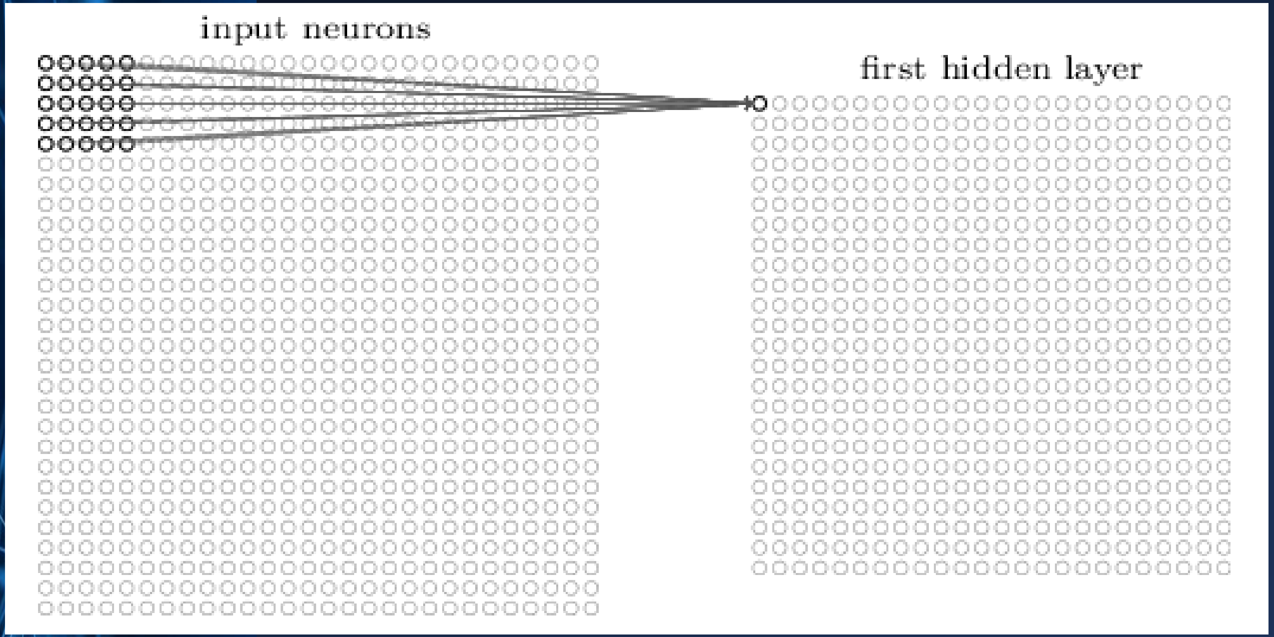
- The most widely used type of deep network
- Convolution
  - An integral that expresses the amount of overlap of one function as it is shifted over another function .

- The MNIST handwritten digits classification
  - High accuracy achieved by classical neural network
  - Fully connected network
  - Did not take into account the spatial structure of the images

Three ideas of convolutional neural networks

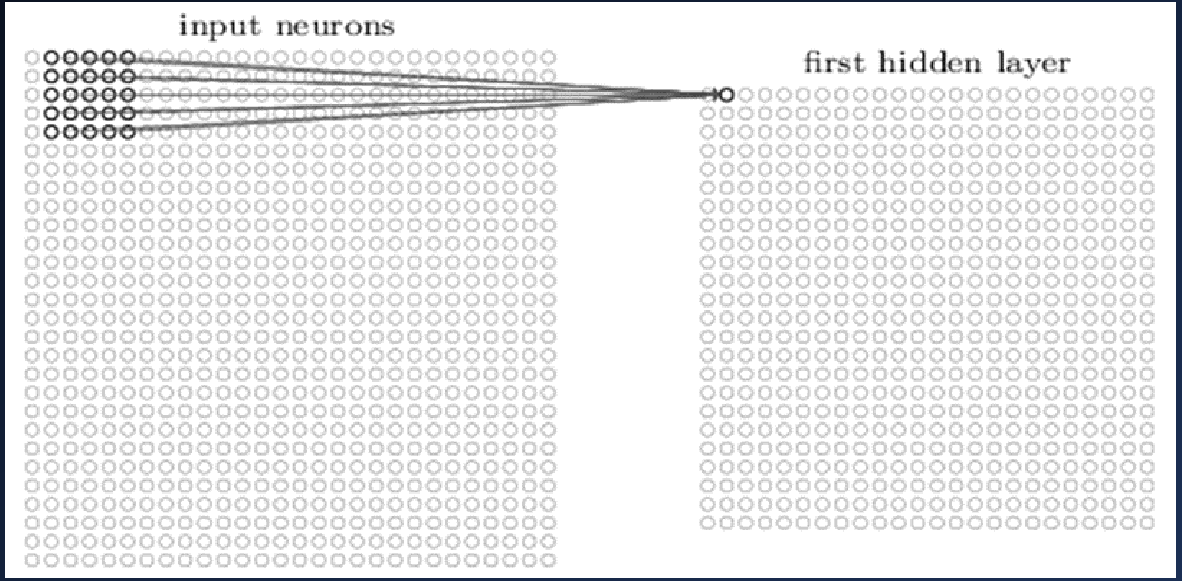
- Local receptive fields
- Shared weights
- pooling

