MACHINE LEARNING
EVOLUTION AND ALGORITHMS

JACK HAN

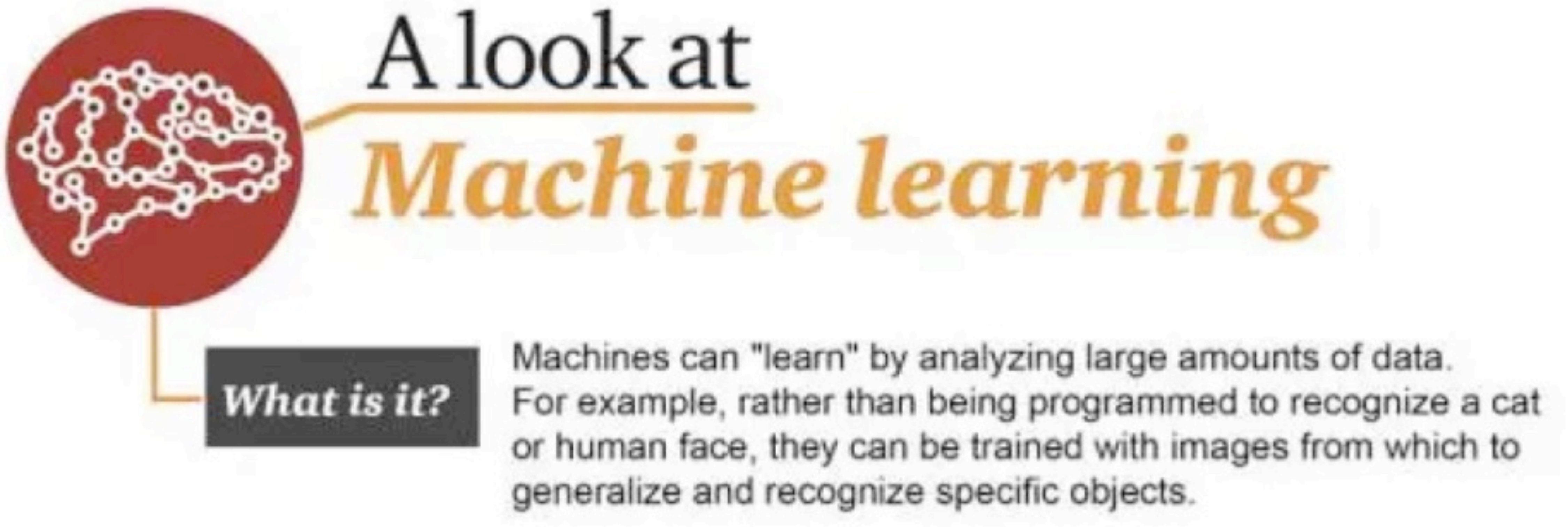
COMPUTER SCIENCE DEPARTMENT

CSUDH

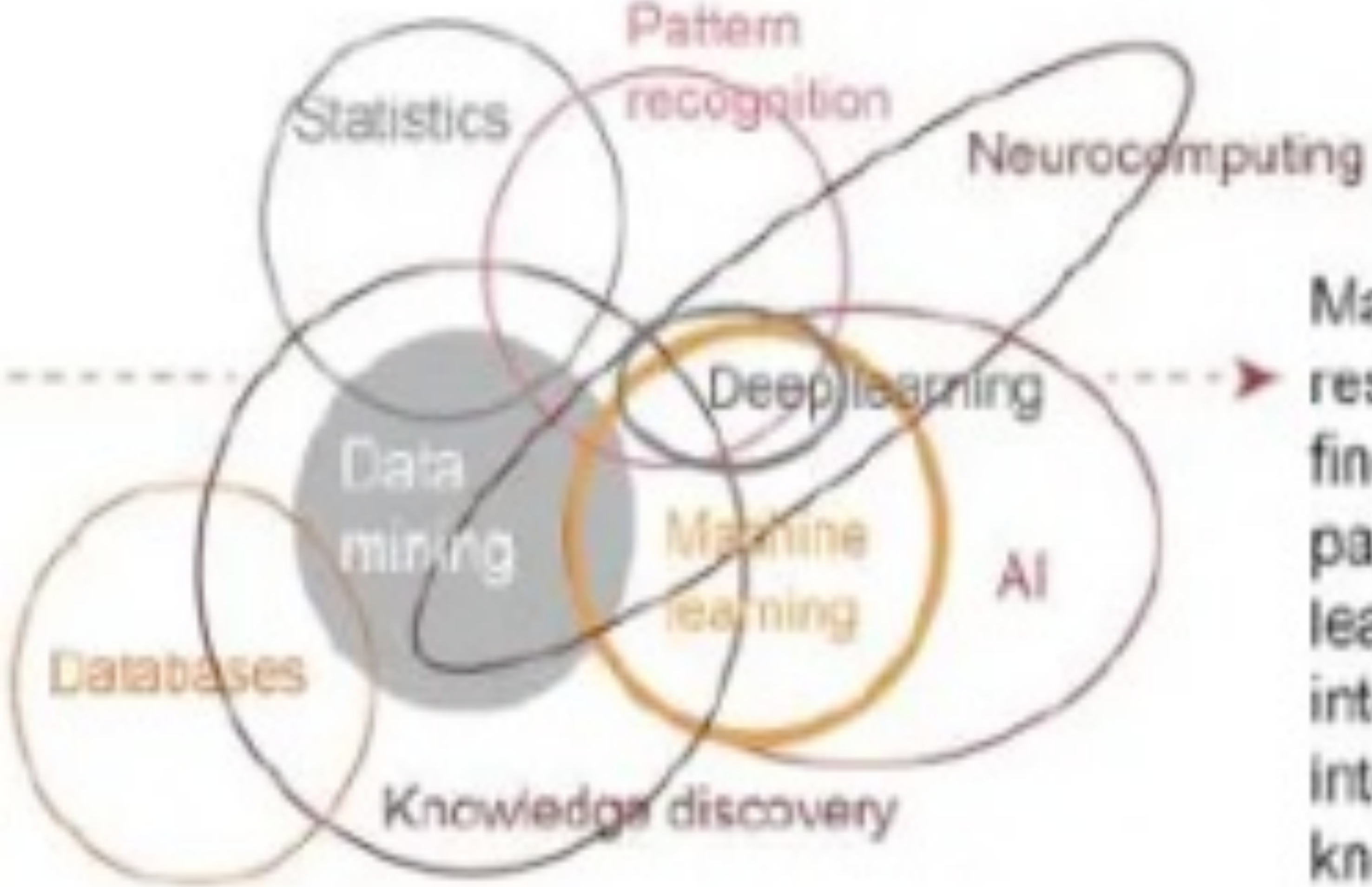
AGENDA

- What is Machine Learning?
- Machine Learning Evolution
- Five Tribes
- Methods/Algorithms

