Data, Inference & Applied Machine Learning

Course: 18-785

Patrick McSharry

patrick@mcsharry.net

www.mcsharry.net

Twitter: @patrickmcsharry

Fall 2022

ICT Center of Excellence
Carnegie Mellon University

Machine Learning

Applied ML

- The aim is to provide the expertise and skills necessary for applying machine learning techniques to large real-world datasets in order to facilitate knowledge discovery, predictive analytics and decision-support.
- A variety of sophisticated techniques for refining, visualizing, exploring and modelling data will be introduced and demonstrated.
- The advantages and disadvantages of linear, nonlinear, nonparametric and ensemble methods will be discussed while exploring the challenges of both supervised and unsupervised learning.

Text Books

- The Elements of Statistical Learning
- T. Hastie, R. Tibshirani, and J. Friedman,
- Springer-Verlag, 2001.

- Pattern Recognition and Machine Learning
- Christopher Bishop
- Springer, 2006.

Course outline

Week	Description
7	Statistical learning
8	Linear models
9	Nonlinear models
10	Supervised learning
11	Unsupervised learning
12	Ensemble approaches

Applied Machine Learning

WEEK 7A

Today's Lecture

No.	Activity	Description	Time
1	Challenge	Machine learning	10
2	Discussion	Learning	10
3	Case study	APGAR score	10
4	Analysis	Sitting rising test	20
5	Demo	Computer evolution	20
6	Q&A	Questions and answers	10

Learning

- Oxford English Dictionary definition:

"The action of receiving instruction or acquiring knowledge; *spec.* in *Psychol.*, a process which leads to the modification of behaviour or the acquisition of new abilities or responses, and which is additional to natural development by growth or maturation."

Machine Learning Context

- Which applications use machine learning?

- **Slido.com**
- **#54377**

Machine Learning

- Autonomous vehicles – Uber/Google
- Voice assistants: Alexa/Siri
- Recommendation systems – Netflix/Amazon
- Customer segmentation – retail