Local receptive fields



- Only make connections in small localized regions (local receptive field) of the input image
- Each connection learns a weight. And the hidden neuron learns an overall bias as well.

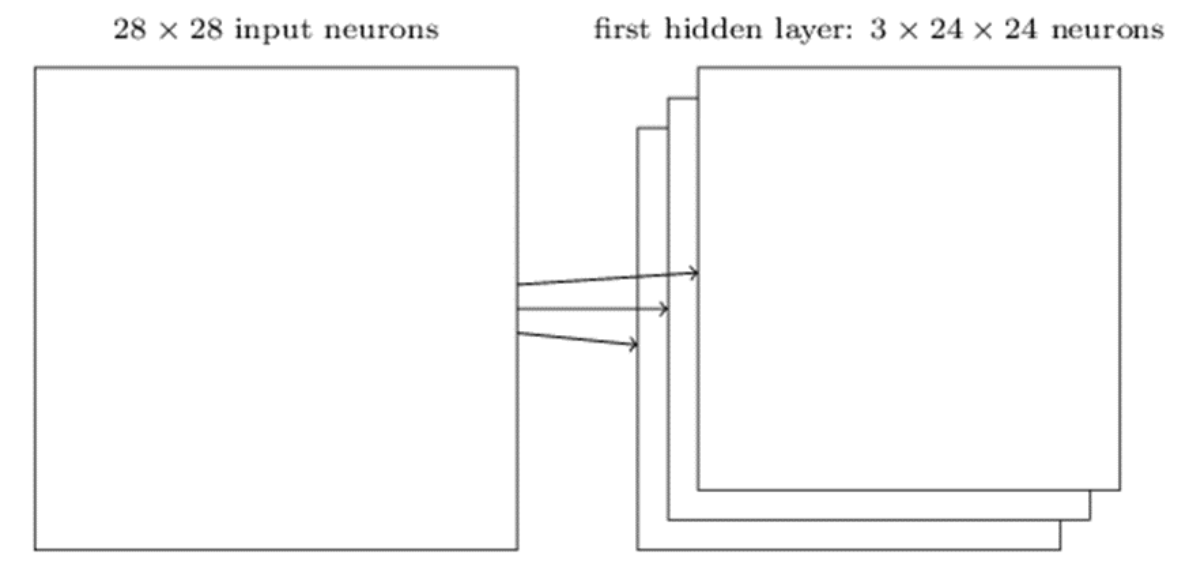
- Slide the local receptive field across the entire input image.



- For a 28×28 input image, and 5×5local receptive fields, then there will be 24×24 neurons in the first hidden layer.
- Same weights and bias for each of the 24×24 hidden neurons
  - all the neurons in the first hidden layer detect exactly the same feature, just at different locations in the input image.
  - greatly reduces the number of parameters involved in a convolutional network.

