DATA SCIENCE IN MANUFACTURING
WEEK 7

ANDREW SHERLOCK, JONATHAN CORNEY, DANAI KORRE

BY THE END OF THIS LECTURE YOU SHOULD:

- Be introduced to Machine Learning (ML) and Artificial Intelligence (AI)
- Get familiar with the uses of ML and AI in manufacturing
- Understand the basic principles behind ML and AI

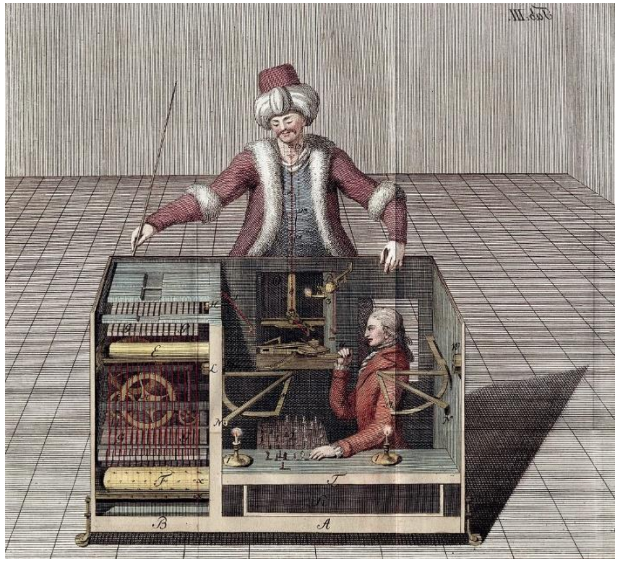
LECTURE: WEEK 7

Introduction to
Artificial Intelligence and Machine Learning

John McCarthy in his paper “What is artificial intelligence” [6] defines AI as “ the science and engineering of **making intelligent machines**, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.”

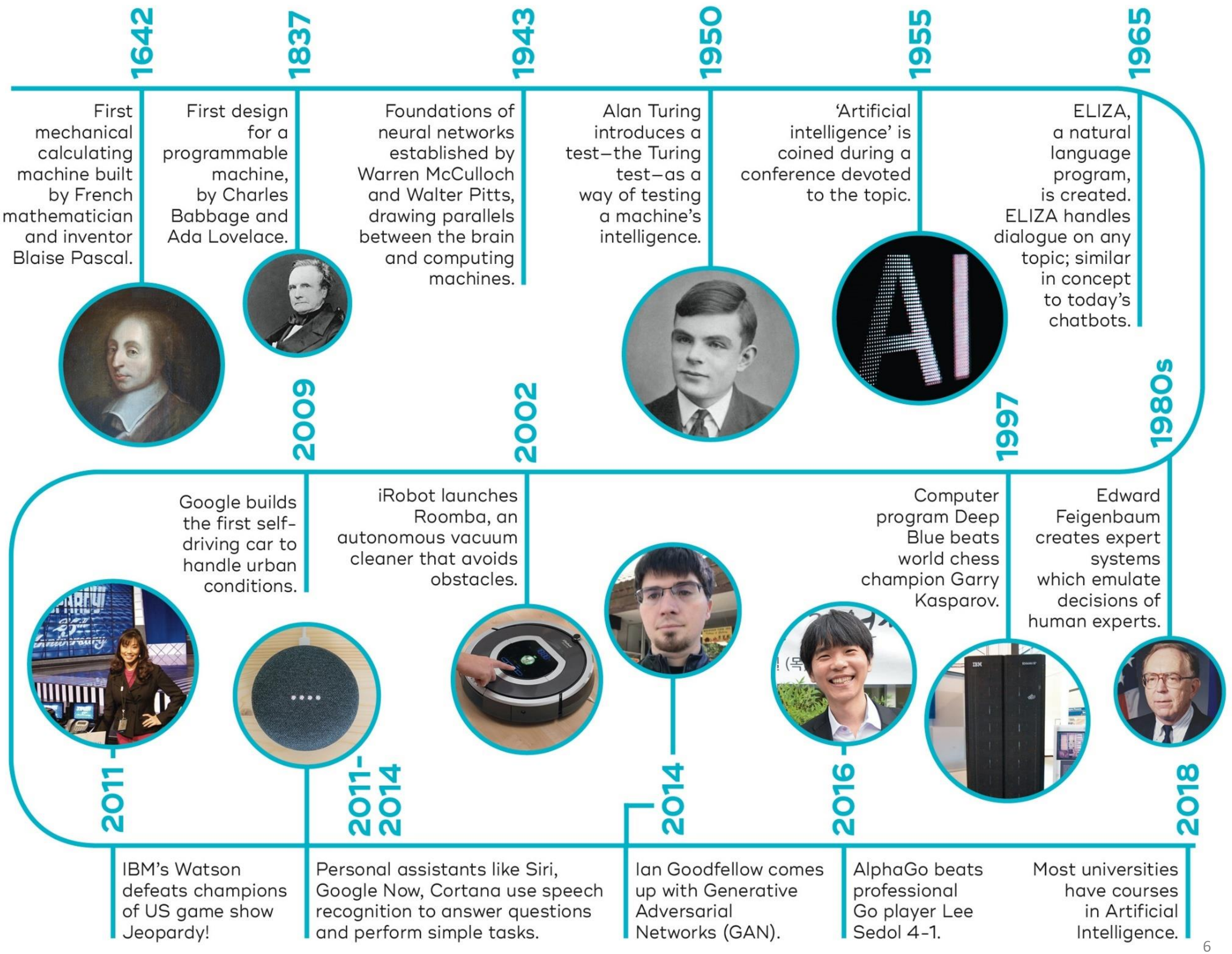
ARTIFICIAL INTELLIGENCE

Making machine “appear” intelligent



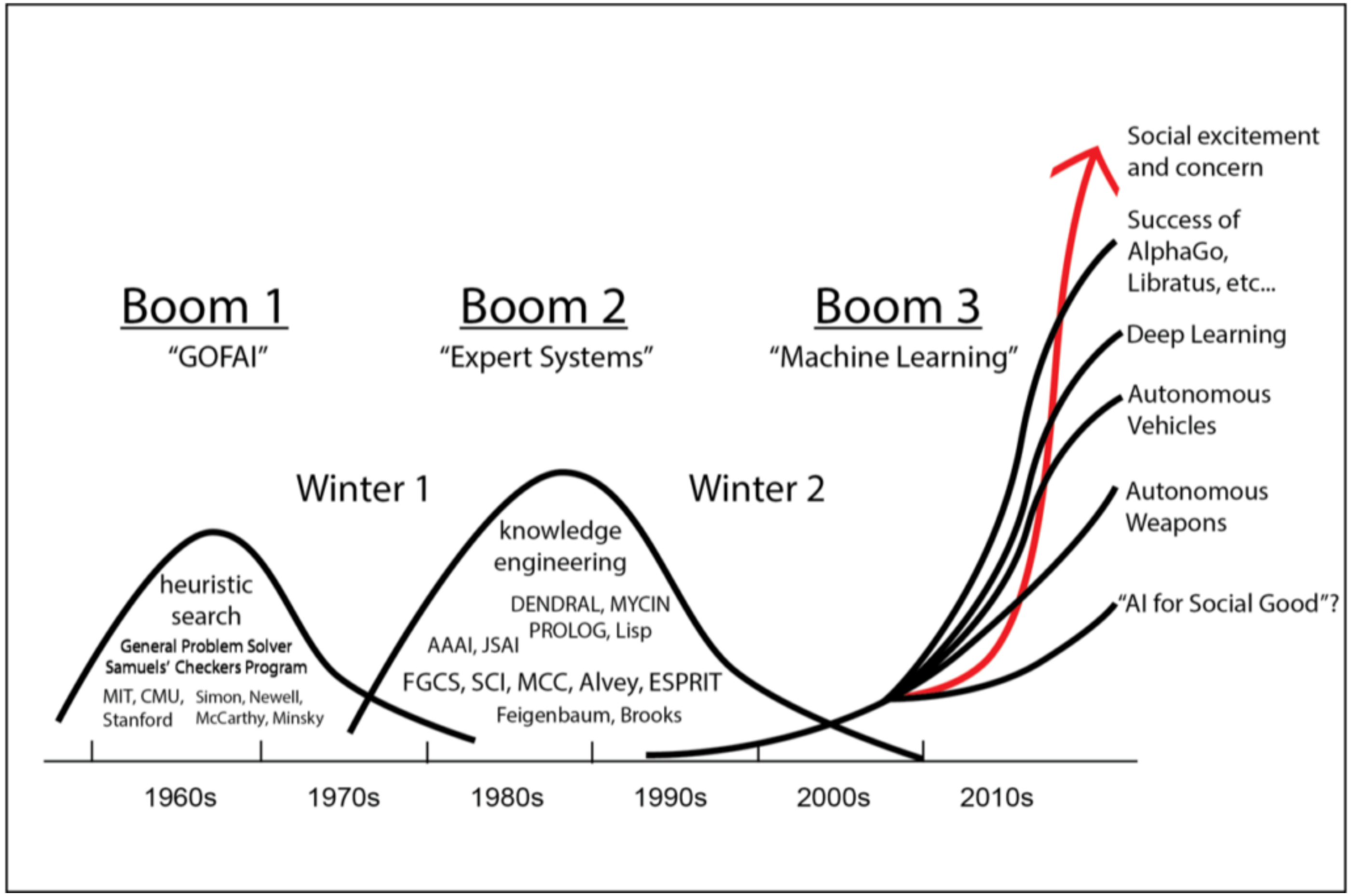
AI TIMELINE

Source: The University of Queensland



THE UNIVERSITY of EDINBURGH
School of Engineering

A History Of Boom And Bust



THE UNIVERSITY *of* EDINBURGH
School of Engineering

SINGULARITY

LAW OF ACCELERATING RETURNS

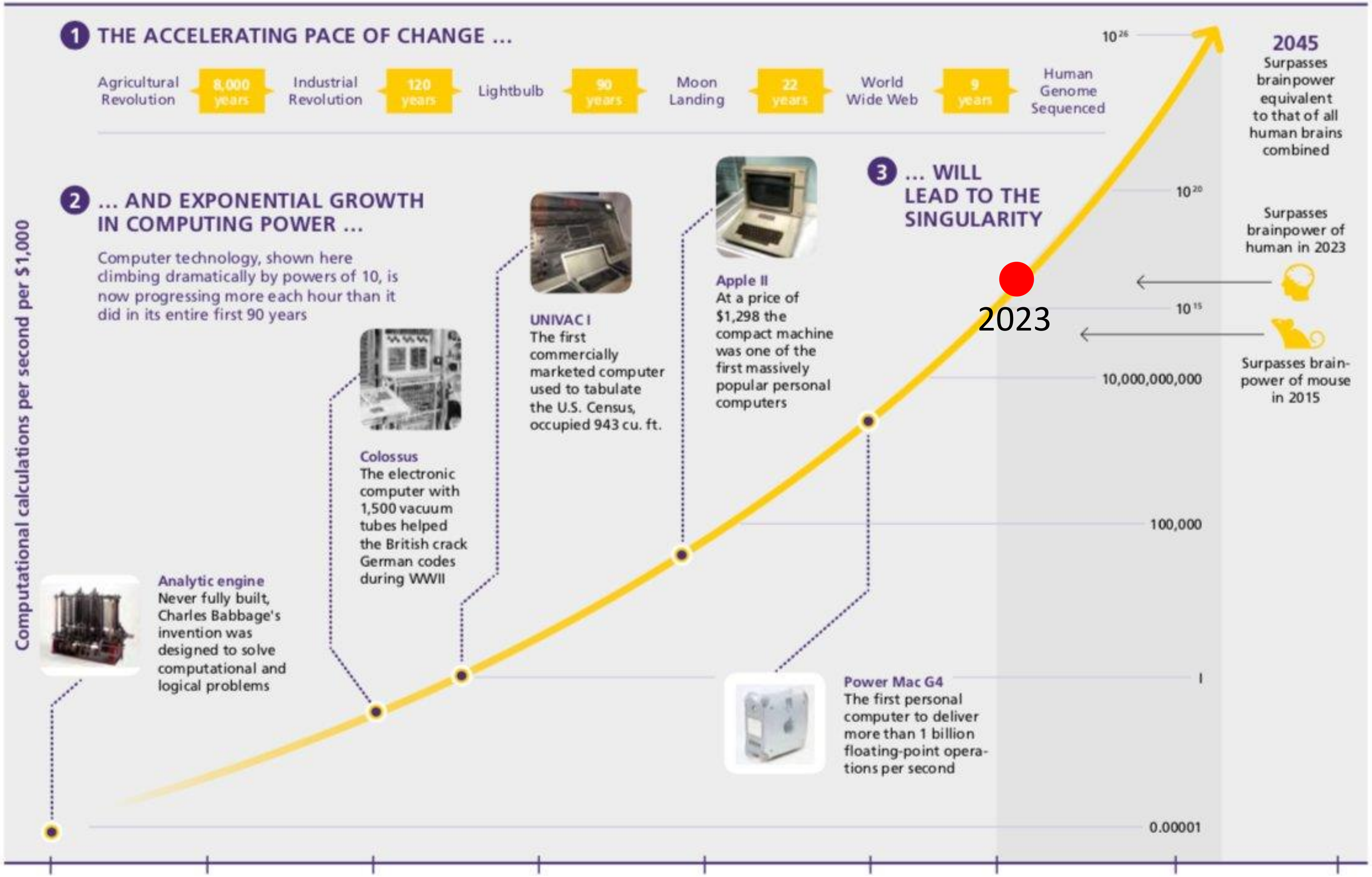


Figure 12: Ray Kurzweil's Law of Accelerating Returns depicts the exponential growth of computer processing power and technology innovations throughout history, and anticipates computers will exceed human intelligence in the future; **Source:** TIME / Wikipedia

source dhl via @mikequindazzi

THE UNIVERSITY of EDINBURGH
School of Engineering

SINGULARITY

"With artificial intelligence we're summoning the demon"
– Elon Musk

"Full artificial intelligence could spell the end of the human race"
– Steven Hawking