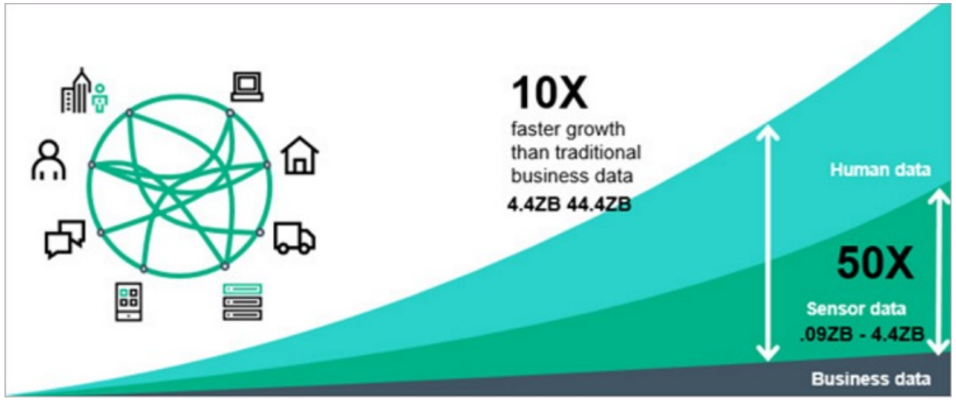
Machine Learning Applications

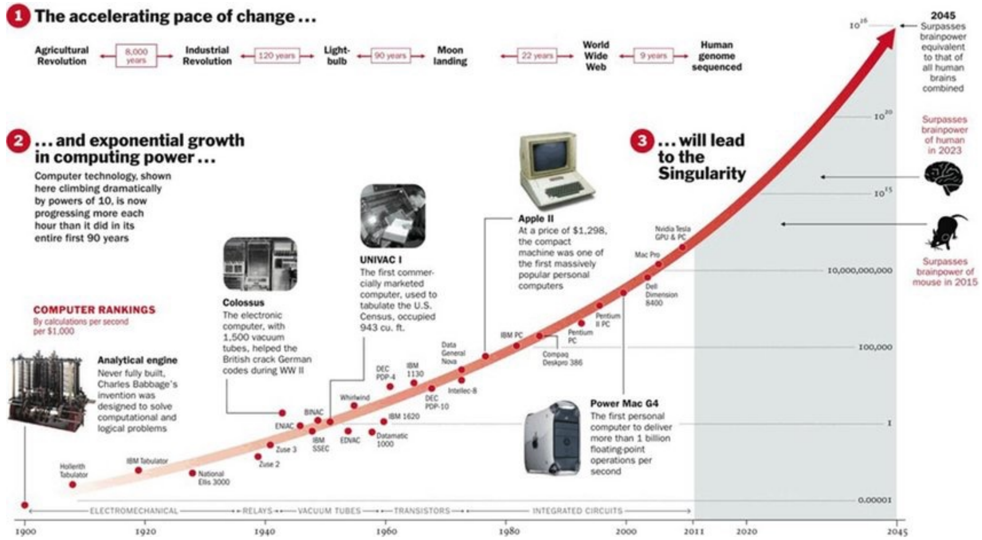
- Spam detection (Google)
- Fraud detection (Banks)
- Speech recognition (Apple's Siri, Amazon's Alexa)
- Driverless cars (Uber)
- Facial recognition (Intelligence agencies)
- Medical diagnosis (IBM)
- Data Mining (Nielson for Retail)
- Credit risk scoring – financial institutions
- Classification and Clustering of data
- Sentiment analysis of social media
- Five-min throat swab COVID tests (Oxford University)

Artificial Intelligence (AI)

- *"AI is the science and engineering of making intelligent machines, especially intelligent computer programs"* – John McCarthy.

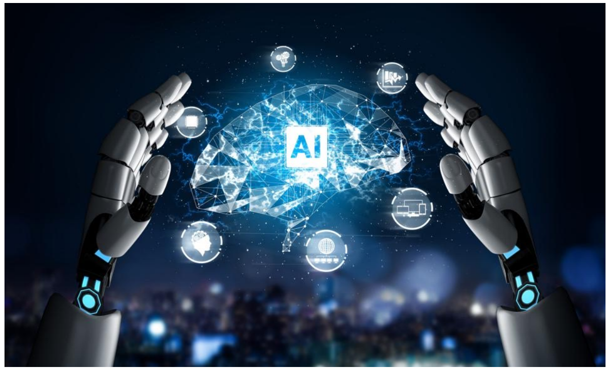
- AI is flourishing at present because of advances in computer power, availability of large amounts of digital information (big data, open data), and enhanced theoretical understanding.





AI is an Enabler of Innovation

- AI describes the ability of a machine to perform activities that normally require human cognition such as learning.