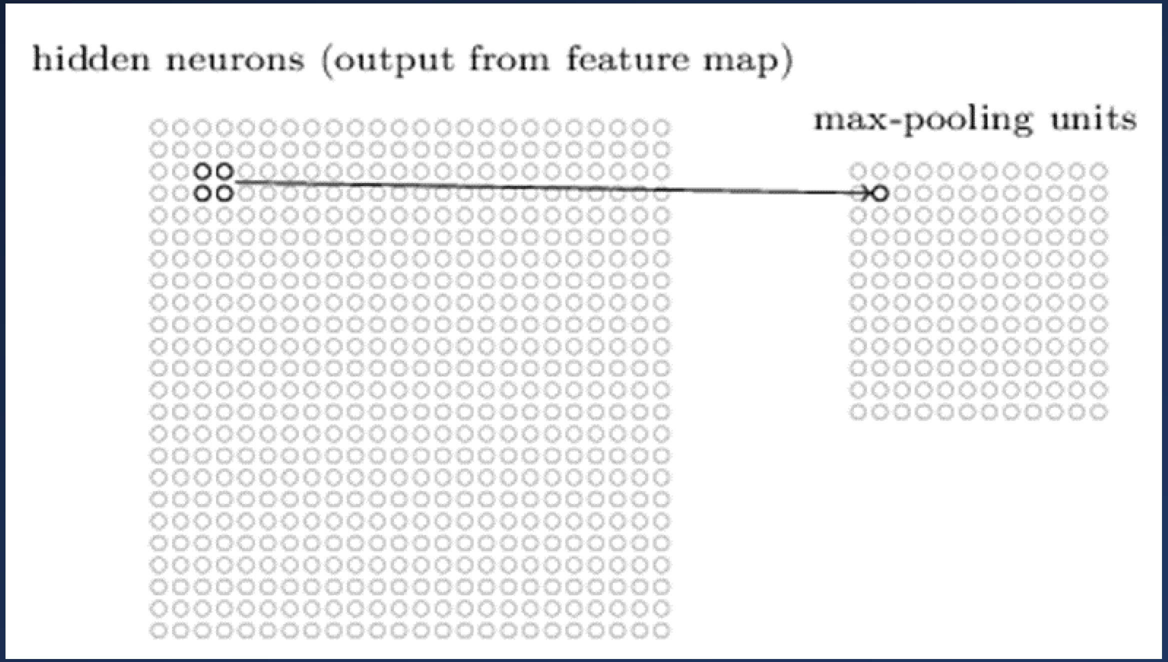
- Some terms:
  - feature map -- map from the input layer to the hidden layer
  - shared weights -- weights defining the feature map
  - shared bias -- bias defining the feature map
  - kernel or filter – defined by the shared weights and bias

- A complete convolutional layer consists of several different feature maps

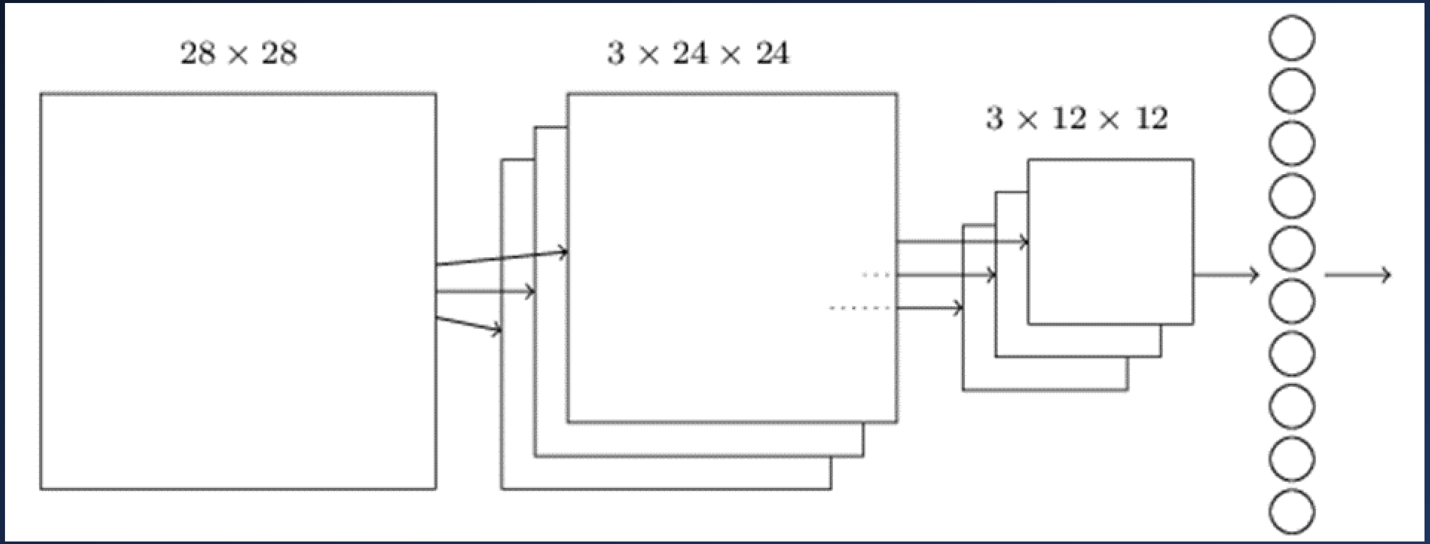


Pooling layers

- Usually used immediately after convolutional layers
- To simplify the information in the output from the convolutional layer