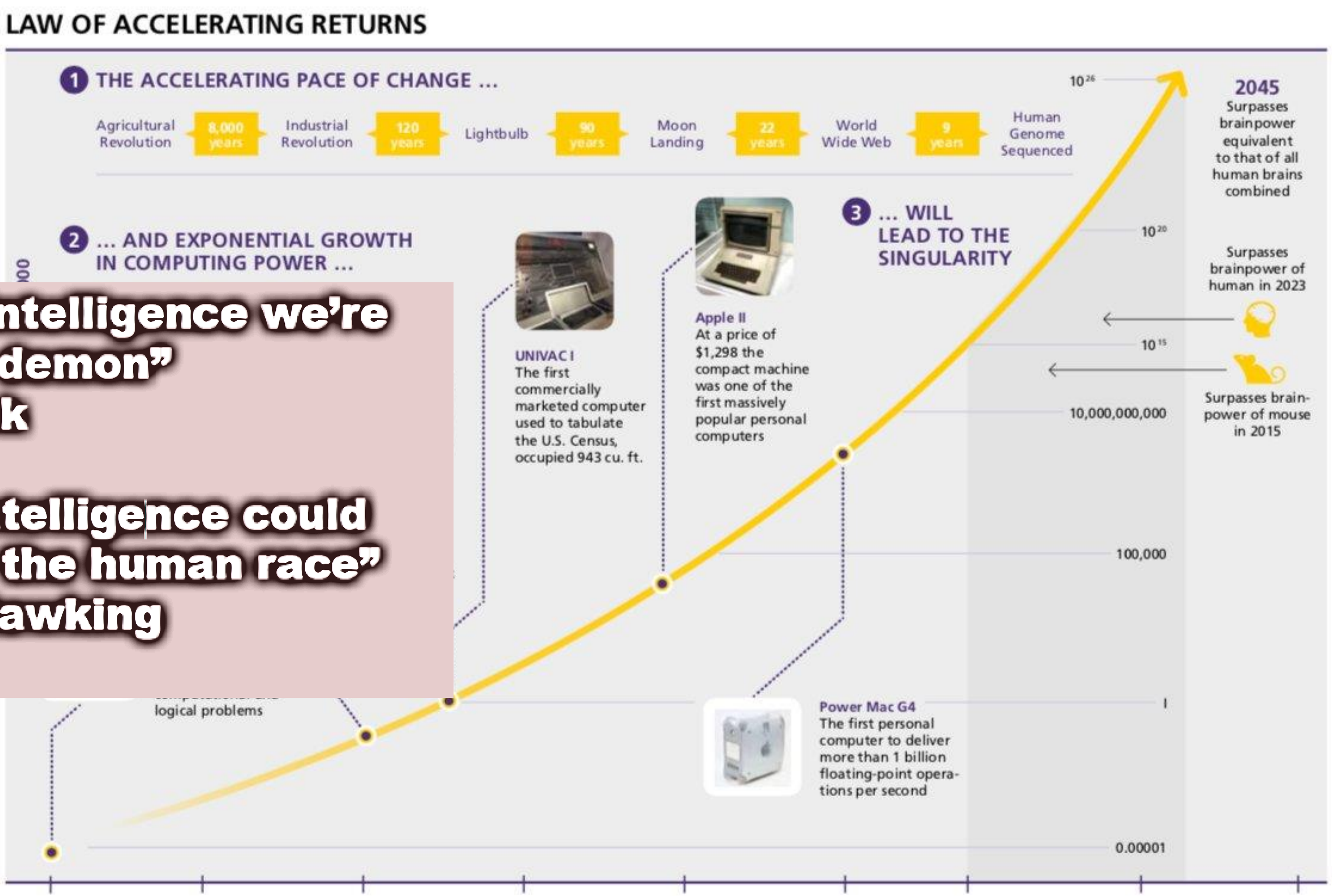
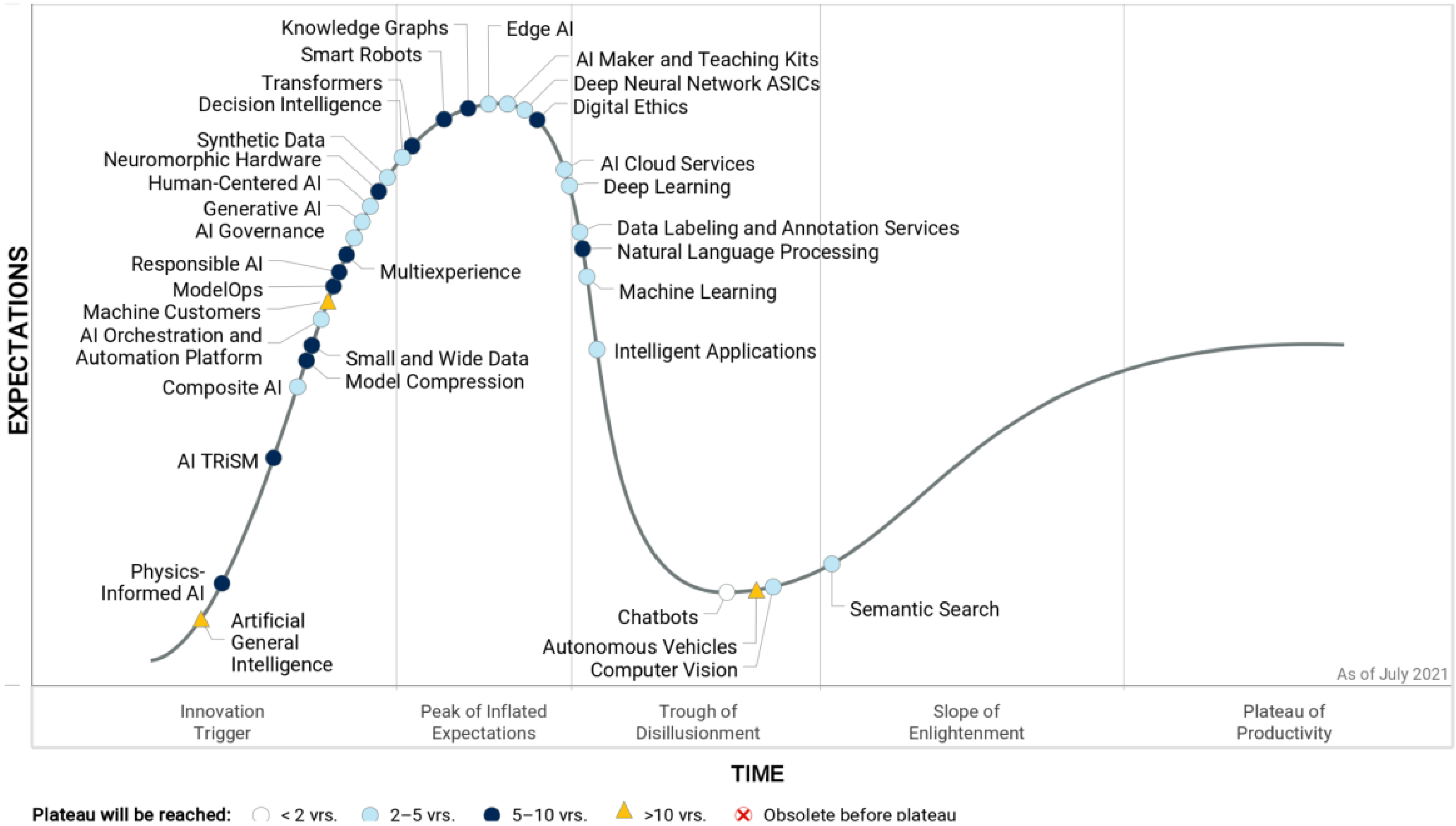


Figure 12: Ray Kurzweil's Law of Accelerating Returns depicts the exponential growth of computer processing power and technology innovations throughout history, and anticipates computers will exceed human intelligence in the future; **Source:** TIME / Wikipedia

source dhl via @mikequindazzi

HYPE CYCLE



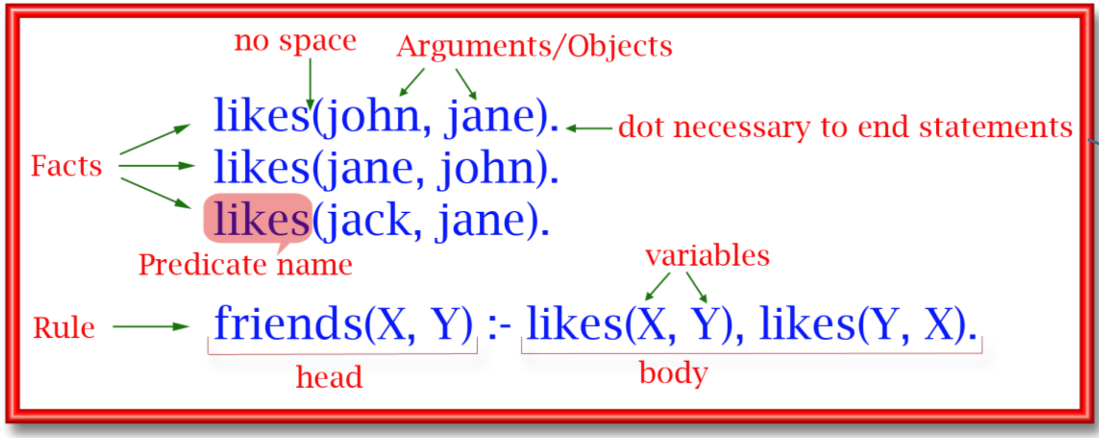
THE UNIVERSITY of EDINBURGH
School of Engineering