TESLA — Intelligent Products

amazon — Intelligent Services

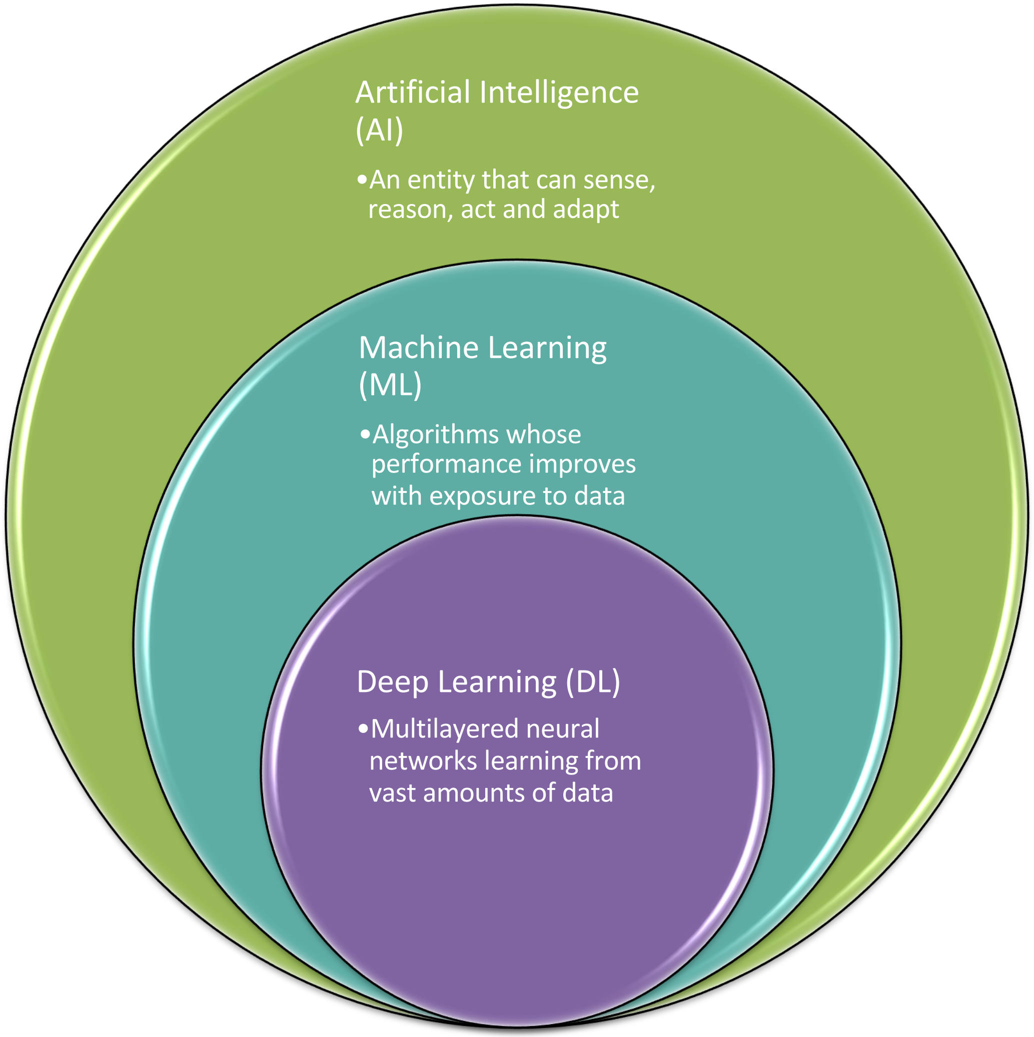
salesforce — Improved Processes

Hierarchy of techniques

- What is the hierarchy of these techniques:
  - Artificial intelligence (AI)
  - Machine Learning (ML)
  - Deep Learning (DL)
- What is the hierachy of these techniques?