- Put it all together

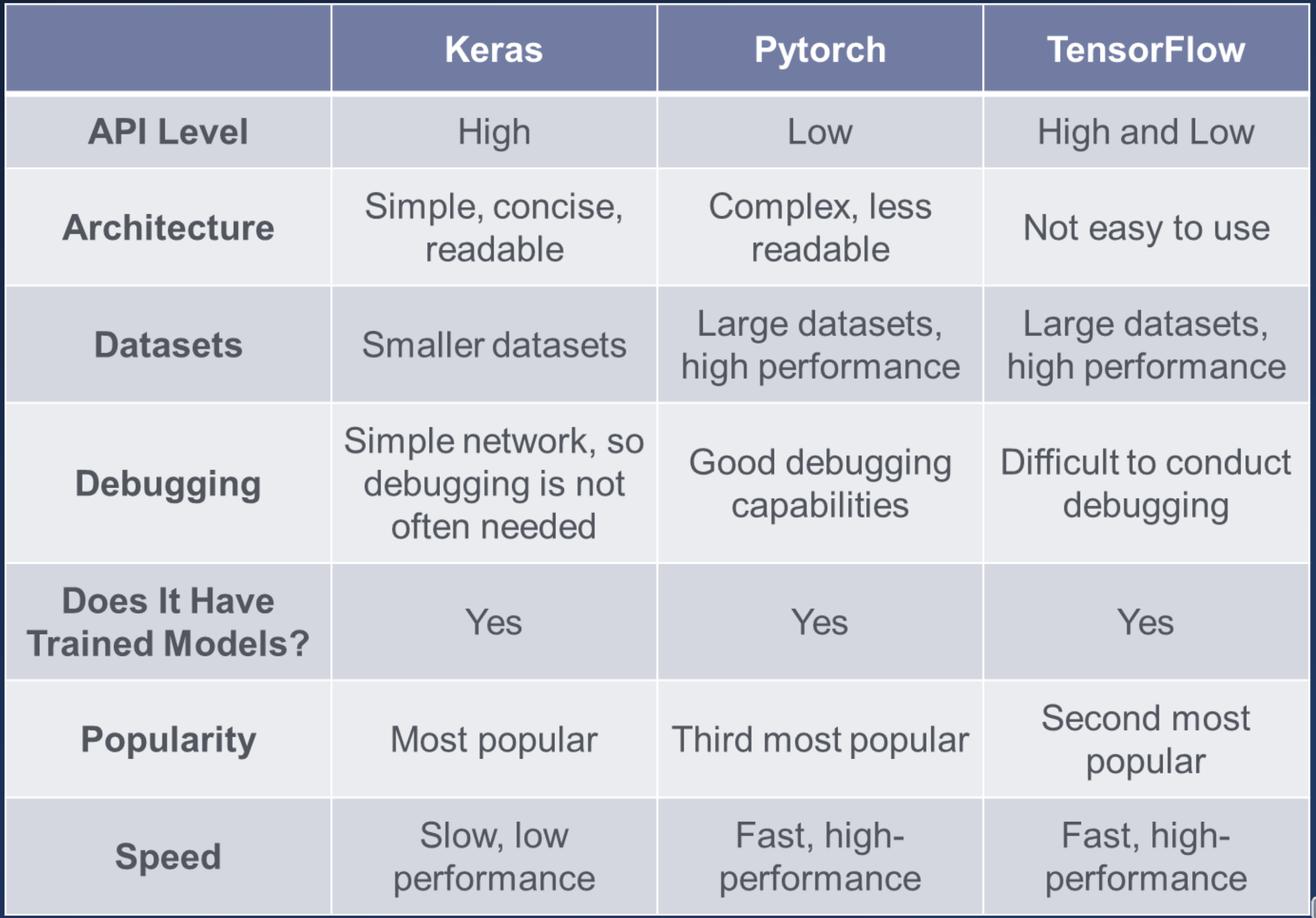


- The final layer of connections in the network is a fully-connected layer.
- Train the network using stochastic gradient descent and backpropagation

Implementation in Python

The most popular Deep Learning frameworks:

- Keras
- Tensorflow
- Pytorch

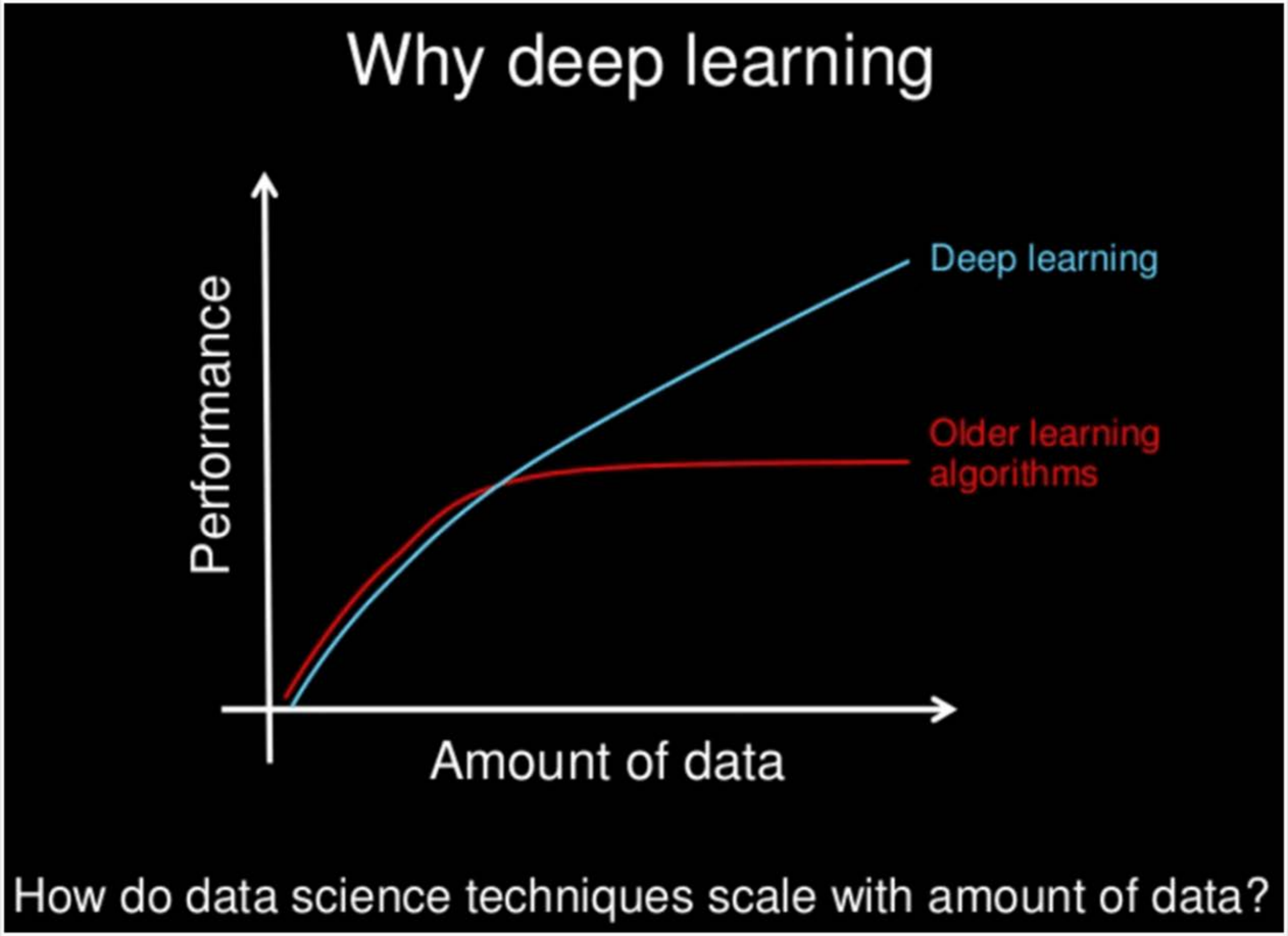


	Keras	Pytorch	TensorFlow
API Level	High	Low	High and Low
Architecture	Simple, concise, readable	Complex, less readable	Not easy to use
Datasets	Smaller datasets	Large datasets, high performance	Large datasets, high performance
Debugging	Simple network, so debugging is not often needed	Good debugging capabilities	Difficult to conduct debugging
Does It Have Trained Models?	Yes	Yes	Yes
Popularity	Most popular	Third most popular	Second most popular
Speed	Slow, low performance	Fast, high-performance	Fast, high-performance

Other approaches to deep neural networks

- Recurrent neural networks (RNN)
- Boltzmann machines
- Generative models
- Transfer learning
- Reinforcement learning
- etc.