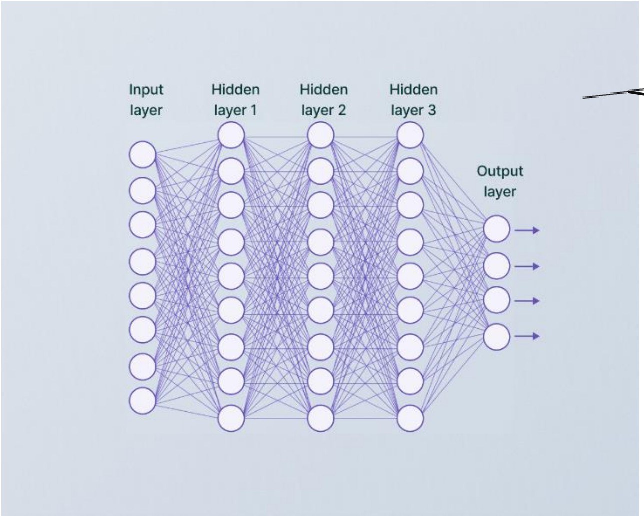
Gartner

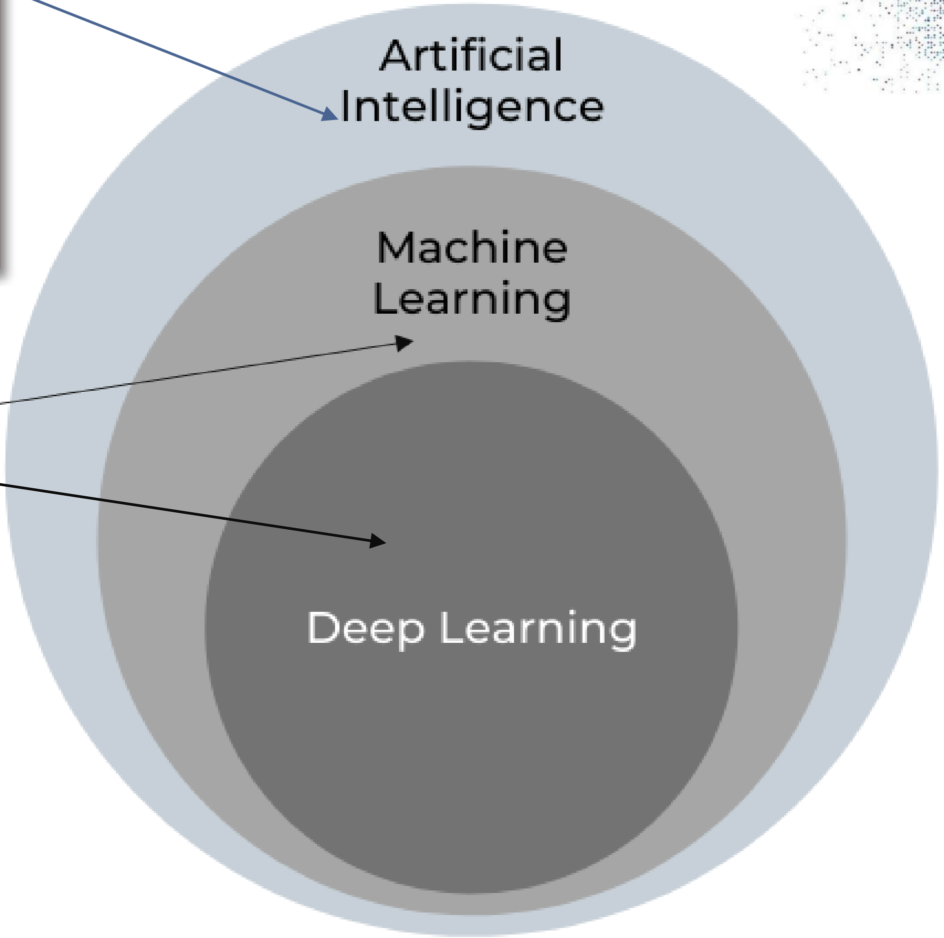
DEEP LEARNING VS. MACHINE LEARNING

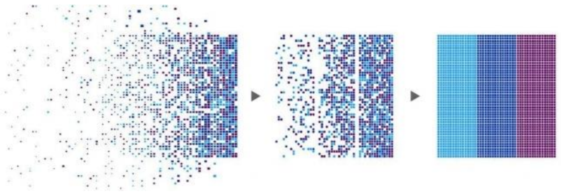
A broad range of technologies (e.g.) Prolog

Program Window

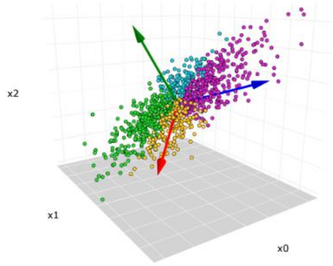




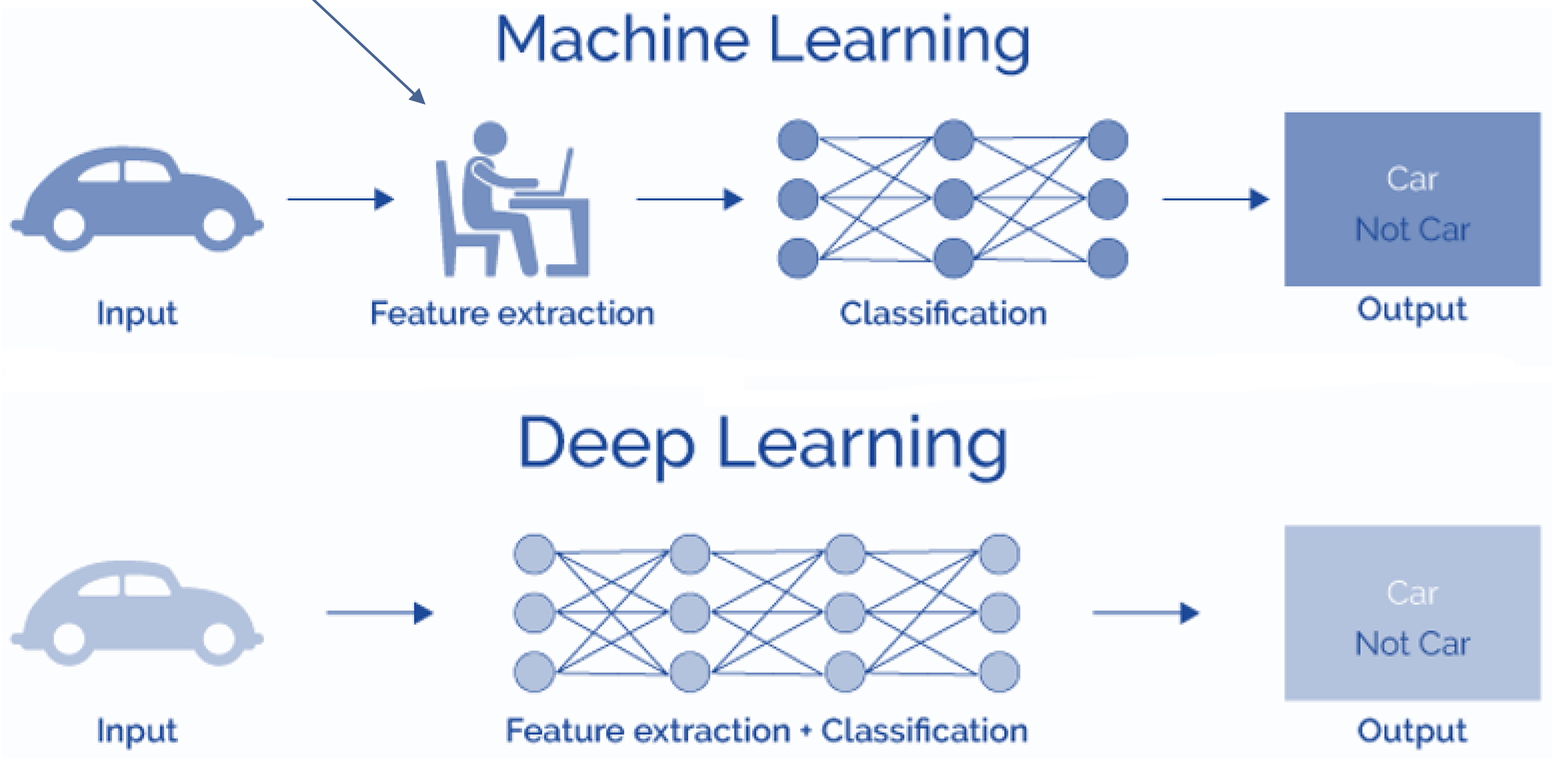




A range of pattern recognition and clustering technologies (lots of maths)!

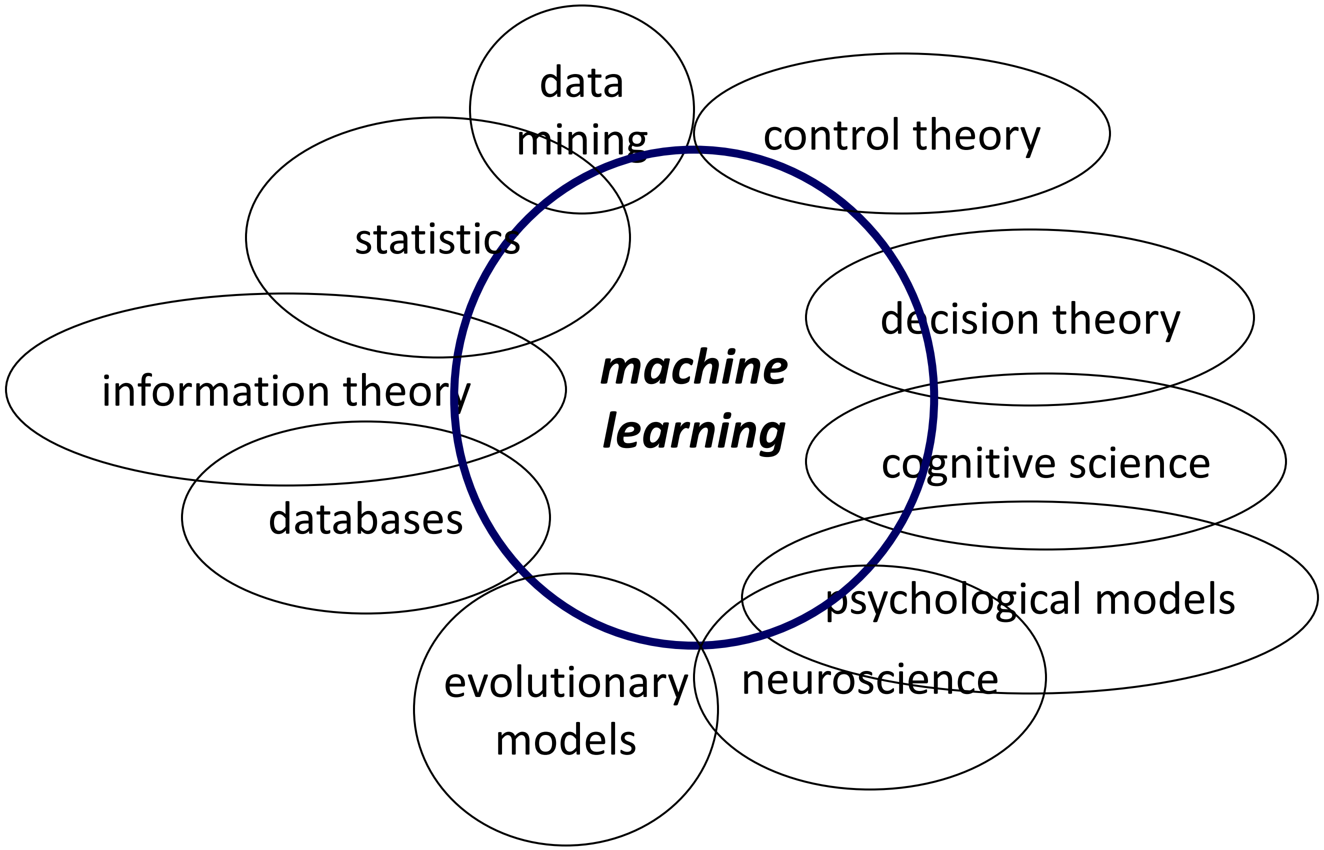



Hand bag
Shopping Bags
Shopping Bag
Shopping Bags
Shopping Bag
Shopping Bag
Shopping Bag


Machine Learning
Input
Feature extraction
Classification
Car
Not Car
Output
Deep Learning
Input
Feature extraction + Classification
Car
Not Car
Output

Machine Learning (ML)

OVERVIEW: RELATED FIELDS

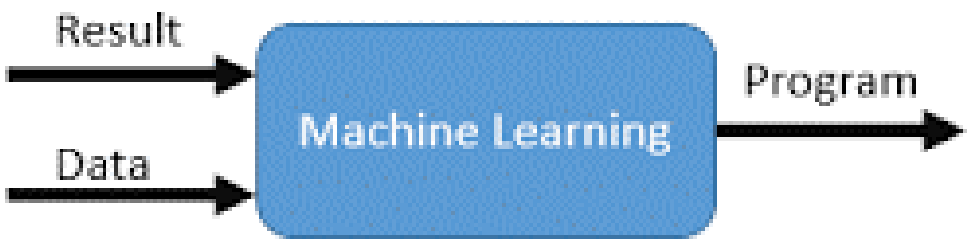


MACHINE LEARNING (ML)

- What is machine learning?
- What is supervised, unsupervised and reinforcement learning?
- What is dimensionality reduction and PCA?
- Top prediction algorithms

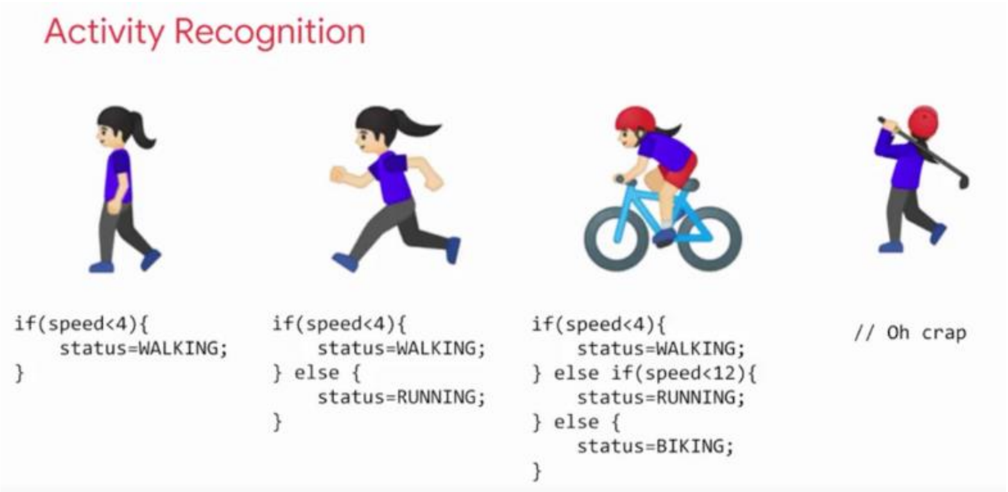
Regardless of the Hype it's a new way of programming!