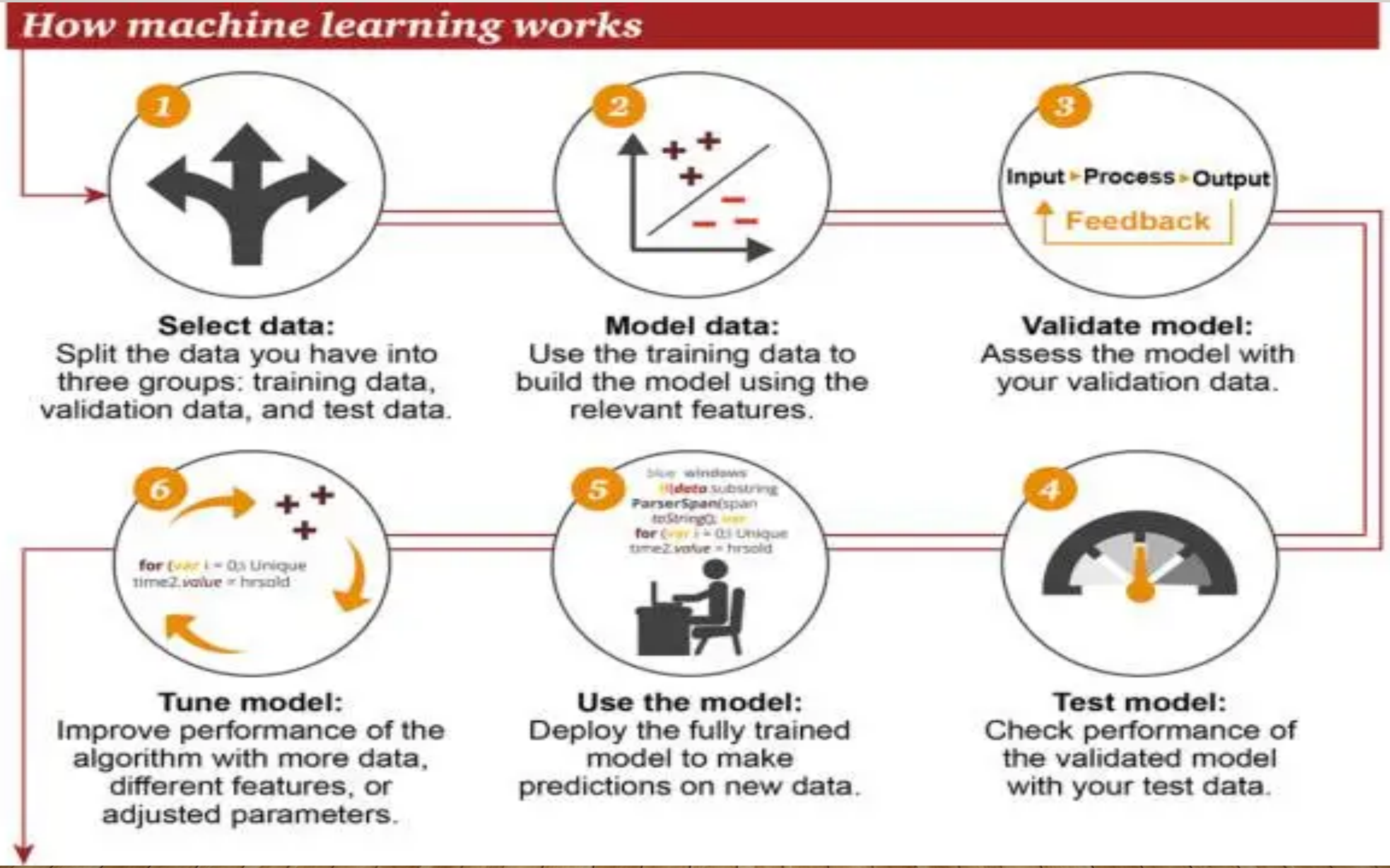
WHAT IS MACHINE LEARNING



MACHINE LEARNING AND OTHER FIELDS



Machine learning is a category of research and algorithms focused on finding patterns in data and using those patterns to make predictions. Machine learning falls within the artificial intelligence (AI) umbrella, which in turn intersects with the broader field of knowledge discovery and data mining.


How machine learning works
1
2
3
Input ▸ Process ▸ Output
Feedback
Select data:
Split the data you have into three groups: training data, validation data, and test data.
Model data:
Use the training data to build the model using the relevant features.
Validate model:
Assess the model with your validation data.
6
5
4
Tune model:
Improve performance of the algorithm with more data, different features, or adjusted parameters.
Use the model:
Deploy the fully trained model to make predictions on new data.
Test model:
Check performance of the validated model with your test data.

How machine learning fits in

1 Traditional programming

The software engineer writes a program that solves a problem.

Data

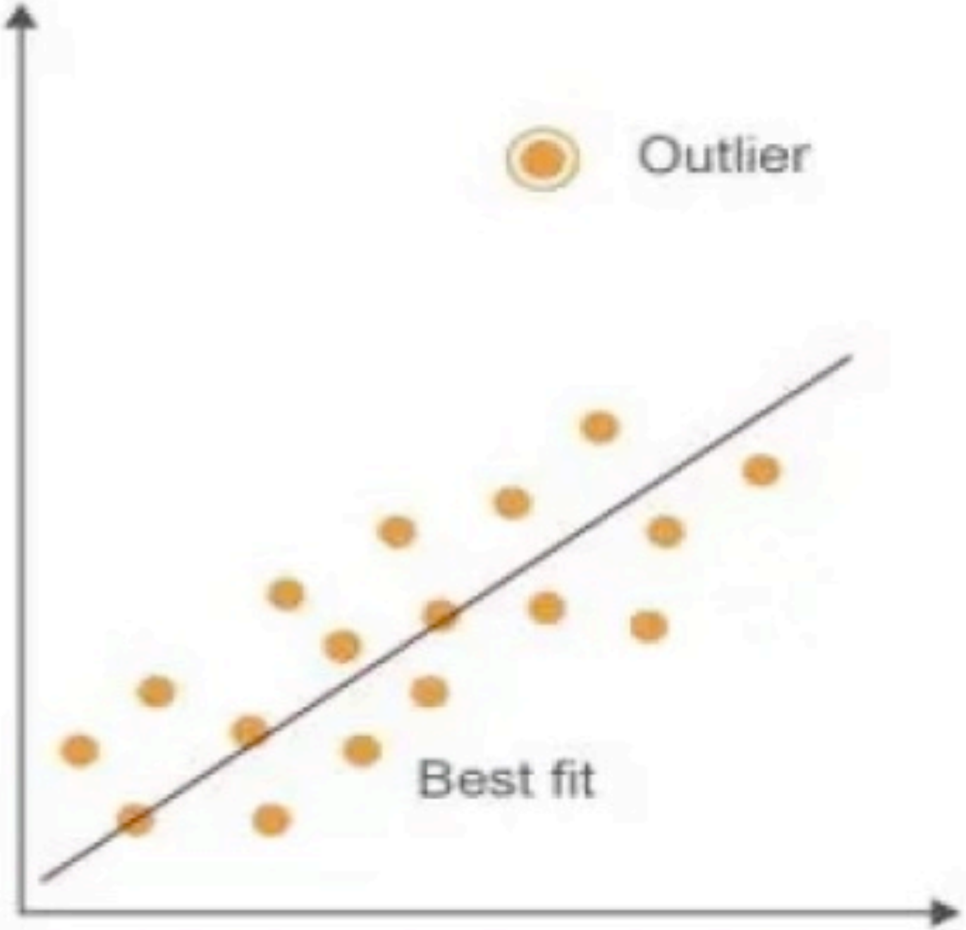
Software engineer writes a procedure that tells the machine what to do to solve the problem.

```
class FREE!!! implements SpamDetector
{
 public function detect($string)
 {
  if (str_word_count($string) = FREE!!!) {
     return true;
   }
   return false;
 }
}
```

Computer follows the procedure and generates a result.

2 Statistics

An analyst compares the relationships of variables.

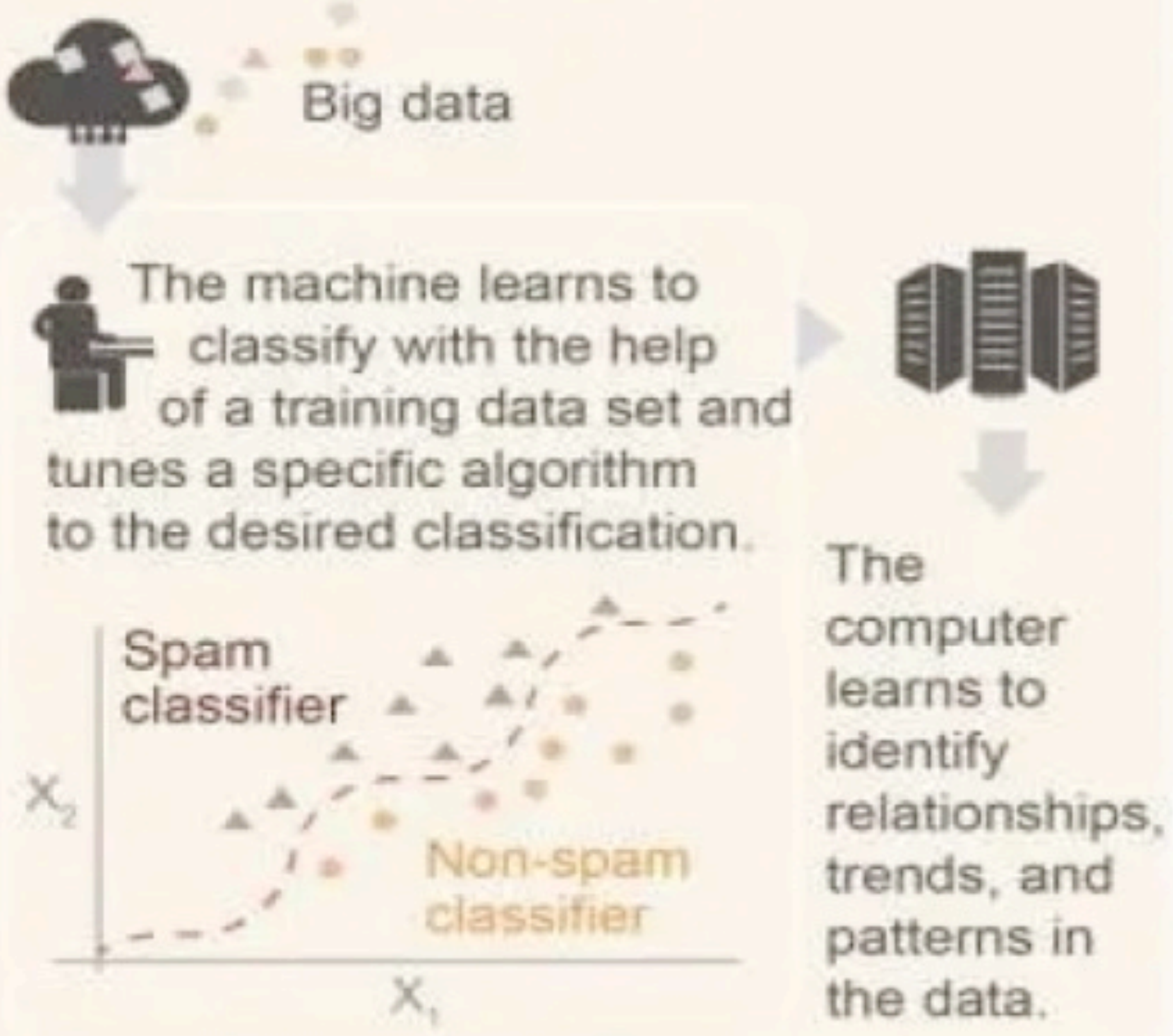


4 Intelligent apps

Intelligent apps leverage the outputs of AI, as in this precision farming example that uses drone-based data collection.