1. DL > AI > ML
2. DL > ML > AI
3. ML > DL > AI
4. ML > AI > DL
5. AI > ML > DL
6. AI > DL > ML

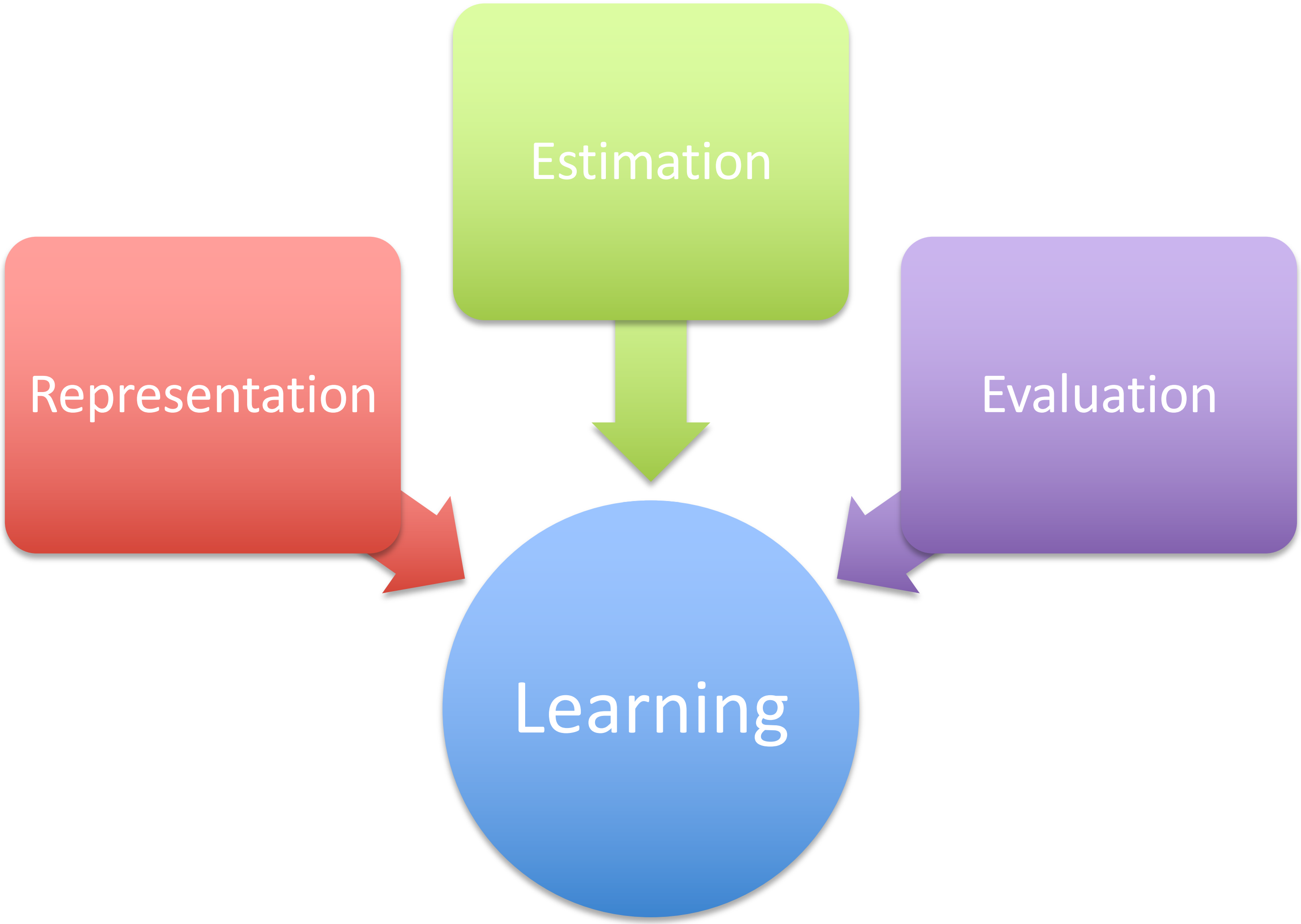
AI, ML and DL


Artificial Intelligence (AI)
•An entity that can sense, reason, act and adapt
Machine Learning (ML)
•Algorithms whose performance improves with exposure to data
Deep Learning (DL)
•Multilayered neural networks learning from vast amounts of data

Qualitative to Quantitative

- In order to make decisions and take definitive action, we often need to transform qualitative information into numerical values.
- The numerical output can then be compared with pre-specified thresholds to decide on what action should be taken.
- This approach attempts to remove emotions.
- The scoring process should be relatively simple and repeatable.
- Different users at different times should arrive at the same conclusion!

Statistical Learning



Rule-based decisions

- Scores for making decisions are often calculated using a rule-based approach.
- This approach will rely on observations or data and the rule for deriving the score may have evolved based on both scientific knowledge and observations.
- Essentially, the rule encapsulates common sense and is both repeatable and objective.

APGAR score

- The Apgar score is the first test given to a newborn, usually in the delivery room.
- The test was designed to quickly evaluate a newborn's physical condition and to see if there's an immediate need for extra medical or emergency care.
- Developed in 1952 by an anesthesiologist named Virginia Apgar.
- Acronym for: Appearance, Pulse, Grimace, Activity, and Respiration.