Conventional Programming is writing a program in a traditional procedural language, such as assembly language or a high-level compiler language (C, C++, Java, JavaScript, Python, etc).

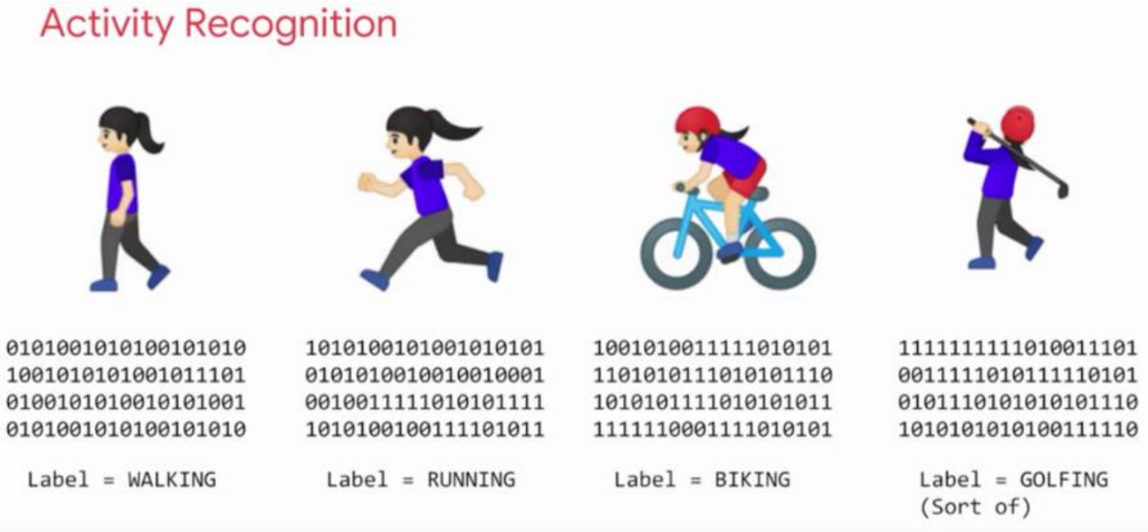


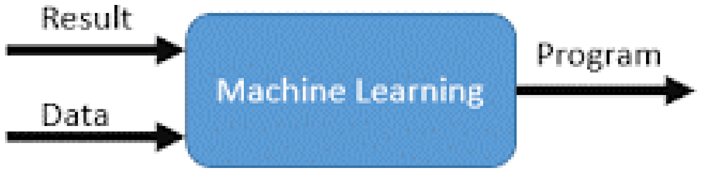
Machine Learning solves this problem by modeling this data with **train data** and **test data** and then ***predict*** the result.

TEACHABLE MACHINES



Conventional Programming

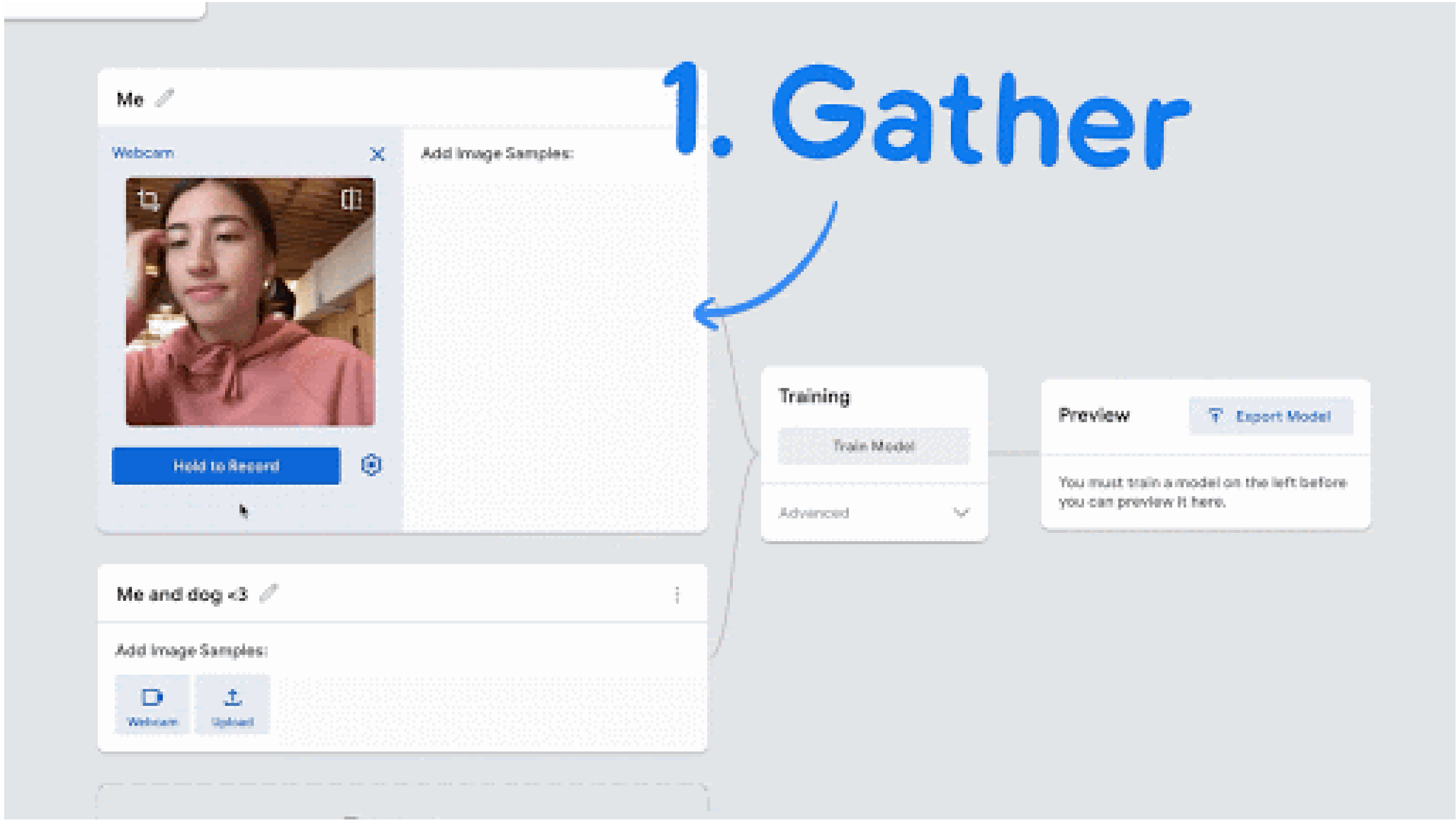




Machine Learning

THE UNIVERSITY of EDINBURGH
School of Engineering

https://teachablemachine.withgoogle.com/



Supervised Learning

THE UNIVERSITY of EDINBURGH
School of Engineering

SUPERVISED VS. UNSUPERVISED LEARNING

Supervised learning: classification is supervised learning from examples.

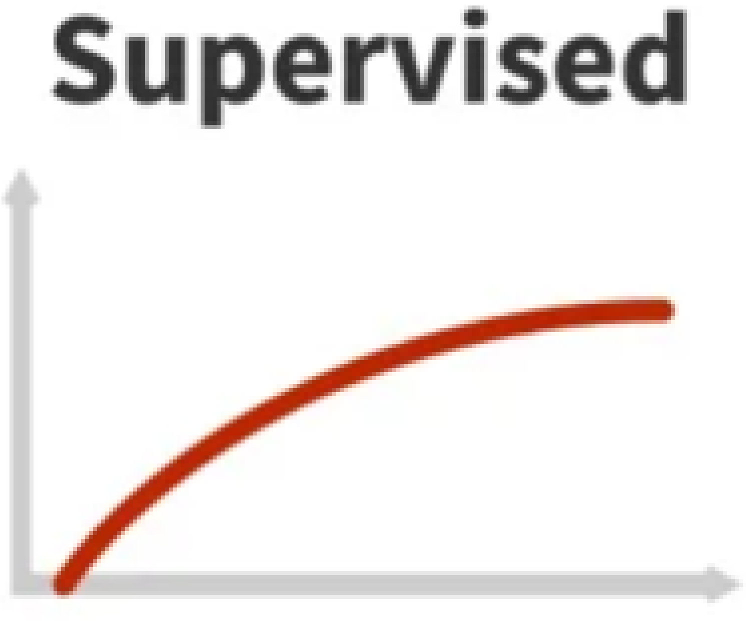
- Supervision: The data (observations, measurements, etc.) are labeled with pre-defined classes, which is
- like a “teacher” gives us the classes (supervision).

Unsupervised learning (clustering)

- Class labels of the data are not given or unknown
- **Goal**: Given a set of data, the task is to establish the existence of classes or clusters in the data

Source: Bing Liu, UIC

WHAT IS SUPERVISED LEARNING?



Source: Ben Freundorfer

Doug Rose defines supervised learning as "When a data scientist acts like a tutor for the machine, training it by showing it basic rules and giving it an overall strategy." [5]

- **Regression model**
- **Classification model**

SUPERVISED MACHINE LEARNING

- We humans learn from past experiences.
- A computer does not "experience."
  - **A computer system learns from data, which** represents "past experiences" in an application domain.
- Our focus: learn **a target function** that can be used to predict the values (**labels**) of a discrete class attribute, e.g.,
  - **high-risk** or **low risk** and **approved** or **not-approved**.
- The task is commonly called: **supervised learning, classification,** or **inductive learning**.

Source: Bing Liu, UIC

EXAMPLE APPLICATION

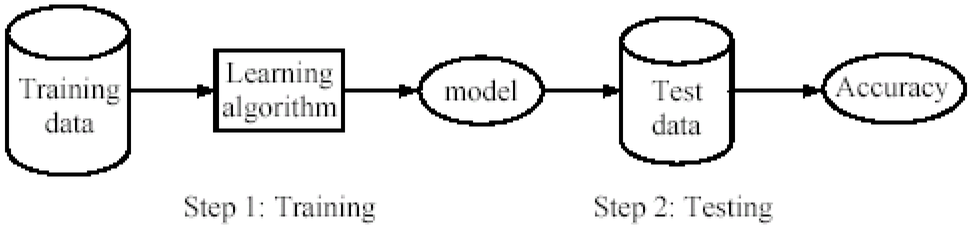
- A credit card company receives thousands of applications for new cards. Each application contains information about an applicant,
  - age
  - annual salary
  - outstanding debts
  - credit rating
  - etc.
- <u>Problem</u>: Decide whether an application should approved, i.e., **classify** applications into two categories, **approved** and **not approved**.