3 Machine learning

A data scientist uses a training data set to teach the computer what to do, and the system carries out the tasks.



Machine learning in practice

Here are just a few of the many ways we've put machine learning to work. How will your company use it?

Rapid 3D mapping and modeling

For a railway bridge reconstruction, PwC data scientists and domain experts applied machine learning to data captured from drones. The combination enabled precise monitoring and quick feedback on work in progress.

Enhanced profiling to mitigate risks

To detect insider trading, PwC combined machine learning with other analytic techniques to develop more comprehensive user profiles and gain deeper insight into complex suspicious behaviors.

Predicting the top performers

PwC used machine learning and other analysis to evaluate the potential of different horses running in the Melbourne Cup.

MACHINE LEARNING EVOLUTION

A look at

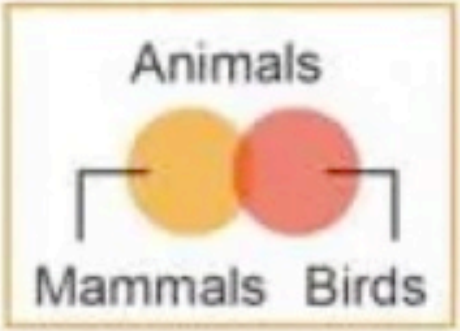
Machine learning evolution

Overview

For decades, individual “tribes” of artificial intelligence researchers have vied with one another for dominance. Is the time ripe now for tribes to collaborate? They may be forced to, as collaboration and algorithm blending are the only ways to reach true artificial general intelligence (AGI). Here’s a look back at how machine learning methods have evolved and what the future may look like.

What are the five tribes?

Symbolists



Use symbols, rules, and logic to represent knowledge and draw logical inference

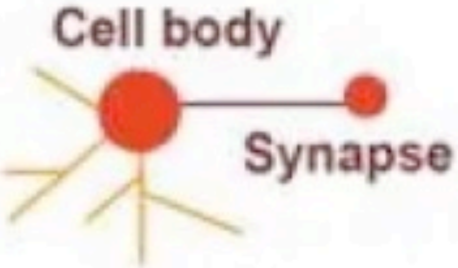
Favored algorithm
Rules and decision trees

Bayesians

Assess the likelihood of occurrence for probabilistic inference

Favored algorithm
Naive Bayes or Markov

Connectionists



Recognize and generalize patterns dynamically with matrices of probabilistic, weighted neurons

Favored algorithm
Neural networks

Evolutionaries

Generate variations and then assess the fitness of each for a given purpose

Favored algorithm
Genetic programs

Analogizers

Optimize a function in light of constraints ("going as high as you can while staying on the road")

Favored algorithm
Support vectors

Phases of evolution

1980s

Predominant tribe
Symbolists

Architecture
Server or mainframe

Predominant theory
Knowledge engineering

Basic decision logic:
Decision support systems with limited utility

1990s to 2000

Predominant tribe
Bayesians

Architecture
Small server clusters

Predominant theory
Probability theory

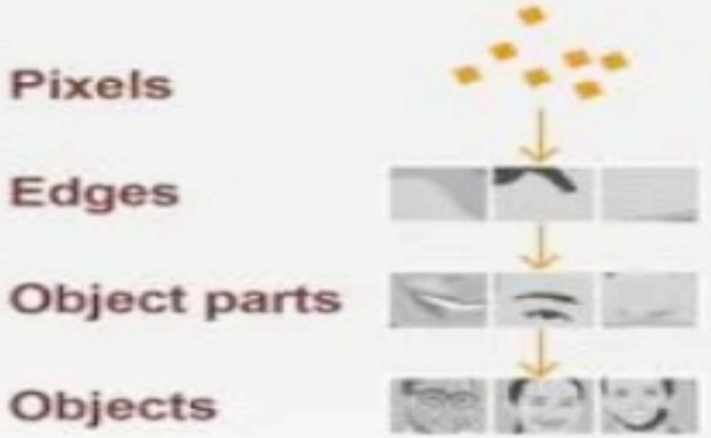
Classification:
Scalable comparison and contrast that's good enough for many purposes