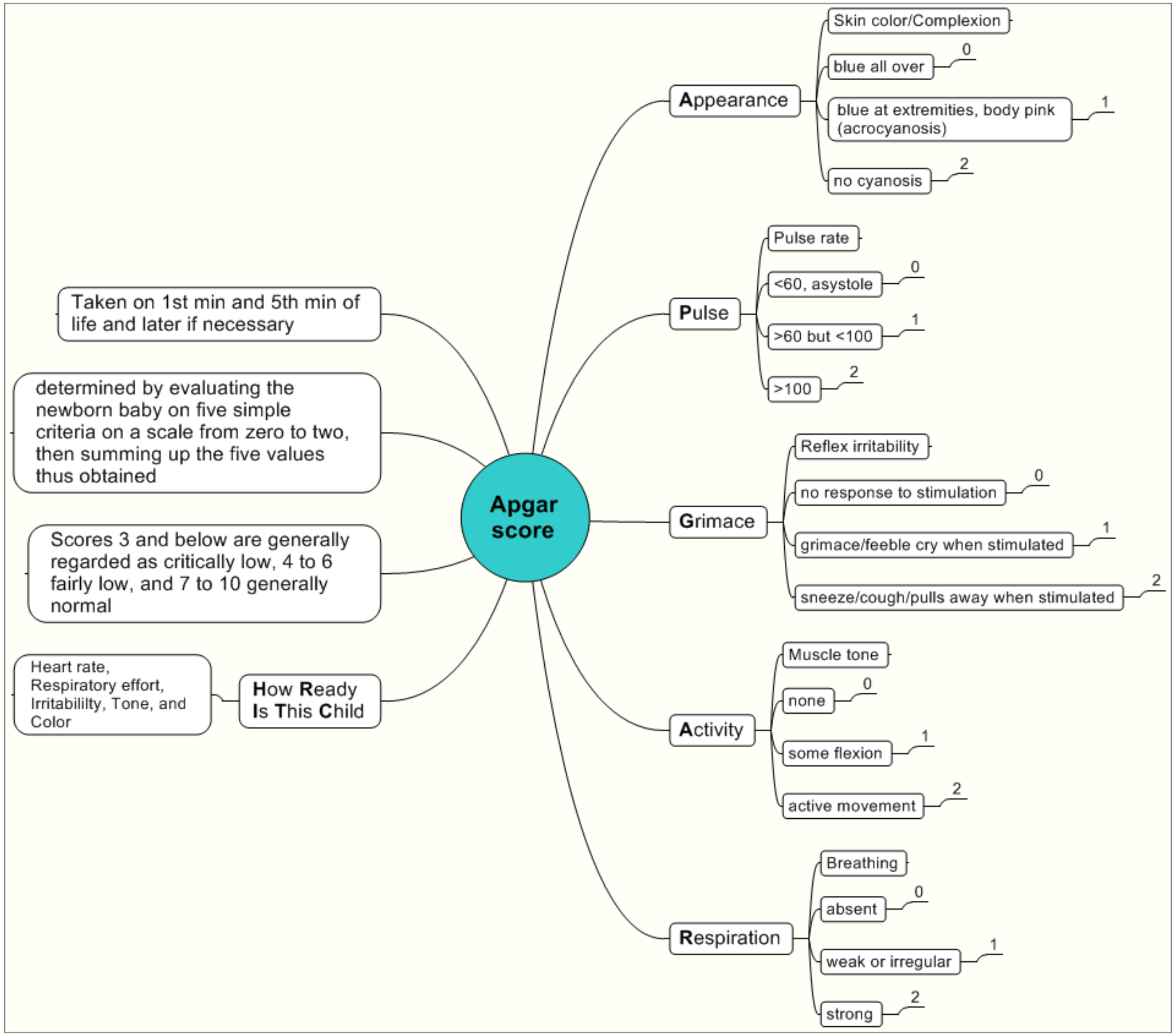
Five criteria of Apgar score

Component\Score	Score of 0	Score of 1	Score of 2
Appearance	Blue or pale all over	Blue at extremities Body pink	Body and extremities pink
Pulse	Absent	<100	>100
Grimace	No response to stimulation	Grimace on suction or aggressive stimulation	Cry on stimulation
Activity	None	Some flexion	Flexed arms and legs that resist extension
Respiration	Absent	Weak, irregular, gasping	Strong, lusty, cry