Source: Bing Liu, UIC

SUPERVISED LEARNING PROCESS: TWO STEPS

- Learning or training: Learn a model using the training data (with labels)
- Testing: Test the model using unseen test data (without labels) to assess the model accuracy

$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$

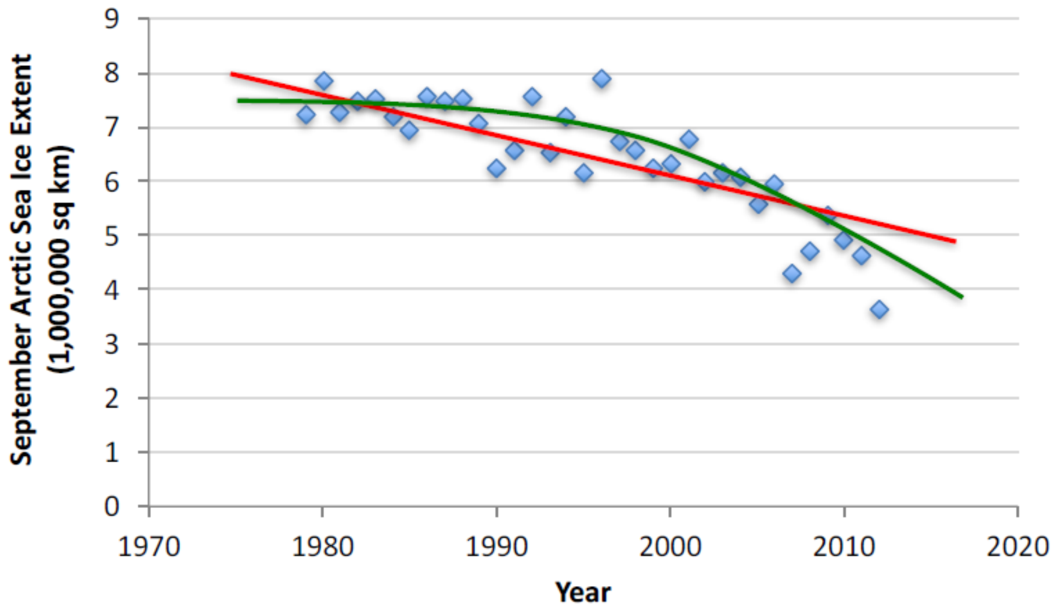


Source: Bing Liu, UIC

A TOUR OF ML TECHNOLOGIES

REGRESSION MODEL

- Given (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression

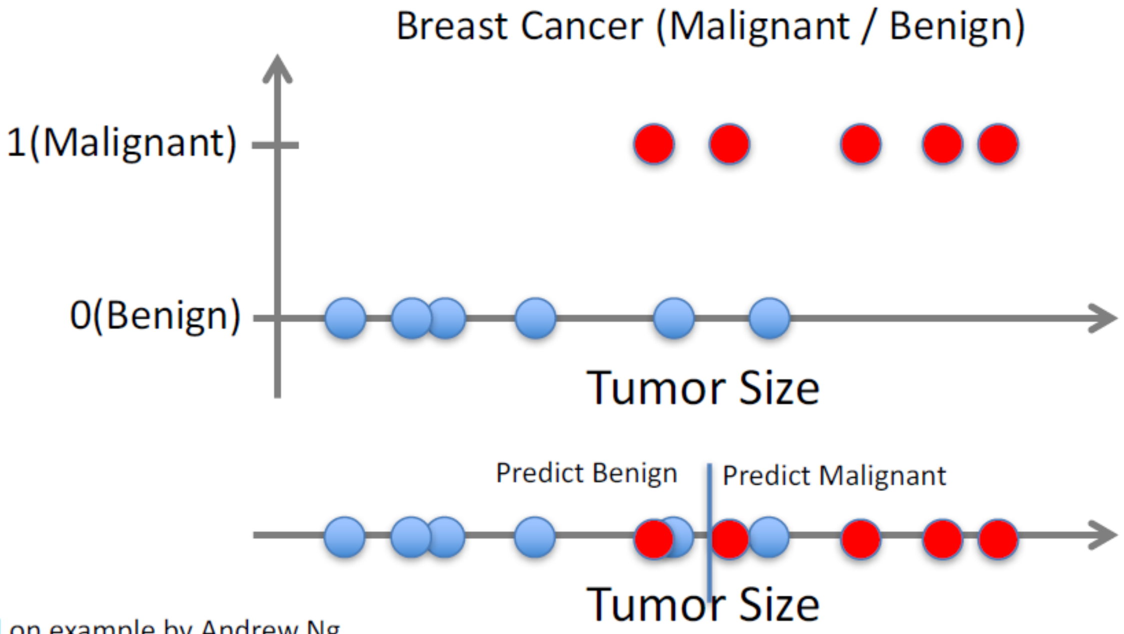


Data from G. Witt. Journal of Statistics Education, Volume 21, Number 1 (2013)

Source: Eric Eaton

CLASSIFICATION MODEL

- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification

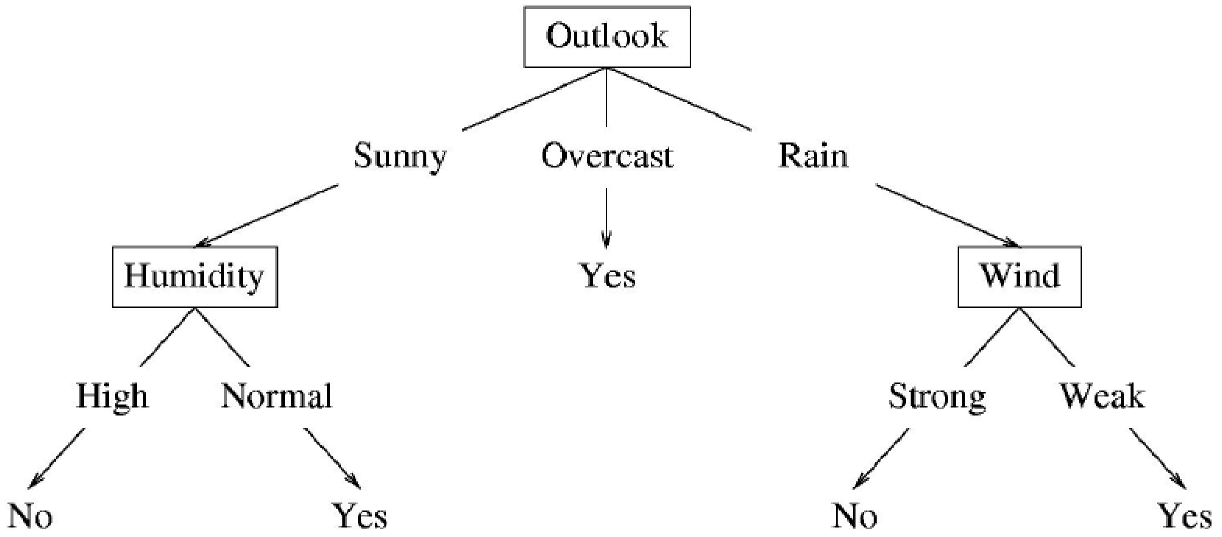


Based on example by Andrew Ng

Source: Eric Eaton

DECISION TREE

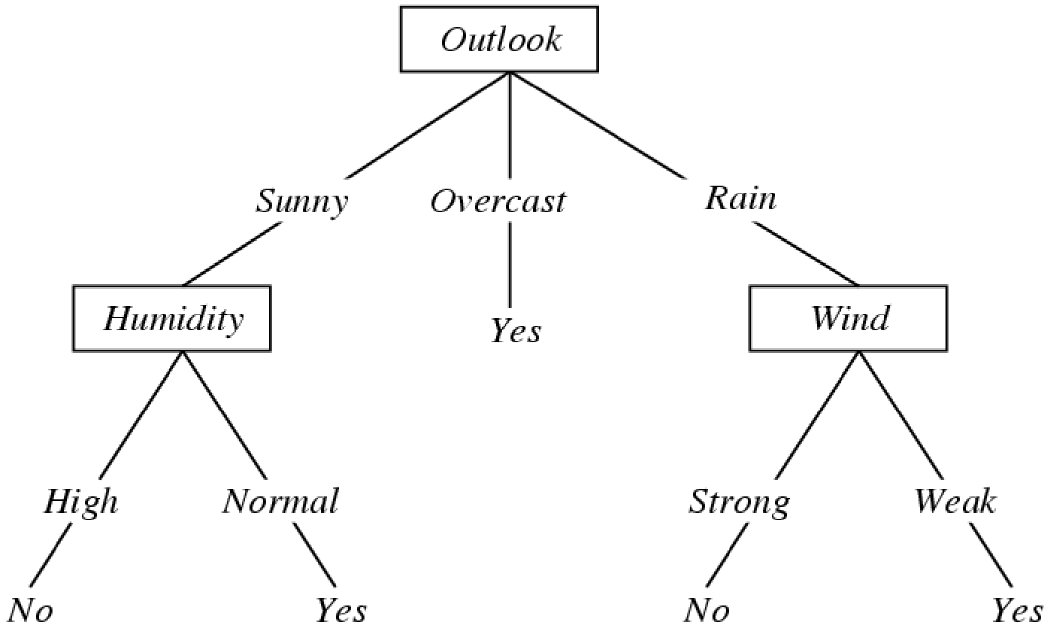
- A possible decision tree for the data:



- Each internal node: test one attribute X_i
- Each branch from a node: selects one value for X_i
- Each leaf node: predict Y (or $p(Y \mid \boldsymbol{x} \in \text{leaf})$)

DECISION TREE LEARNING

Decision Tree for *PlayTennis*

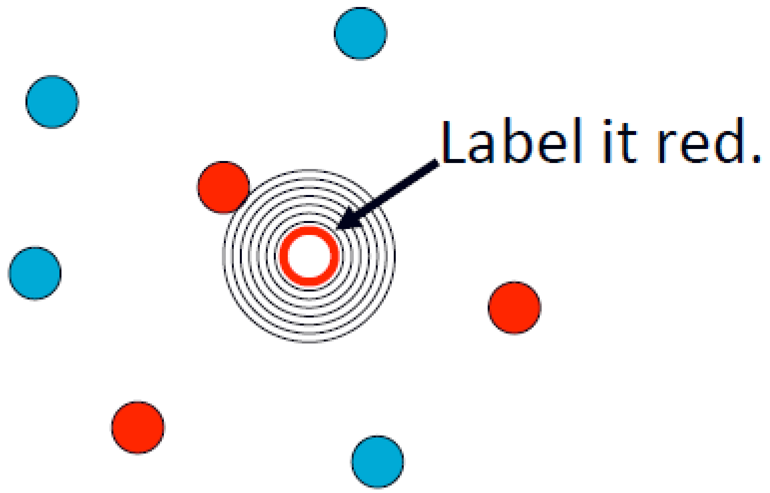


Problem Setting:

- Set of possible instances *X*
  - each instance *x* in *X* is a feature vector
  - e.g., <*Humidity=low, Wind=weak, Outlook=rain, Temp=hot*>
- Unknown target function *f* : *X*➡*Y*
  - *Y* is discrete valued
- Set of function hypotheses *H*={ *h* | *h* : *X*➡*Y* }
- each hypothesis *h* is a decision tree
- trees sorts *x* to leaf, which assigns *y*

K-NEAREST NEIGHBOUR

- **1-Nearest Neighbour**
- One of the simplest of all machine learning classifiers
- Simple idea: label a new point the same as the closest known point



Source: Eric Eaton

K-NEAREST NEIGHBOUR

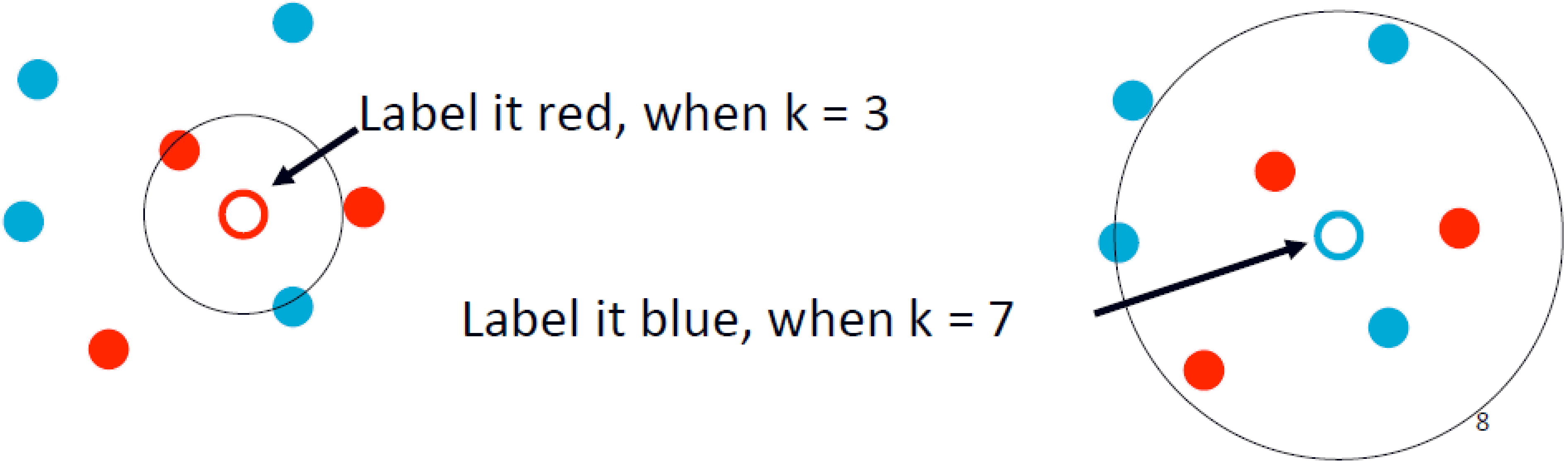
- **Four Aspects of an Instance---Based Learner:**
- 1. A distance metric
- 2. How many nearby neighbours to look at?
- 3. A weighting function (optional)
- 4. How to fit with the local points?