Early to mid-2010s

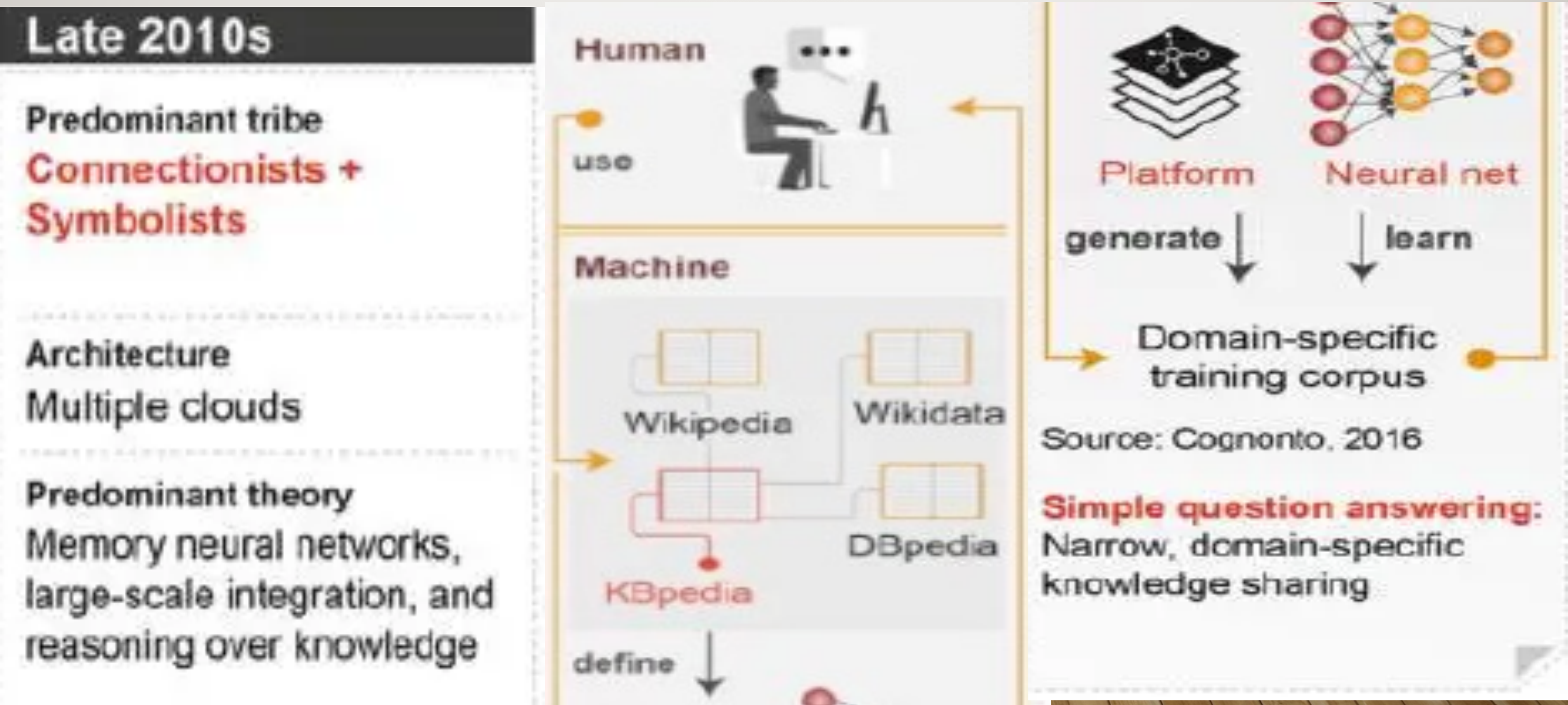
Predominant tribe
Connectionists

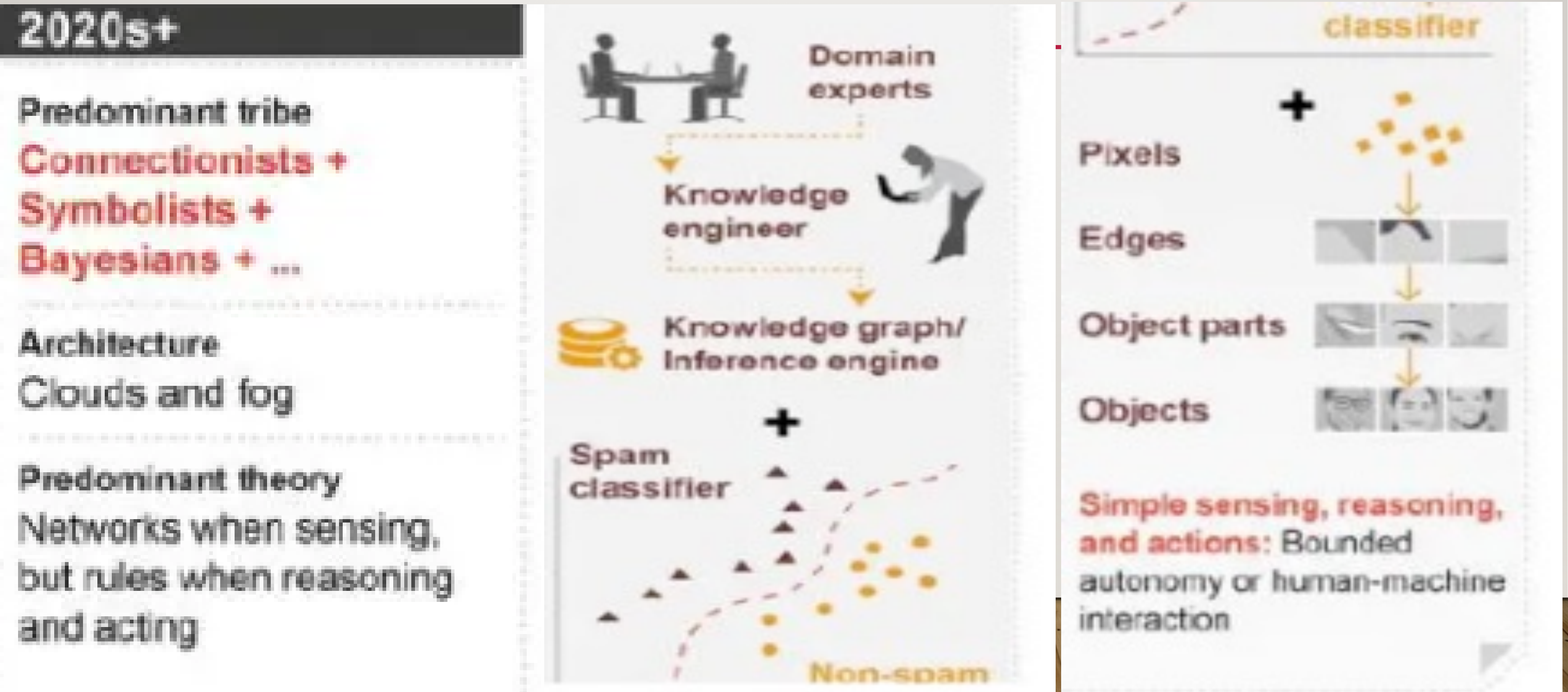
Architecture
Large server farms (the cloud)

Predominant theory
Neuroscience and probability



Recognition:
More precise image and voice recognition, translation, sentiment analysis, etc.


Late 2010s
Predominant tribe
Connectionists + Symbolists
Architecture
Multiple clouds
Predominant theory
Memory neural networks, large-scale integration, and reasoning over knowledge
Human
use
Machine
Wikipedia
Wikidata
KBpedia
DBpedia
define
Platform
Neural net
generate
learn
Domain-specific training corpus
Source: Cognonto, 2016
Simple question answering:
Narrow, domain-specific knowledge sharing


2020s+
Predominant tribe
Connectionists + Symbolists + Bayesians + ...
Architecture
Clouds and fog
Predominant theory
Networks when sensing, but rules when reasoning and acting
Domain experts
Knowledge engineer
Knowledge graph/ Inference engine
+
Spam classifier
Non-spam
classifier
+
Pixels
Edges
Object parts
Objects
Simple sensing, reasoning, and actions: Bounded autonomy or human-machine interaction