- 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.

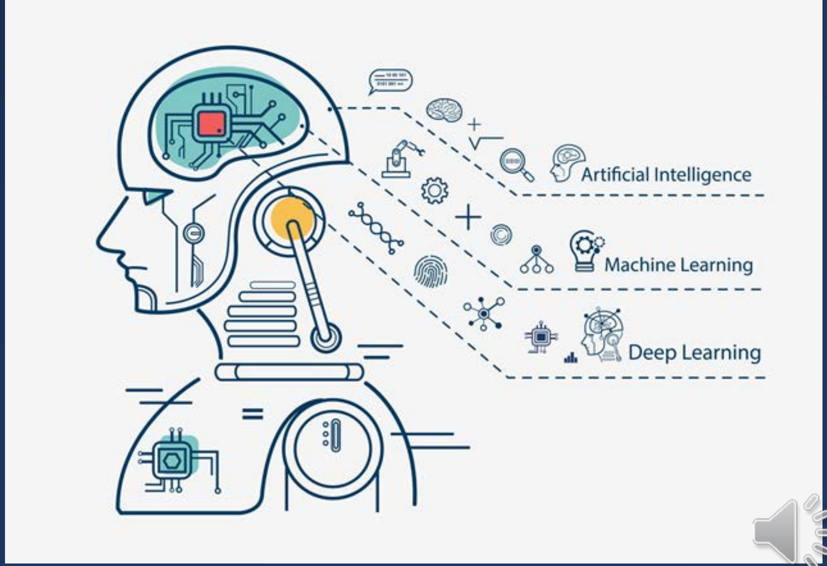


Why deep learning

How do data science techniques scale with amount of data?

Why Deep Learning?
Slide by Andrew Ng, all rights reserved.