Adapted from “Instance---Based Learning” lecture slides by Andrew Moore, CMU.

Source: Eric Eaton

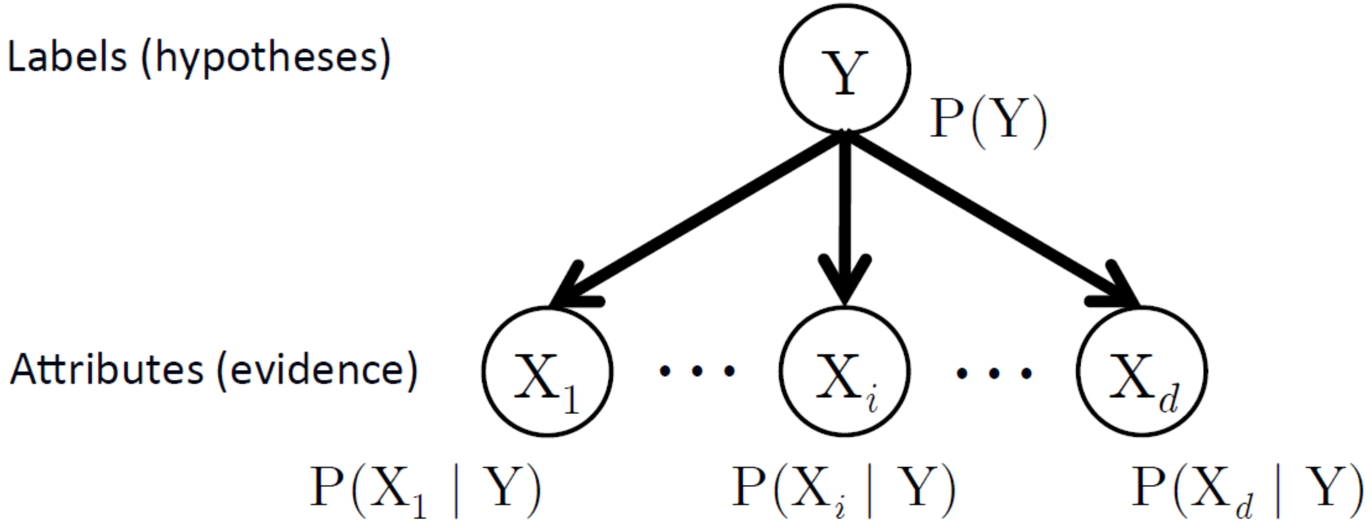
K-NEAREST NEIGHBOUR

- Generalizes 1---NN to smooth away noise in the labels
- A new point is now assigned the most frequent label of its *k* nearest neighbours



Source: Eric Eaton

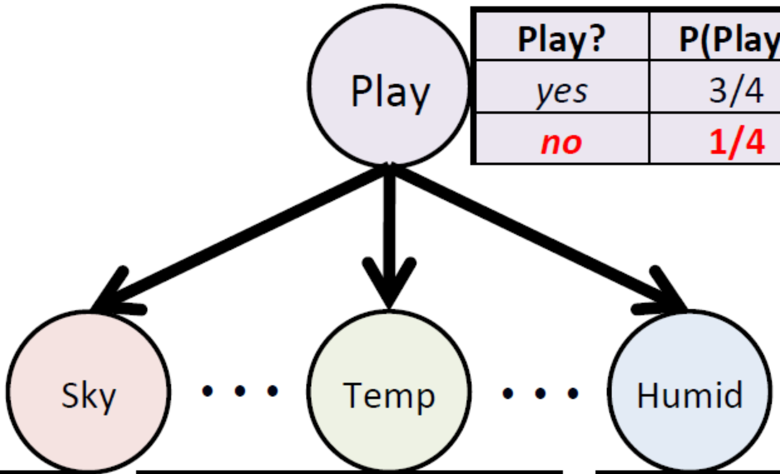
The Naïve Bayes Graphical Model



- Nodes denote random variables
- Edges denote dependency
- Each node has an associated conditional probability table (CPT), conditioned upon its parents

Source: Eric Eaton

Example NB Graphical Model



Play?	P(Play)
yes	3/4
no	**1/4**

Sky	Play?	P(Sky | Play)
sunny	*yes*	4/5
rainy	***yes***	**1/5**
sunny	*no*	1/3
rainy	***no***	**2/3**

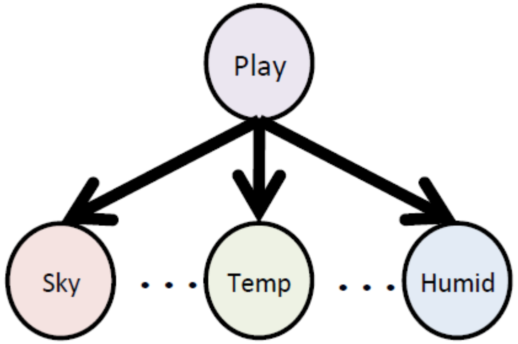
Temp	Play?	P(Temp | Play)
warm	*yes*	4/5
cold	***yes***	**1/5**
warm	*no*	1/3
cold	***no***	**2/3**

Humid	Play?	P(Humid | Play)
high	*yes*	3/5
norm	***yes***	**2/5**
high	*no*	2/3
norm	***no***	**1/3**

- Some **redundancies** in CPTs that can be eliminated

Source: Eric Eaton

Example Using NB for Classification



Play?	P(Play)
yes	3/4
no	1/4

Temp	Play?	P(Temp | Play)
warm	*yes*	4/5
cold	*yes*	1/5
warm	*no*	1/3
cold	*no*	2/3

Sky	Play?	P(Sky | Play)
sunny	*yes*	4/5
rainy	*yes*	1/5
sunny	*no*	1/3
rainy	*no*	2/3

Humid	Play?	P(Humid | Play)
high	*yes*	3/5
norm	*yes*	2/5
high	*no*	2/3
norm	*no*	1/3

$$h(\mathbf{x}) = \arg\max_{y_k} \; \log P(Y = y_k) + \sum_{j=1}^{d} \log P(X_j = x_j \mid Y = y_k)$$

Goal: Predict label for $\mathbf{x}$ = (rainy, warm, normal)

NAIVE BAYES

- **Advantages:**
  - Fast to train (single scan through data)
  - Fast to classify
  - Not sensitive to irrelevant features
  - Handles real and discrete data
  - Handles streaming data well
- **Disadvantages:**
  - Assumes independence of features

WHAT IS UNSUPERVISED LEARNING?

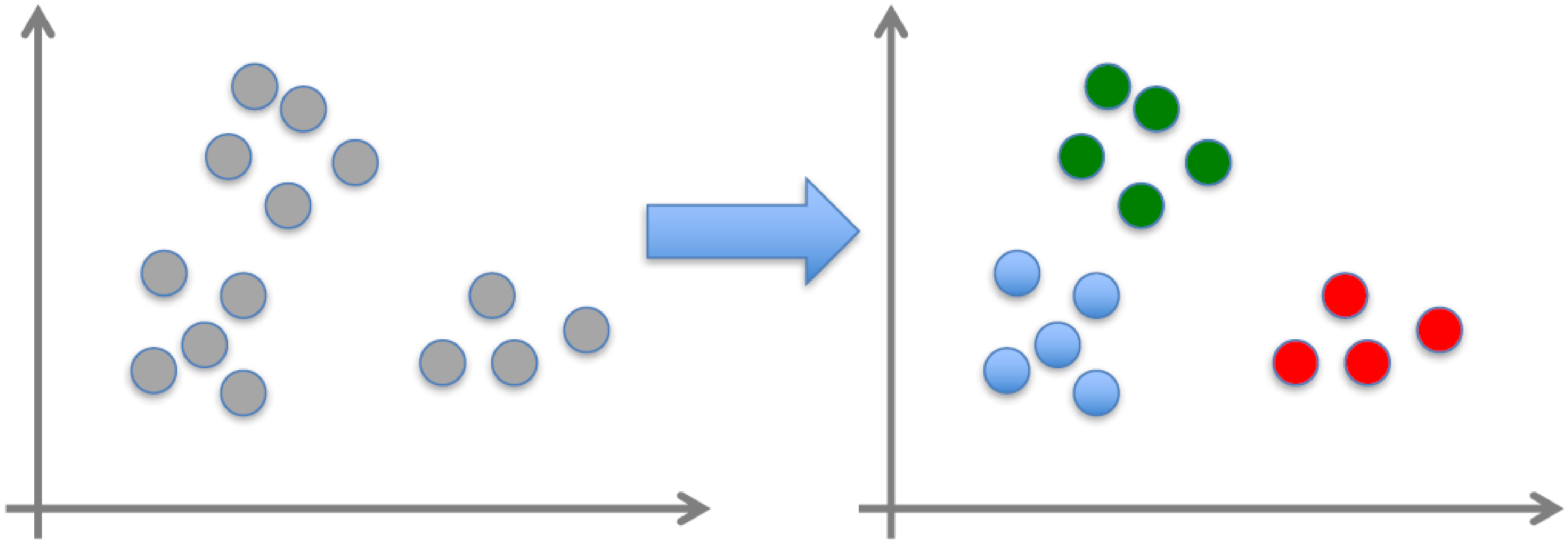
Doug Rose defines unsupervised learning as “The machine makes all the observations on its own. It might now know all the different names and labels, but it will find patterns on its own.” [5]

- Clustering
- Anomaly detection
- Association mining
- Latent variable models

Source: Ben Freundorfer

UNSUPERVISED LEARNING

- Given x_1, x_2, ..., x_n (without labels)
- Output hidden structure behind the x's
  - E.g., clustering



Source: Eric Eaton

CLUSTERING

Clustering is one main approach to unsupervised learning.

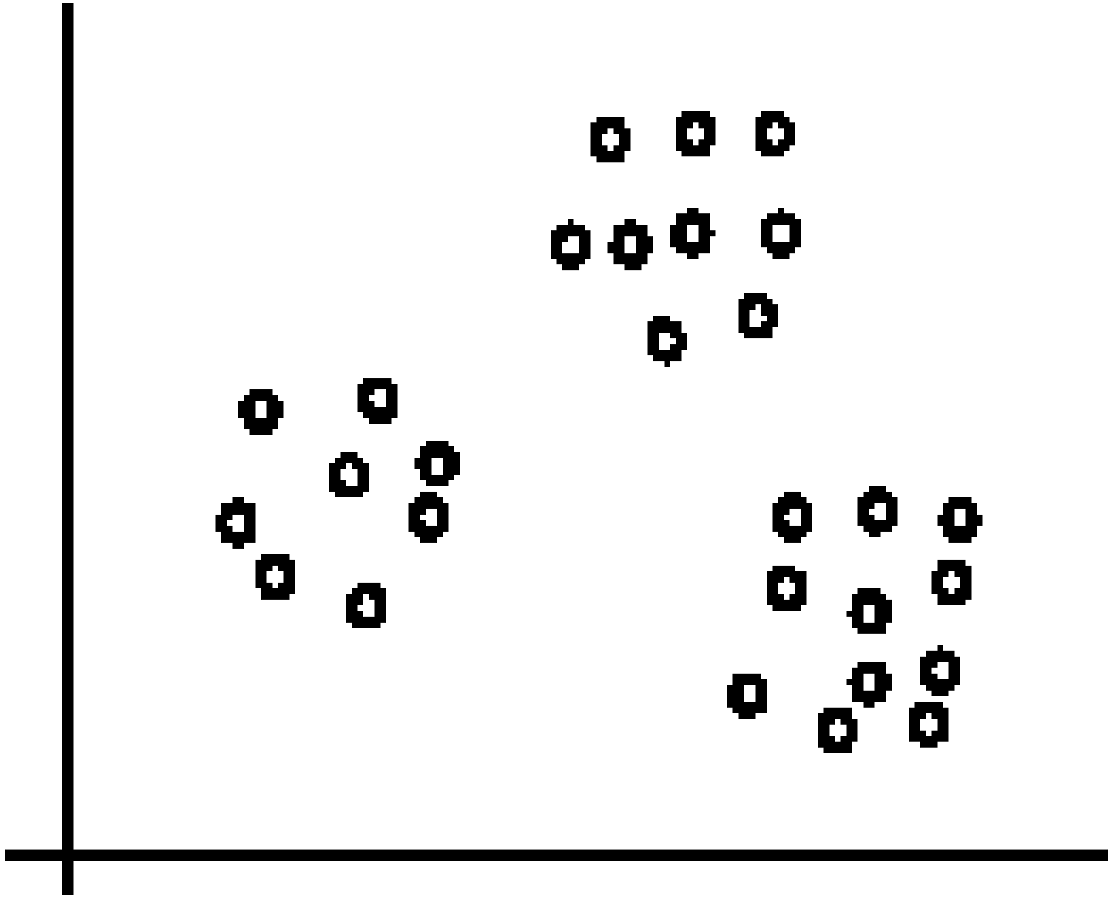
- It finds similarity groups in data, called clusters,
  - it groups data instances that are similar to (near) each other in one cluster and data instances that are very different (far away) from each other into different clusters.