2040s+

Predominant tribe

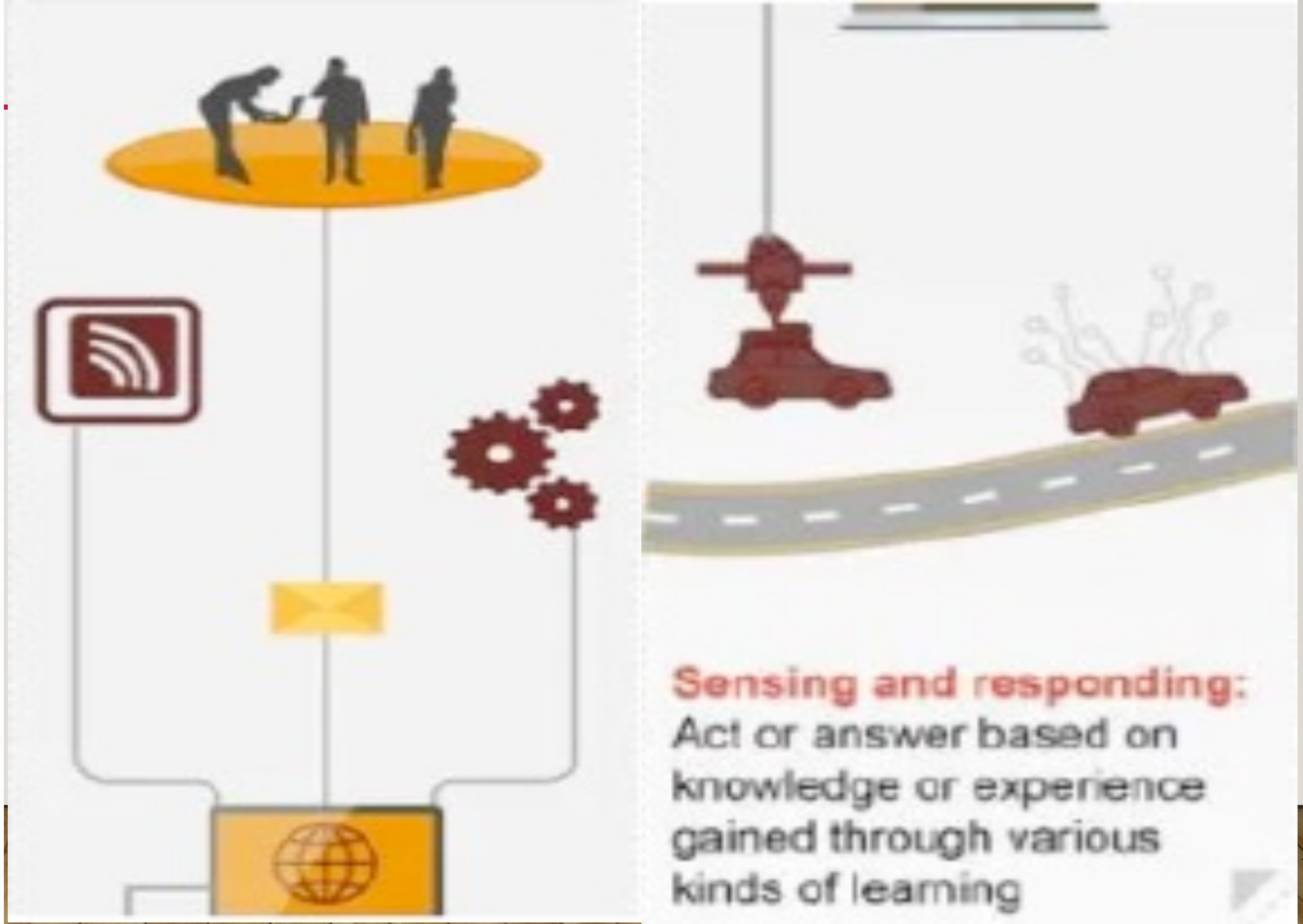
Algorithmic convergence

Architecture

Server ubiquity

Predominant theory

Best-of-breed meta-learning



MACHINE LEARNING METHODS

A look at

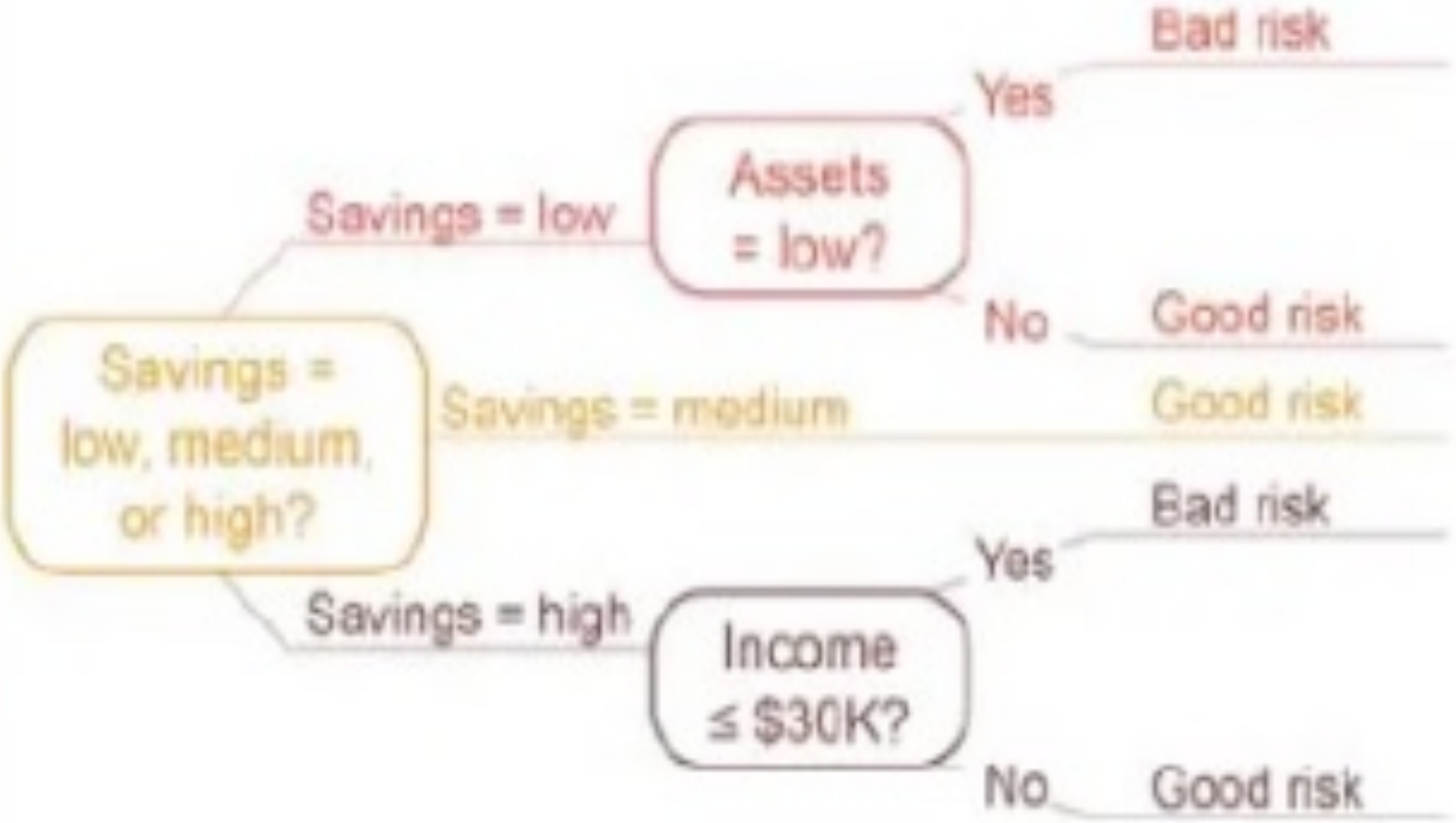
Machine learning methods

Introduction

Which machine learning algorithm should you use? A lot depends on the characteristics and the amount of the available data, as well as your training goals, in each particular use case. Avoid using the most complicated algorithms unless the end justifies more expensive means and resources. Here are some of the more common algorithms ranked by ease of use.

Decision trees

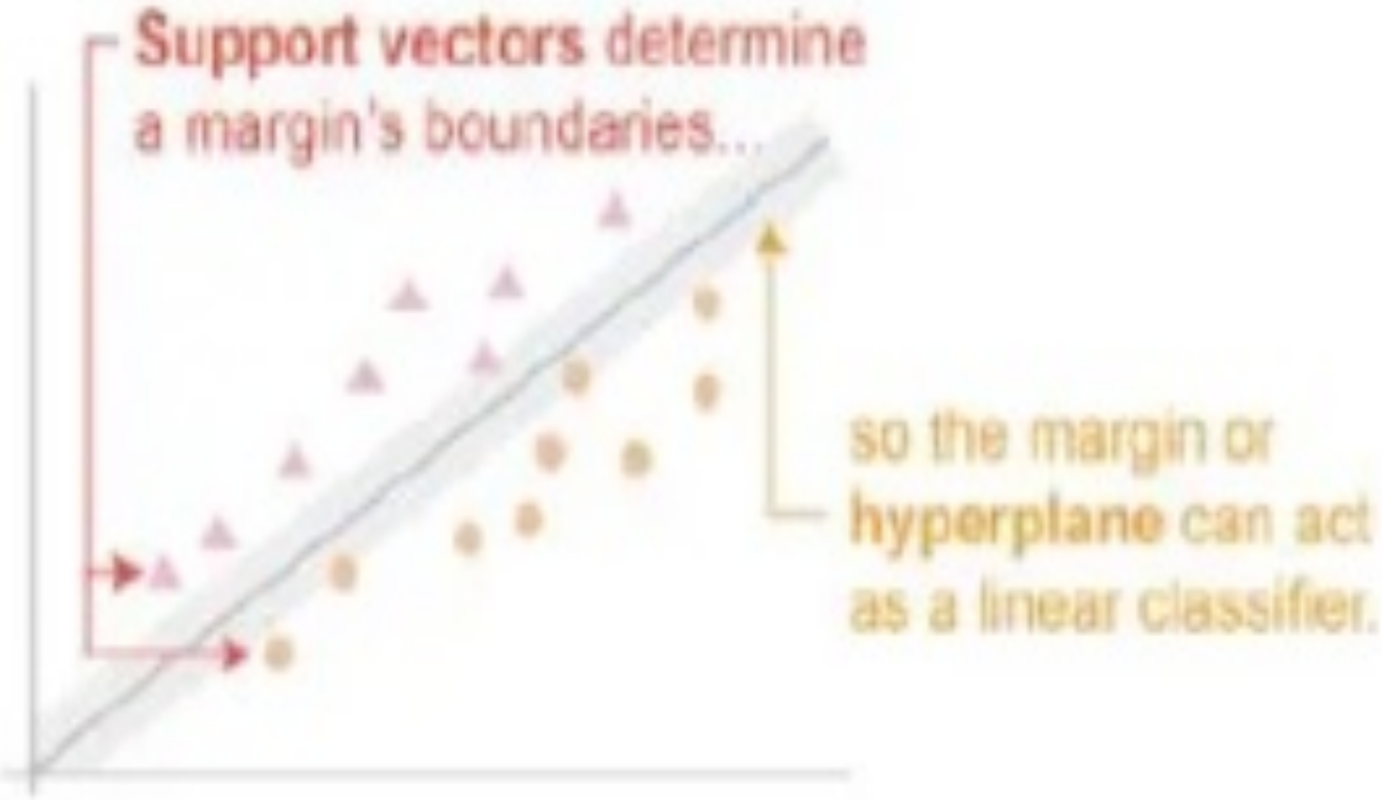
Decision tree analysis typically uses a hierarchy of variables or decision nodes that, when answered step by step, can classify a given customer as creditworthy or not, for example.



Advantages	Use cases
Decision trees are useful when evaluating lists of distinct features, qualities, or characteristics of people, places, or things.	Rule-based credit risk assessment, horse race performance prediction