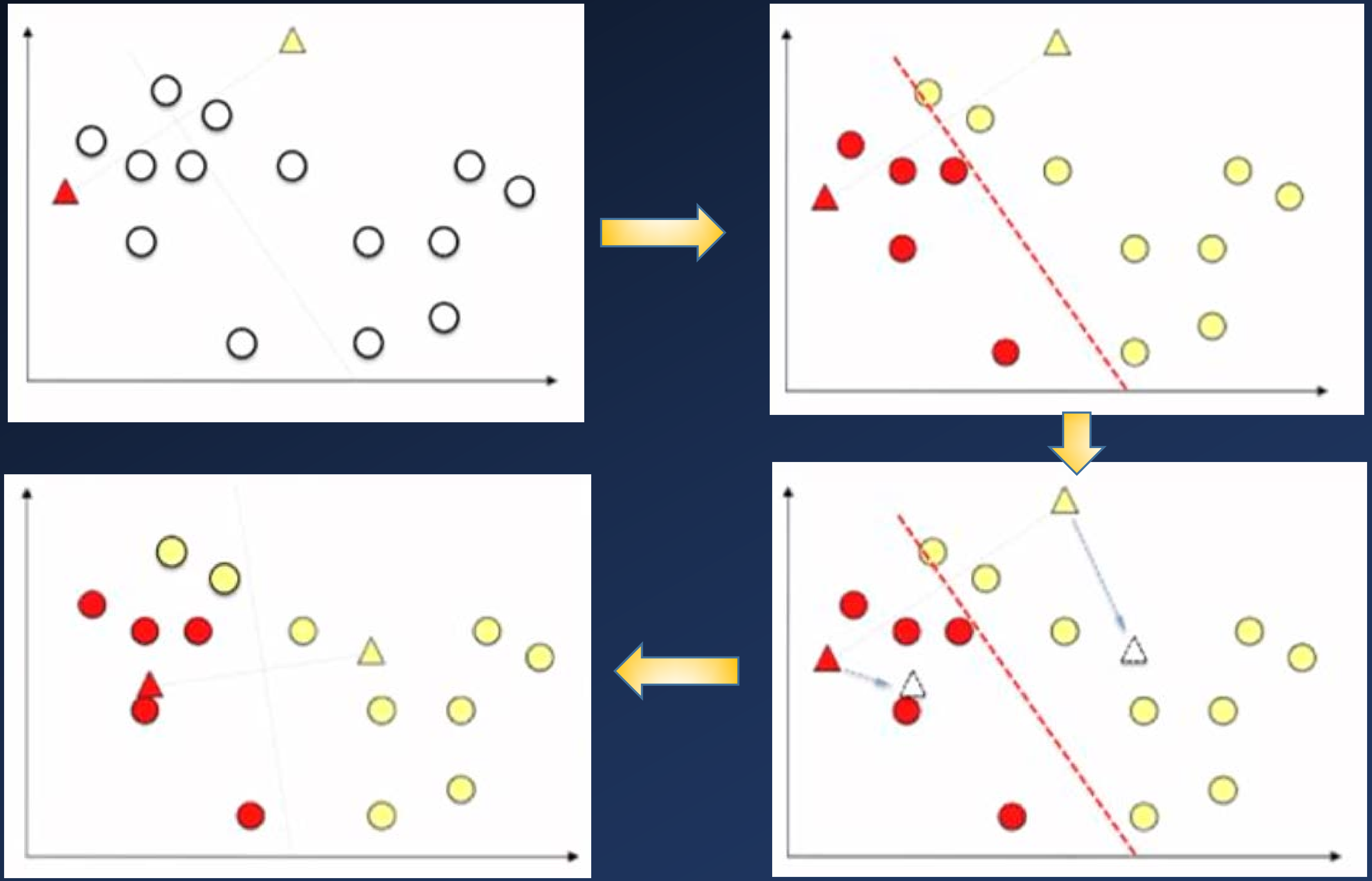
K-means

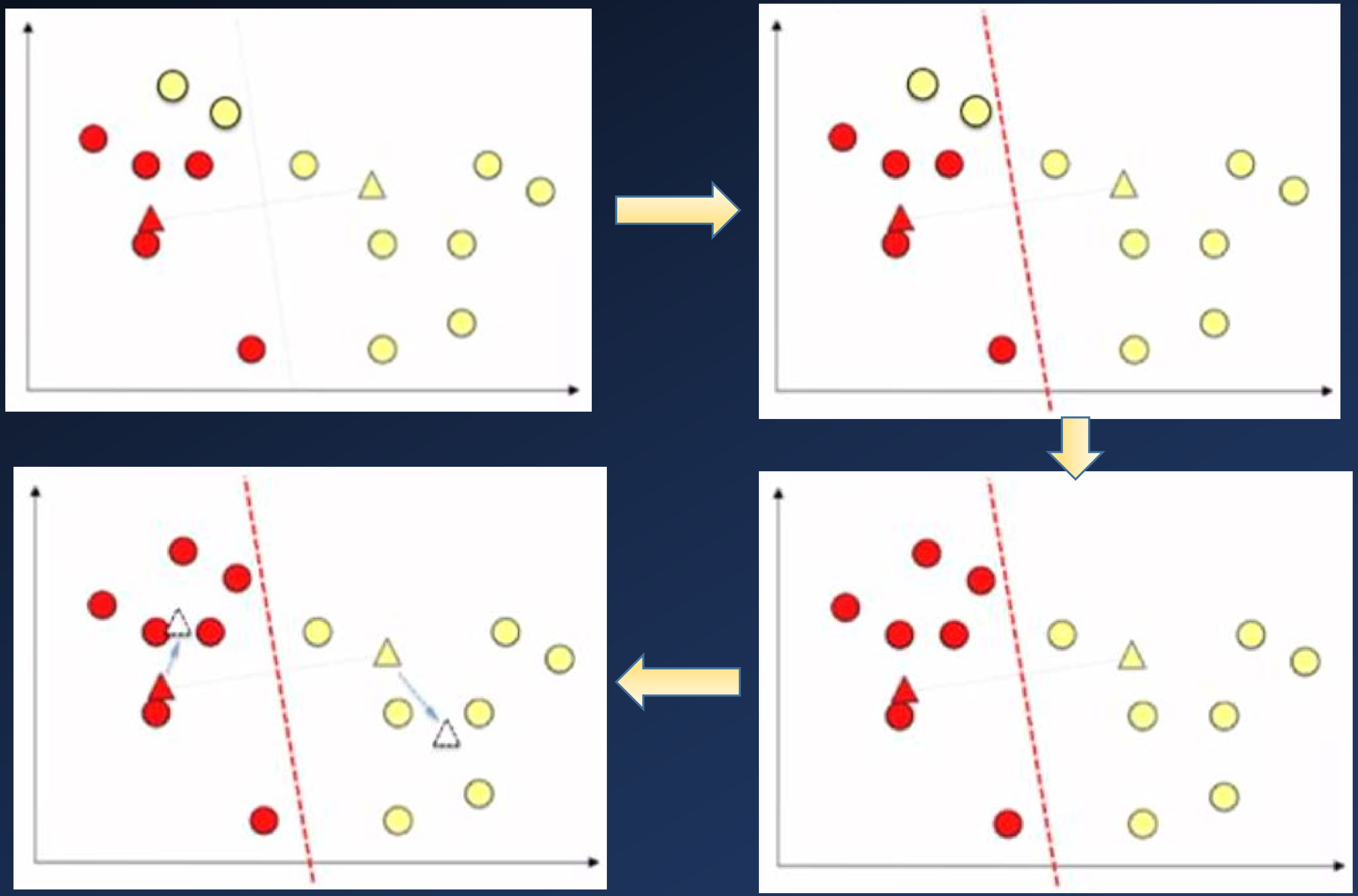


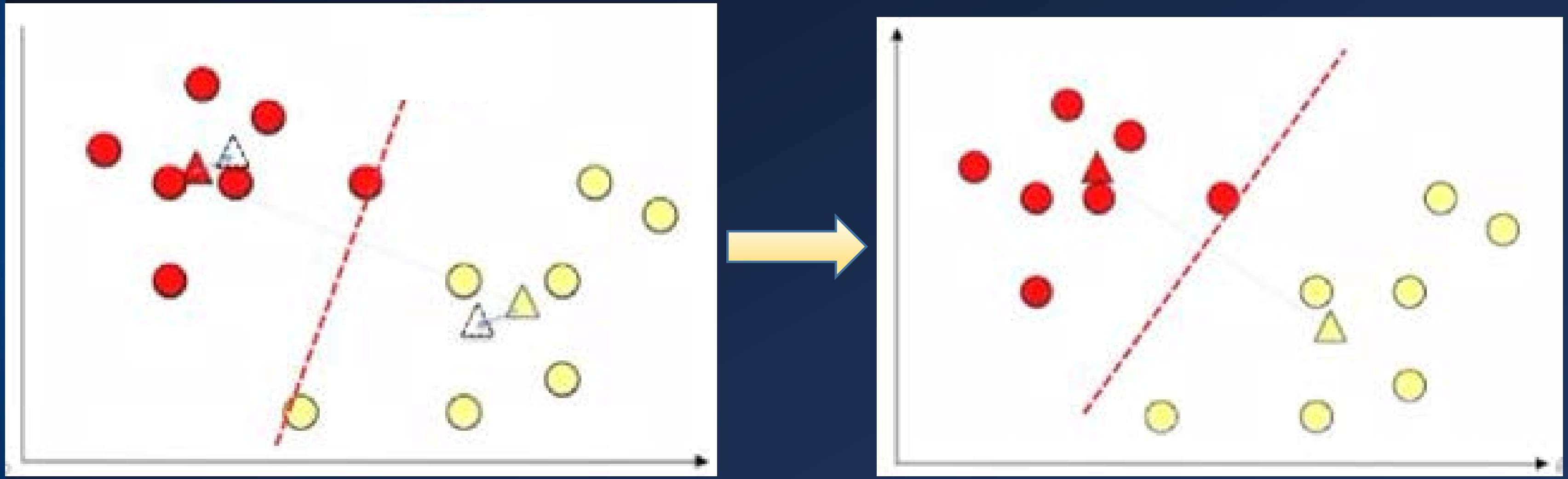
K-means clustering

- The most used clustering algorithm
- Runs on distance calculations, uses “Euclidean Distance” for this purpose.
- Tries to improve the inter group similarity while keeping the groups as far as possible from each other.
- “K” in K-Means represents the number of clusters in which we want our data to divide into.
- The basic restriction: your data should be continuous in nature. It won’t work if data is categorical in nature.

An example...







K-means algorithm

- Input:
    - K (number of clusters)
    - Training set {x^((1)),x^((2)),…,x^((m)) }, where x^((i)) ϵ〖 R〗^n

- Randomly initialize K cluster centroids $u_1, u_2, \ldots,\ u_k \epsilon R^n$

- Repeat until convergence [none of the cluster assignments change]

```
{
  for $i$=1 $to\ m$                                   [for each point]
        $c^{(i)}$ : = index (from 1 $to\ K$) of cluster centroid closest to $x^{(i)}$
  for $k$=1 $to\ K$                                   [for each cluster]
        $u_k$ := average (mean) of points assigned to cluster $k$
}
```

Summary

- Deep Learning

- K-means clustering