Clustering is often considered synonymous with unsupervised learning.

- But, association rule mining is also unsupervised

This chapter focuses on clustering.

The data set has three natural groups of data points, i.e., 3 natural clusters.



Source: Bing Liu, UIC

WHAT IS CLUSTERING FOR?

Example: Given a collection of text documents, we want to organize them according to their content similarities,

- To produce a topic hierarchy

In fact, clustering is one of the most utilized data mining techniques.

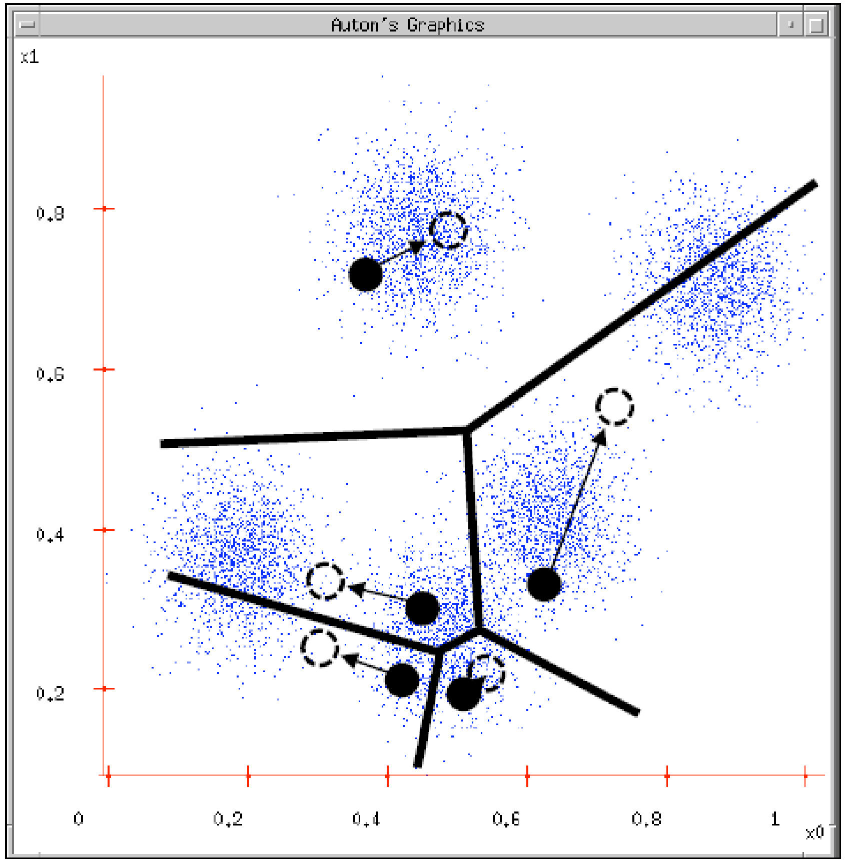
- It has a long history, and been used in almost every field, e.g., medicine, psychology, botany, sociology, biology, archeology, marketing, insurance, libraries, etc.
- In recent years, due to the rapid increase of online documents, text clustering becomes important.

Source: Bing Liu, UIC

K-MEAN CLUSTERING

K-Means (k , X)

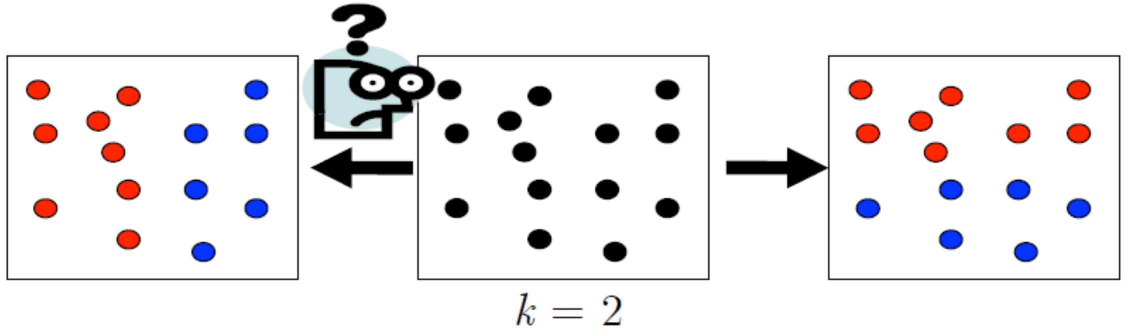
- Randomly choose k cluster center locations (centroids)
- Loop until convergence
  - Assign each point to the cluster of the closest centroid
  - Re-estimate the cluster centroids based on the data assigned to each cluster



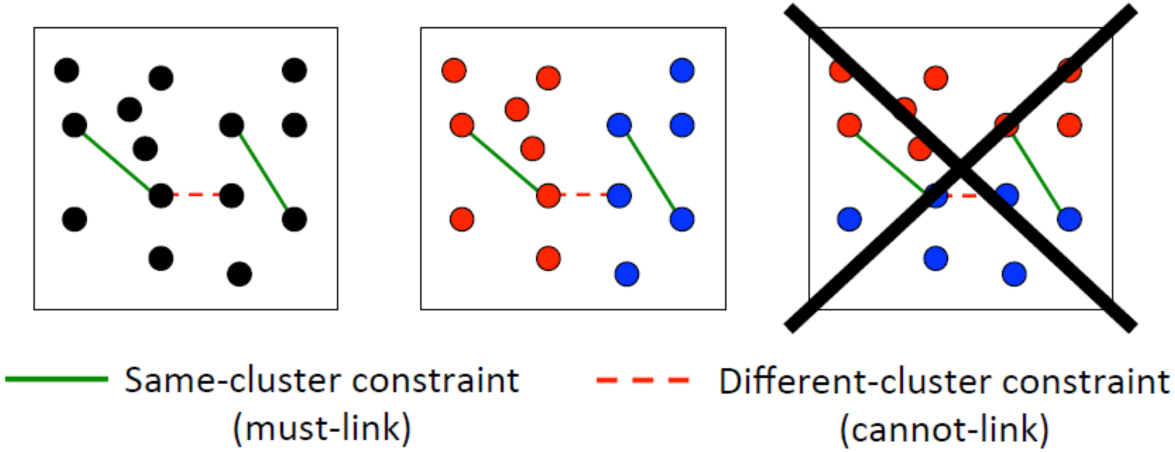
Source: Eric Eaton

PROBLEMS WITH K-MEANS

- How do you tell it which clustering you want?



Constrained clustering techniques (semi-supervised)



Source: Eric Eaton

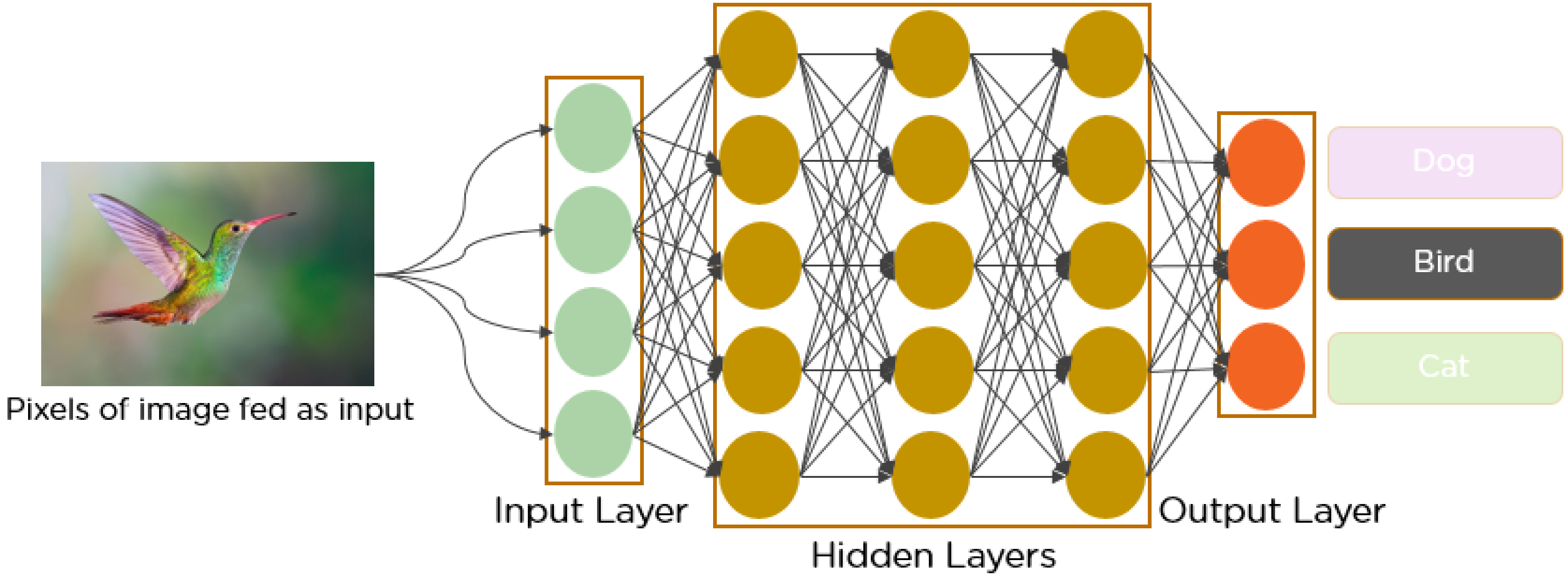
NEURAL NETWORKS

Neural networks refer to a biological phenomenon comprised of interconnected neurons that exchange messages with each other. This idea has now been adapted to the world of machine learning and is called ANN (Artificial Neural Networks).

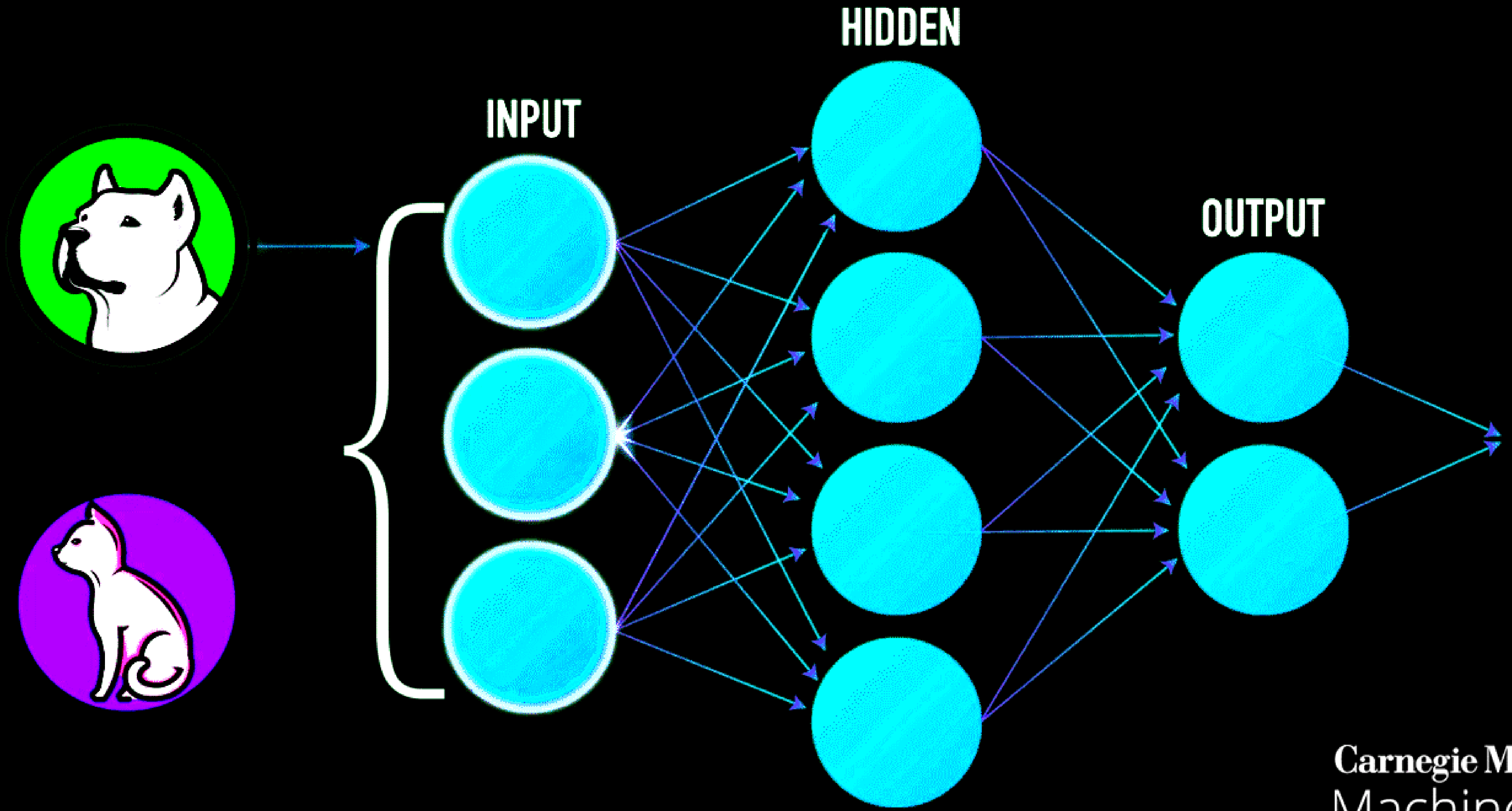
Deep learning, which you've heard a lot about, can be done with several layers of neural networks put one after the other. ANNs are a family of models that are taught to adopt cognitive skills.