Support vector machines

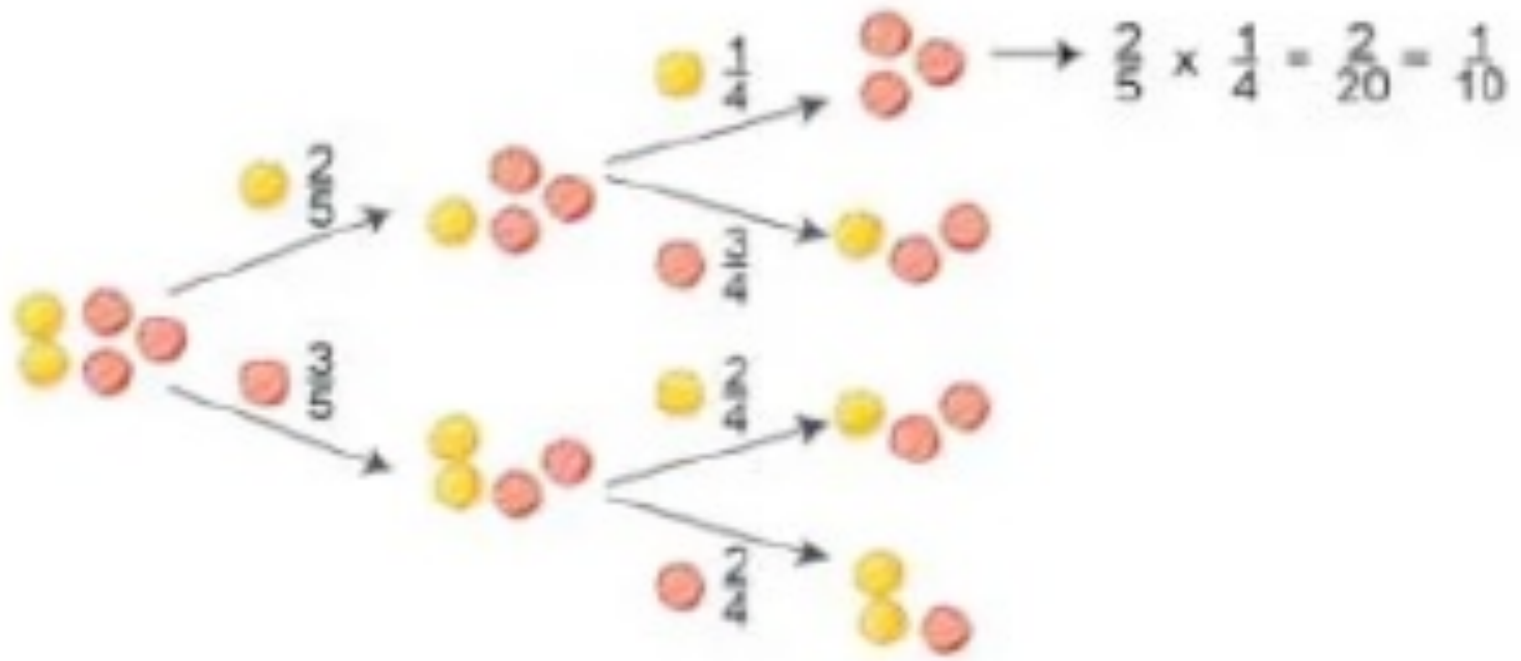
Support vector machines classify groups of data with the help of hyperplanes.



Advantages	Use cases
Support vector machines are good for the binary classification of X versus other variables and are useful whether or not the relationship between variables is linear.	News categorization, handwriting recognition

Naive Bayes classification

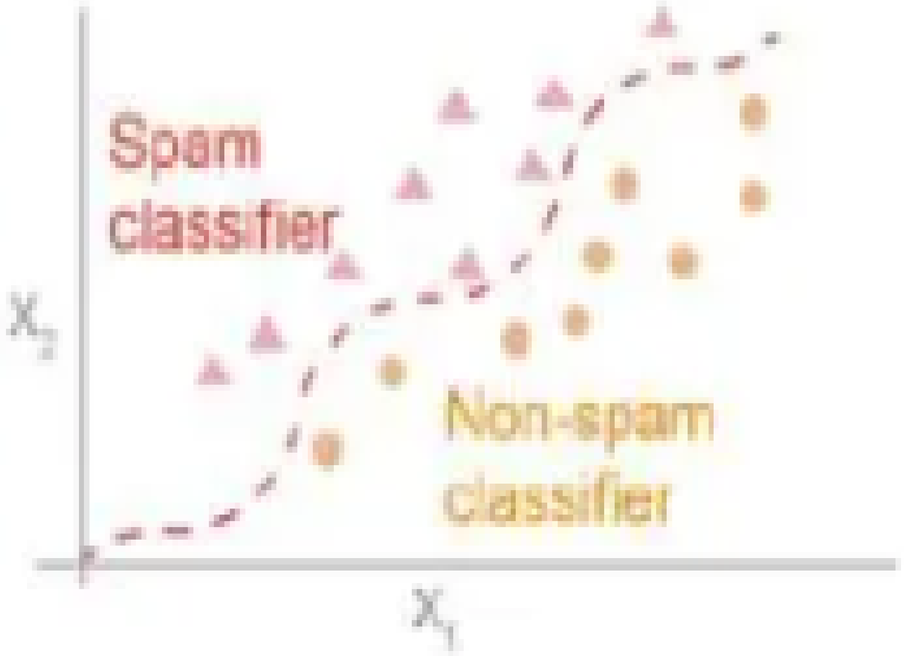
Naive Bayes classifiers compute probabilities, given tree branches of possible conditions. Each individual feature is "naive" or conditionally independent of, and therefore does not influence, the others. For example, what's the probability you would draw two yellow marbles in a row, given a jar of five yellow and red marbles total? The probability, following the topmost branch of two yellow in a row, is one in ten. Naive Bayes classifiers compute the combined, conditional probabilities of multiple attributes.



Advantages	Use cases
Naive Bayes methods allow the quick classification of relevant items in small data sets that have distinct features.	Sentiment analysis, consumer segmentation

Regression

Regression maps the behavior of a dependent variable relative to one or more dependent variables. In this example, logistic regression separates spam from non-spam text.

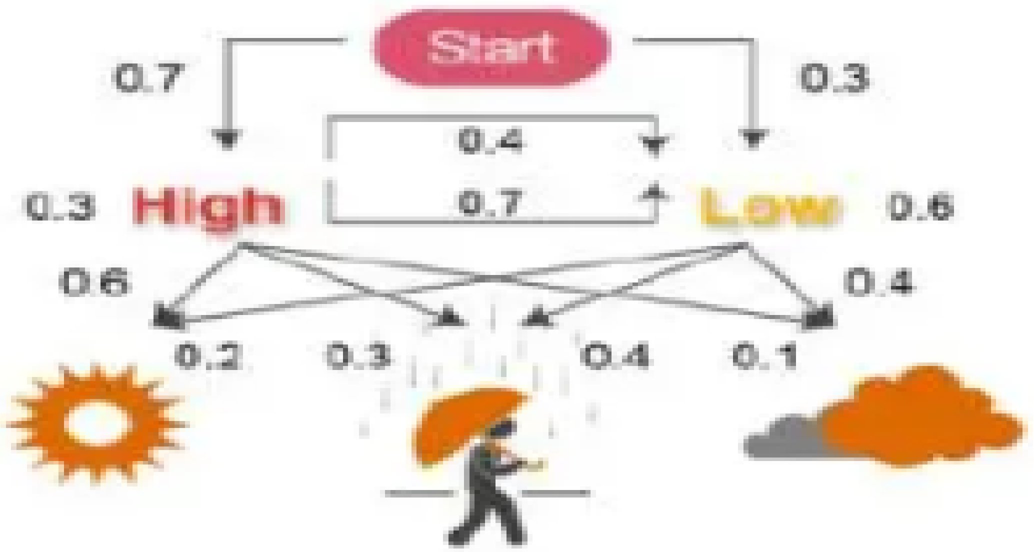


Advantages	Use cases
Regression is useful for identifying continuous (not necessarily distinct) relationships between variables.	Traffic flow analysis, email filtering

Hidden Markov models

Observable Markov processes are purely deterministic—one given state always follows another given state. Traffic light patterns are an example.

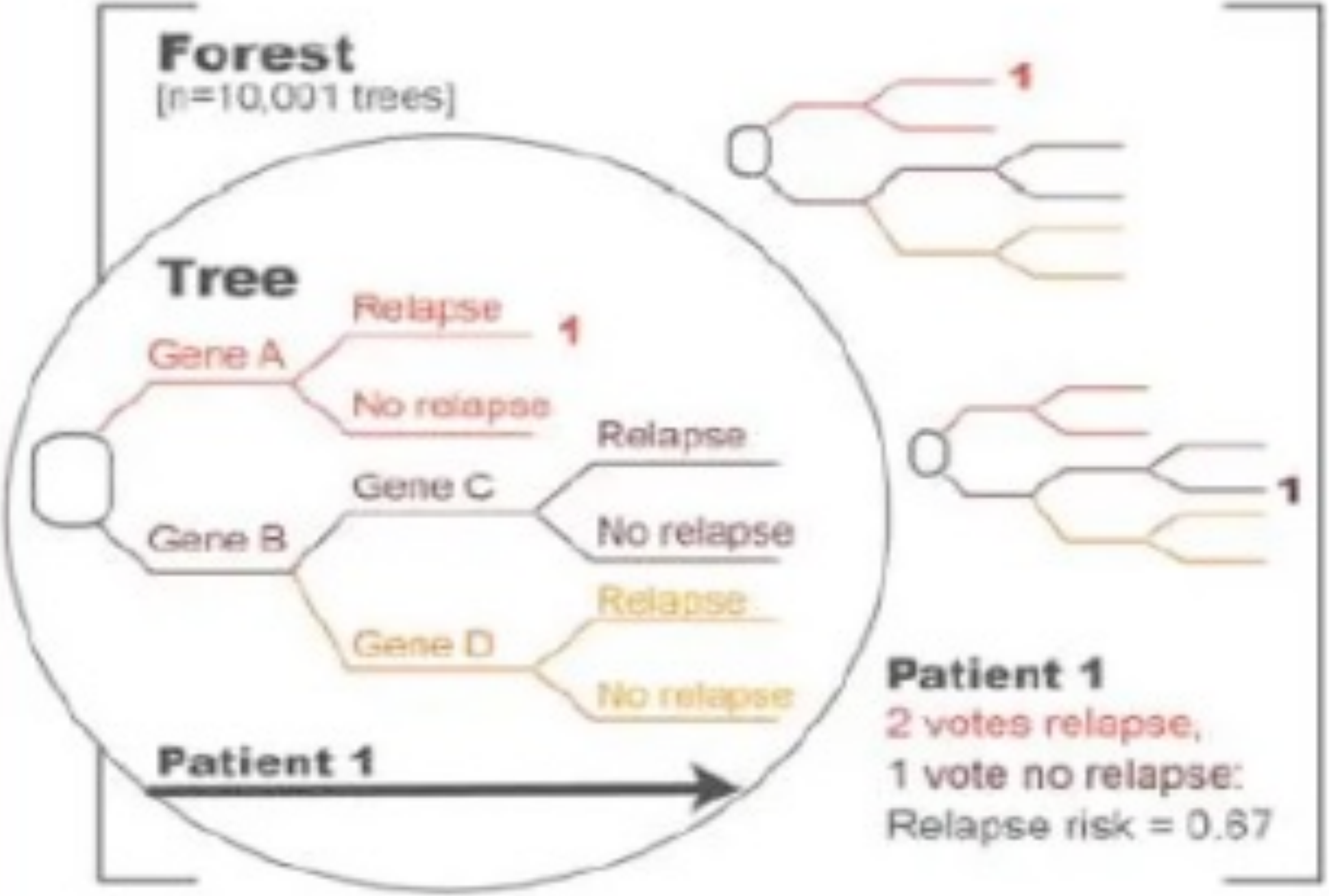
Hidden Markov models, by contrast, compute the probability of hidden states occurring by analyzing observable data, and then estimating the likely pattern of future observation with the help of the hidden state analysis. In this example, the probability of high or low pressure (the hidden state) is used to predict the likelihood of sunny, rainy, or cloudy weather.



Advantages	Use cases
Tolerates data variability and effective for recognition and prediction.	Facial expression analysis, weather prediction

Random forest

Random forest algorithms improve the accuracy of decision trees by using multiple trees with randomly selected subsets of data. This example reviews the expression levels of various genes associated with breast cancer relapse and computes a relapse risk.

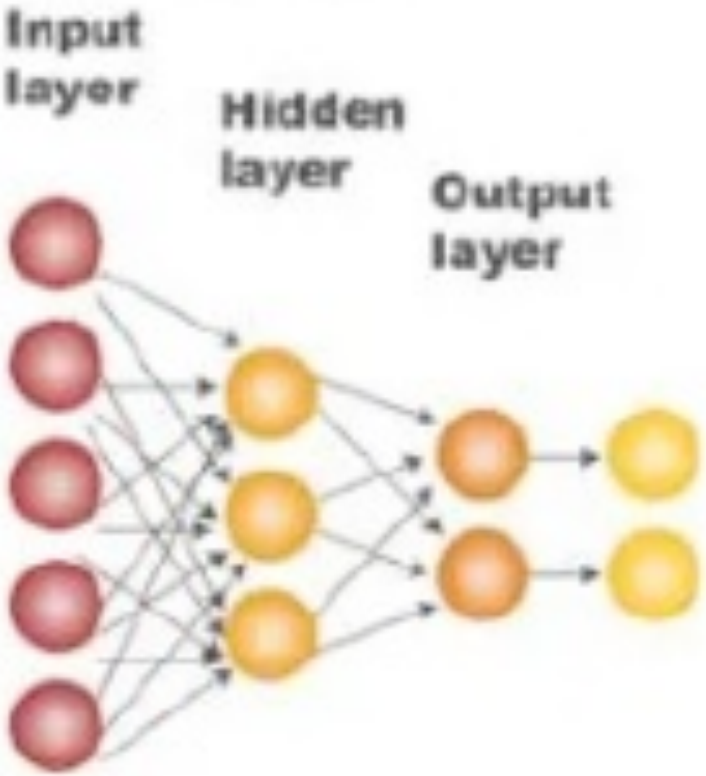


Advantages	Use cases
Random forest methods prove useful with large data sets and items that have numerous and sometimes irrelevant features.	Customer churn analysis, risk assessment

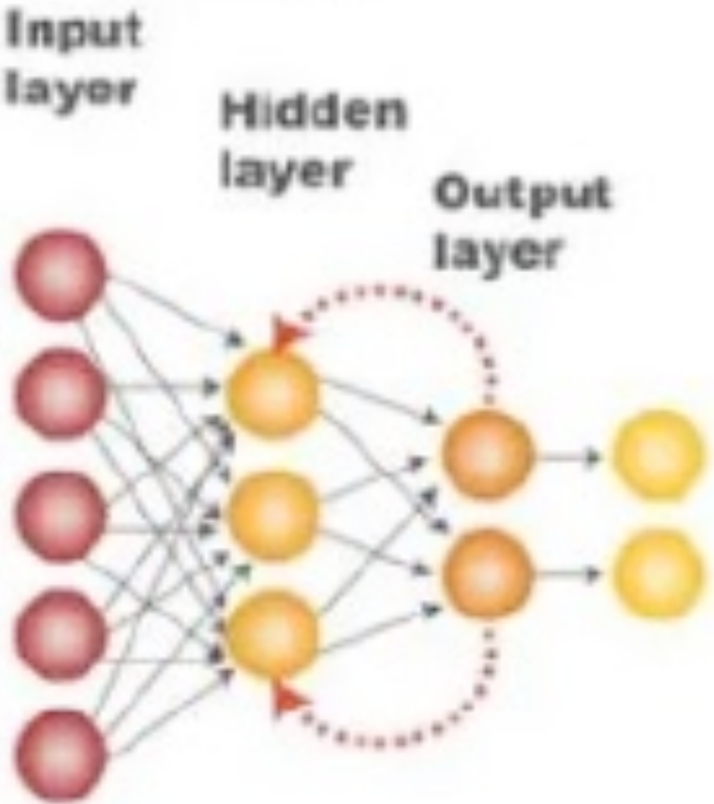
Recurrent neural networks

Each neuron in any neural network converts many inputs into single outputs via one or more hidden layers. Recurrent neural networks [RNNs] additionally pass values from step to step, making step-by-step learning possible. In other words, RNNs have a form of memory, allowing previous outputs to affect subsequent inputs.

Non-recurrent feed-forward neural network



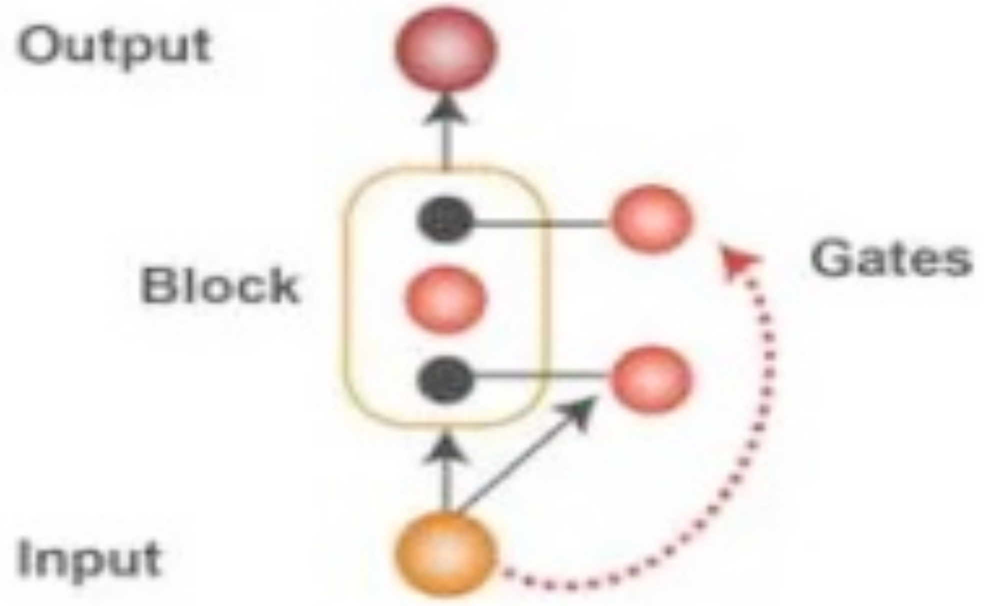
Recurrent neural network—includes loops



Advantages	Use cases
Recurrent neural networks have predictive power when used with large amounts of sequenced information.	Image classification and captioning, political sentiment analysis