Source: Dataiku

MACHINE LEARNING CLASSIFICATION NETWORK

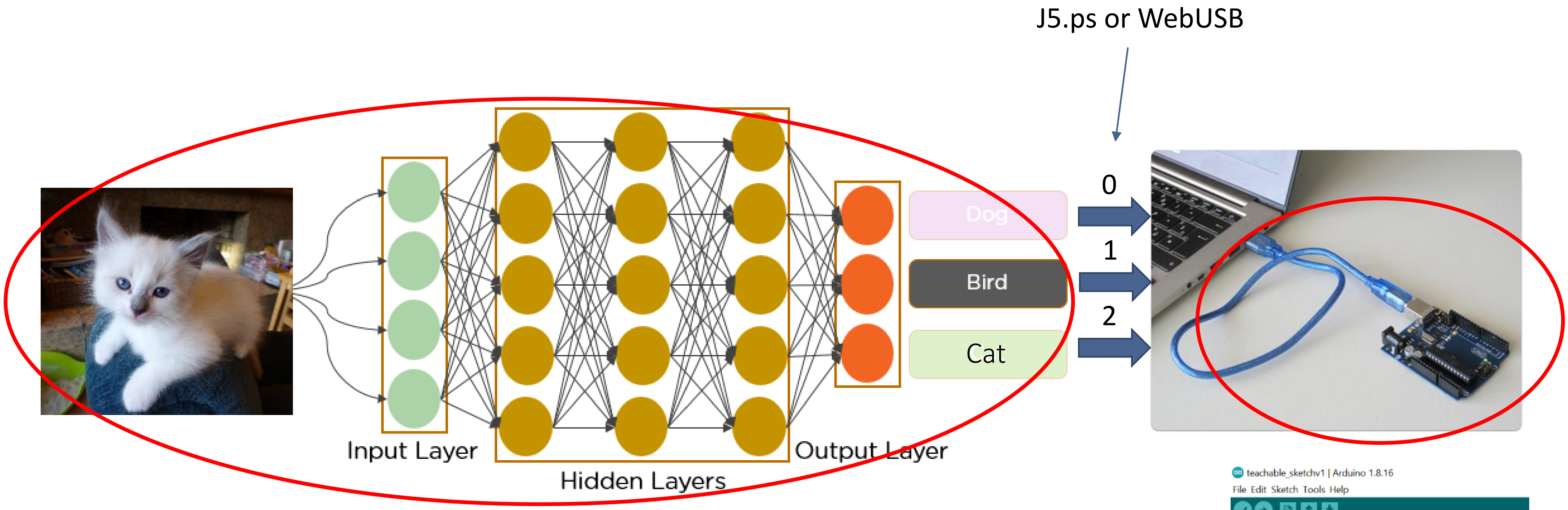


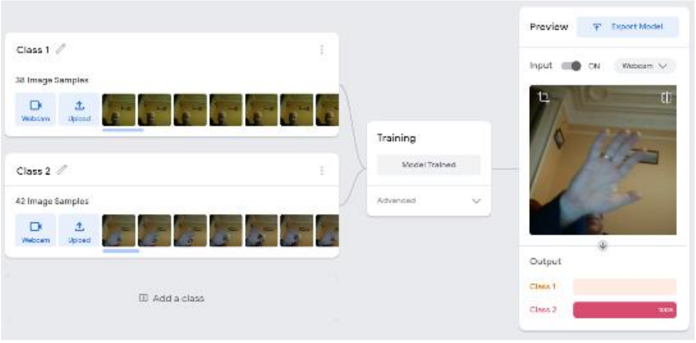
The training process assigns weights to the arcs of the network so it output the right classifications



Carnegie Mellon University
Machine Learning

TEACHABLE MACHINE SCHEMATIC





Teachable Machine Website

Arduino IDE

MACHINE LEARNING PROCESS IN PYTHON

- Load dataset
- Summary statistics
- Visualisation
- Algorithm evaluation
- Predictive models

TOP PREDICTION ALGORITHMS

	TYPE	NAME	DESCRIPTION	ADVANTAGES	DISADVANTAGES
Linear		**Linear Regression**	**The "best fit" line** through all data points. Predictions are numerical.	Easy to understand — you clearly see what the biggest drivers of the model are.	Sometimes too simple to capture complex relationships between variables. Does poorly with correlated features.
Linear		**Logistic Regression**	The adaptation of **linear regression** to problems of classification (e.g., yes/no questions, groups, etc.	Also easy to understand.	Sometimes too simple to capture complex relationships between variables. Does poorly with correlated features.

TOP PREDICTION ALGORITHMS

Tree-Based		**Decision Tree**	**A series of yes/no rules based on the features,** forming a tree, to match all possible outcomes of a decision.	Easy to understand.	Not often used on its own for prediction because it's also often too simple and not powerful enough for complex data.
		Random Forest	Takes advantage of many decision trees, with rules created from subsamples of features. Each tree is weaker than a full decision tree, but **by combining them we get better overall performance.**	A sort of "wisdom of the crowd". Tends to result in very high quality models. Fast to train.	Models can get very large. Not easy to understand predictions.
		Gradient Boosting	Uses even weaker decision trees, that are increasingly **focused on "hard" examples.**	High-performing.	A small change in the feature set or training set can create radical changes in the model. Not easy to understand predictions.

Source: data iku

TOP PREDICTION ALGORITHMS

Neural Networks	**Neural Networks**	Interconnected "neurons" that pass messages to each other. Deep learning uses several **layers of neural networks stacked on top of one another.**	Can handle extremely complex tasks — no other algorithm comes close in image recognition.	Very slow to train, because they often have a very complex architecture. Almost impossible to understand predictions.

Source: data iku

TRAINING VS. TEST DISTRIBUTION

- We generally assume that the training and test examples are independently drawn from the same overall distribution of data
  - We call this “i.i.d” which stands for “independent and identically distributed”
- If examples are not independent, requires ***collective classification***
- If test distribution is different, requires ***transfer learning***

Slide credit: Ray Mooney

EVALUATING MODELS

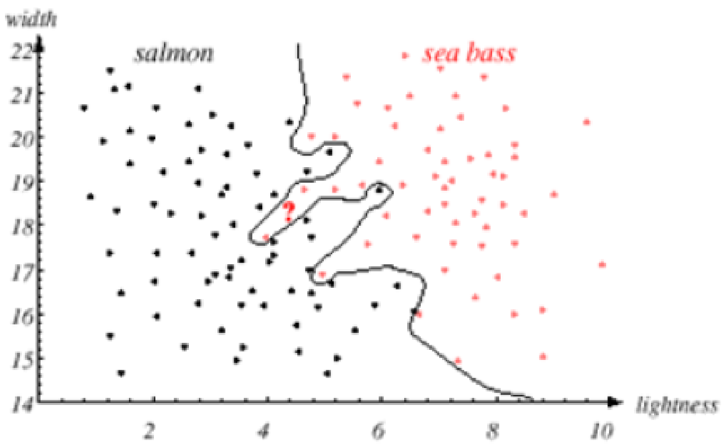
- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- etc.

Source: Pedro Domingos

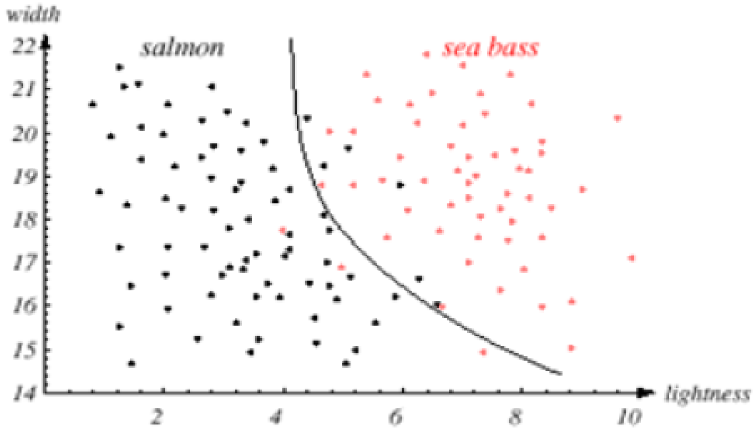
EVALUATING MODELS

Source: Ben Freundorfer

OVERFITTING AND REGULARISATION



Overly complex models lead to complicated decision boundaries. It leads to perfect classification on the training examples, but would lead to poor performance on new examples.



The decision boundary might represent the optimal tradeoff between performance on the training set and simplicity of classifier, therefore giving highest accuracy on new examples.

(Duda et al. Pattern Classification 2000)

AVOIDING OVERFITTING

How can we avoid overfitting?

- Stop growing when data split is not statistically significant
- Acquire more training data
- Remove irrelevant attributes (manual process not always possible)
- Grow full tree, then post-prune
- How to select "best" tree:
  - Measure performance over training data
  - Measure performance over separate validation data set
  - Add complexity penalty to performance measure

Based on Slide by Pedro Domingos

REFERENCES

1. Zhydik, O., 2021. *Six Powerful Use Cases for Machine Learning in Manufacturing*. [online] Eleks.com. Available at: https://eleks.com/blog/machine-learning-in-manufacturing/
2. Green, J., 2020. *5 ways you can use Machine Learning in manufacturing*. [online] Ancoris.com. Available at: https://www.ancoris.com/blog/5-ways-machine-learning-manufacturing
3. Doty, C., 2021. *10 Ways Machine Learning Will Reshape Manufacturing in 2021 - RapidMiner*. [online] RapidMiner. Available at: https://rapidminer.com/blog/10-ways-machine-learning-in-manufacturing-2021/
4. Rai, R., Tiwari, M., Ivanov, D. and Dolgui, A., 2021. Machine learning in manufacturing and industry 4.0 applications. *International Journal of Production Research*, 59(16), pp.4773-4778.
5. GLOSSARY, Artificial Intelligence Foundations: Machine Learning, LinkedinLearn
6. McCarthy, J., 2004. WHAT IS ARTIFICIAL INTELLIGENCE?. Computer Science Department Stanford University Stanford.

RESOURCES

AAAI. Machine Learning. *http://www.aaai.org/Pathfinder/html/machine.html*

Dietterich, T. (2003). Machine Learning. *Nature Encyclopedia of Cognitive Science.*

Doyle, P. Machine Learning. *http://www.cs.dartmouth.edu/~brd/Teaching/AI/Lectures/Summaries/learning.html*

Dyer, C. (2004). Machine Learning. *http://www.cs.wisc.edu/~dyer/cs540/notes/learning.html*

Mitchell, T. (1997). *Machine Learning*.

Nilsson, N. (2004). Introduction to Machine Learning. *http://robotics.stanford.edu/people/nilsson/mlbook.html*

Russell, S. (1997). Machine Learning. *Handbook of Perception and Cognition*, Vol. 14, Chap. 4.

Russell, S. (2002). *Artificial Intelligence: A Modern Approach*, Chap. 18-20. *http://aima.cs.berkeley.edu*