Long short-term memory & gated recurrent unit neural networks

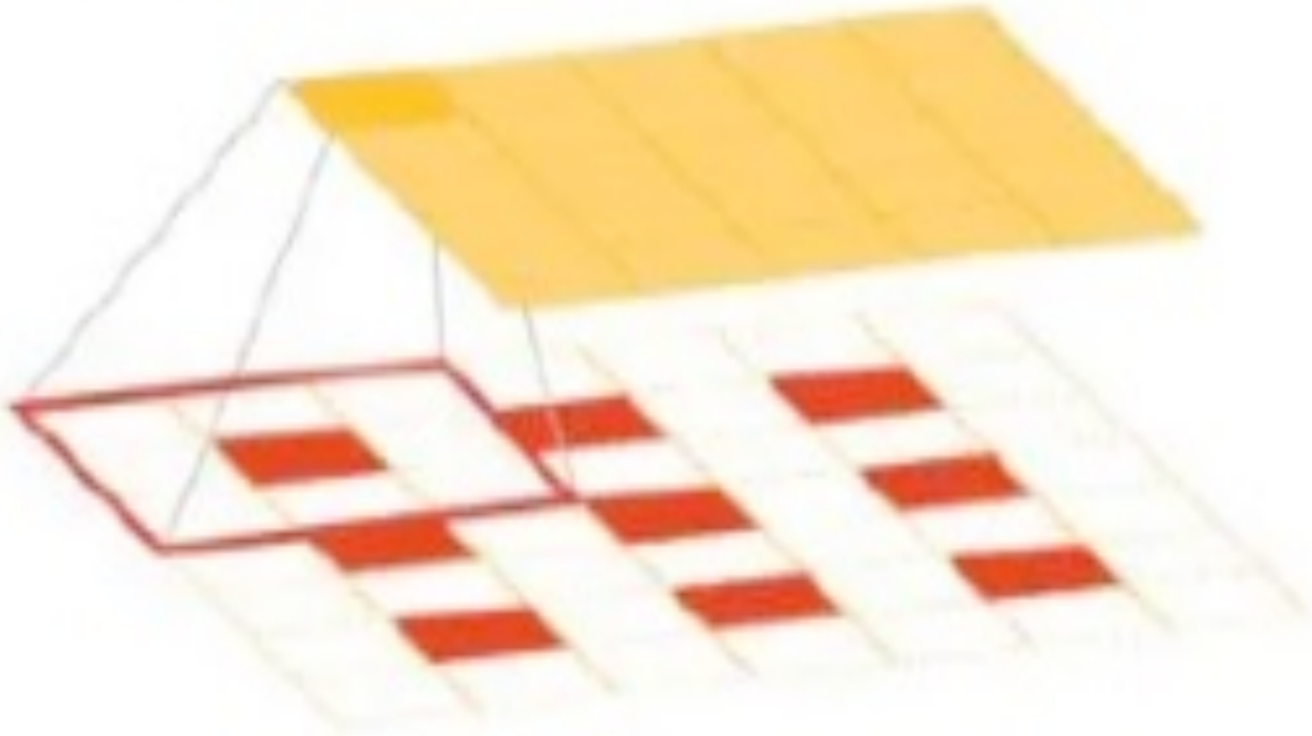
Older forms of RNNs can be lossy. While these older recurrent neural networks only allow small amounts of older information to persist, newer long short-term memory (LSTM) and gated recurrent unit (GRU) neural networks have both long- and short-term memory. In other words, these newer RNNs have greater memory control, allowing previous values to persist or to be reset as necessary for many sequences of steps, avoiding “gradient decay” or eventual degradation of the values passed from step to step. LSTM and GRU networks make this memory control possible with memory blocks and structures called gates that pass or reset values as appropriate.



Advantages	Use cases
Long short-term memory and gated recurrent unit neural networks have the same advantages as other recurrent neural networks and are more frequently used than other recurrent neural networks because of their greater memory capabilities.	Natural language processing, translation

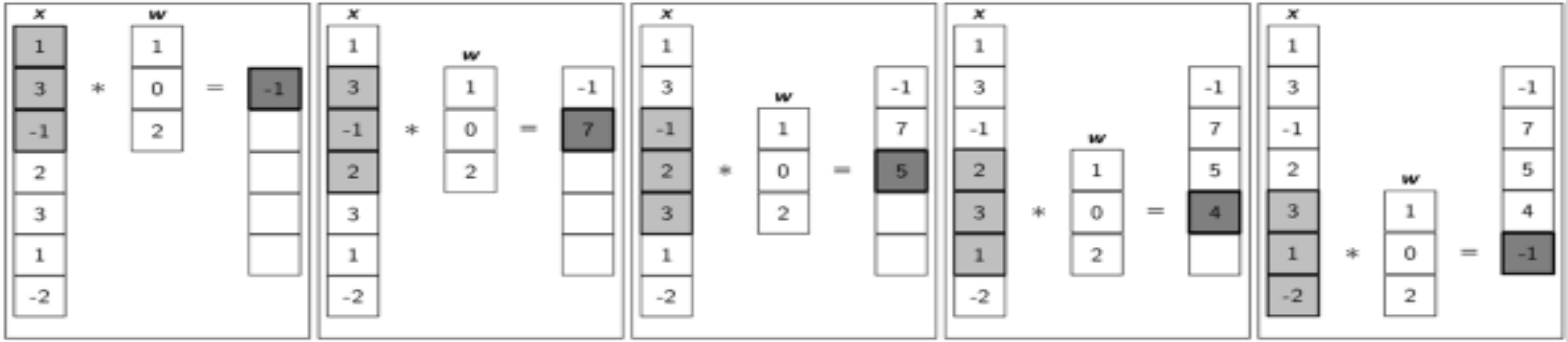
Convolutional neural networks

Convolutions are blends of weights from a subsequent layer that are used to label the output layer.

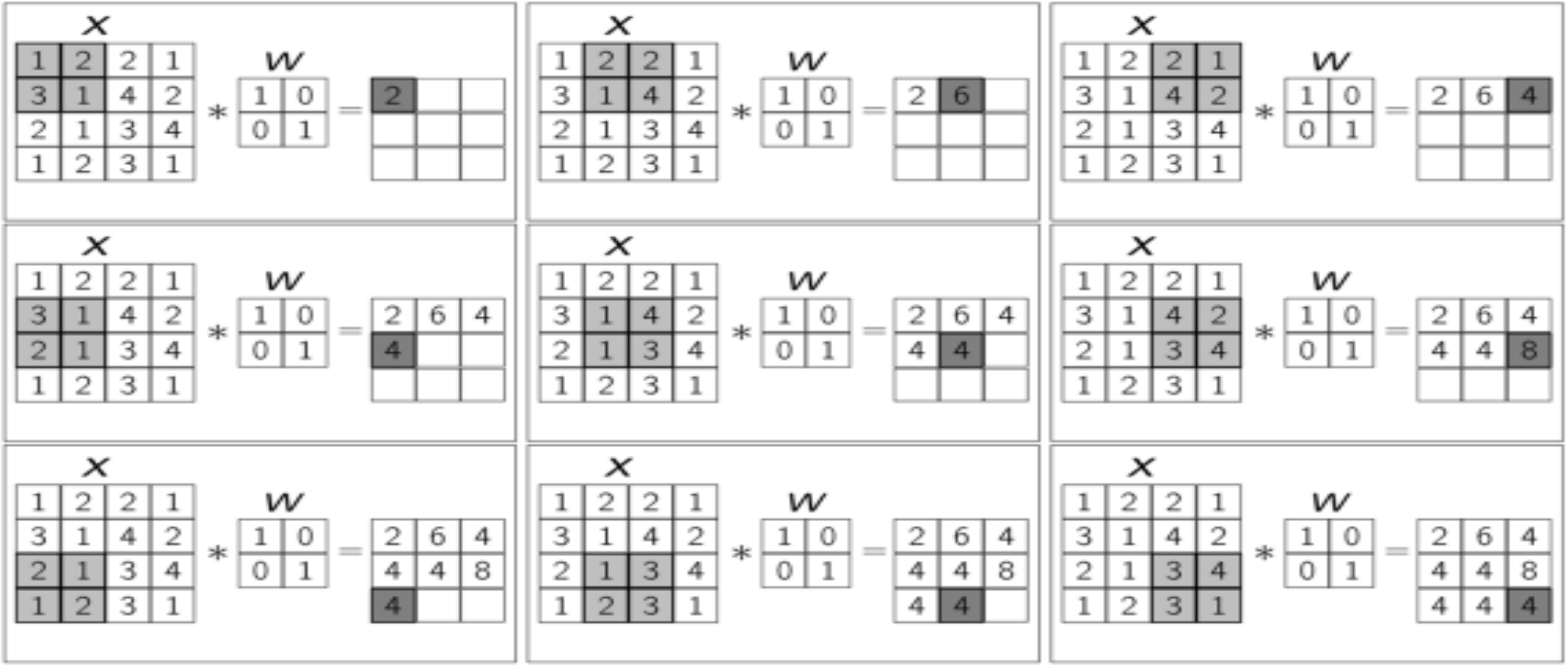


Advantages	Use cases
Convolutional neural networks are most useful with very large data sets, large numbers of features, and complex classification tasks.	Image recognition, text to speech, drug discovery

1D Convolution



2D Convolution



THANKS!

Questions?