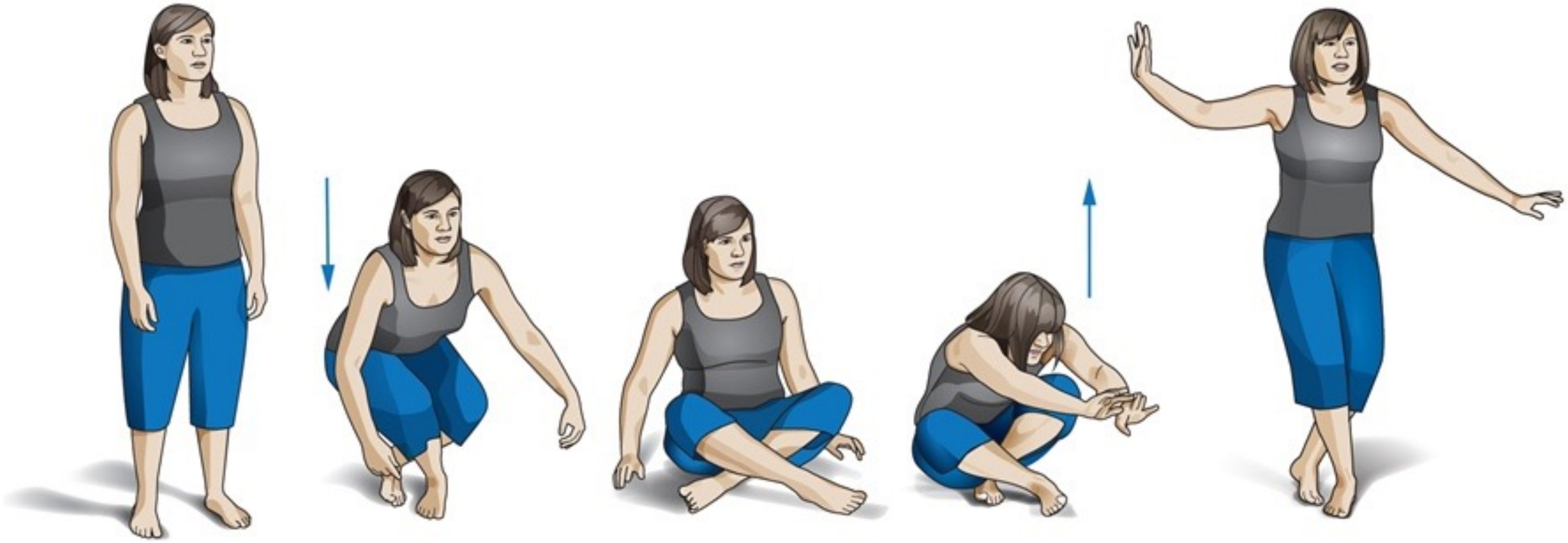
Interpretation of Apgar

Apgar Score	Interpretation	Response to score after 10-30 minutes
1	Critically Low	Risk of neurological damage and small increase in risk of cerebral palsy
2		
3		
4	Fairly Low	Baby requires medical attention
5		
6		
7	Normal	No medical treatment required
8		
9		
10		

Apgar Score Mind Map

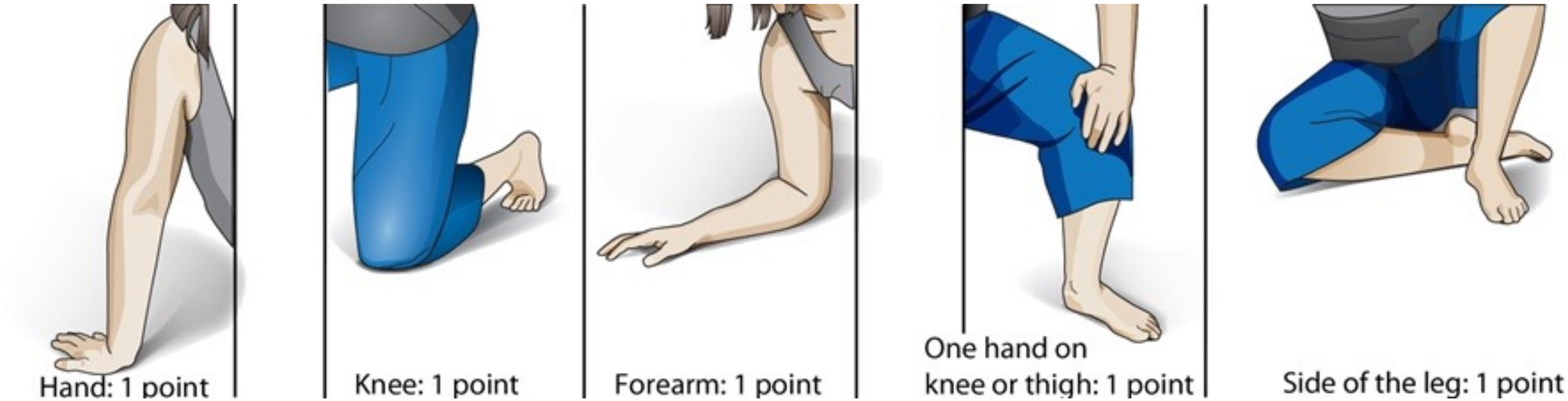


The sitting-rising test



Claudio Gil Araujo et al., *European Journal of Cardiology (2012).*

Scoring


Hand: 1 point
Knee: 1 point
Forearm: 1 point
One hand on knee or thigh: 1 point
Side of the leg: 1 point

Ability to sit and rise from the floor as a predictor of all-cause mortality

- Musculoskeletal fitness, as assessed by SRT, was a significant predictor of mortality in patients aged between 51 and 80.
- People can score a maximum of 10 points, with 1 point deducted for putting a hand or leg for stability, and half a point docked for wobbling.
- Those who scored three points or less out of 10, were more than five times as likely to die within six years, as those who scored more than eight points.
- Every point increase in the test, was linked to a 21% decrease in mortality from all causes.

Why are you interested in machine learning?

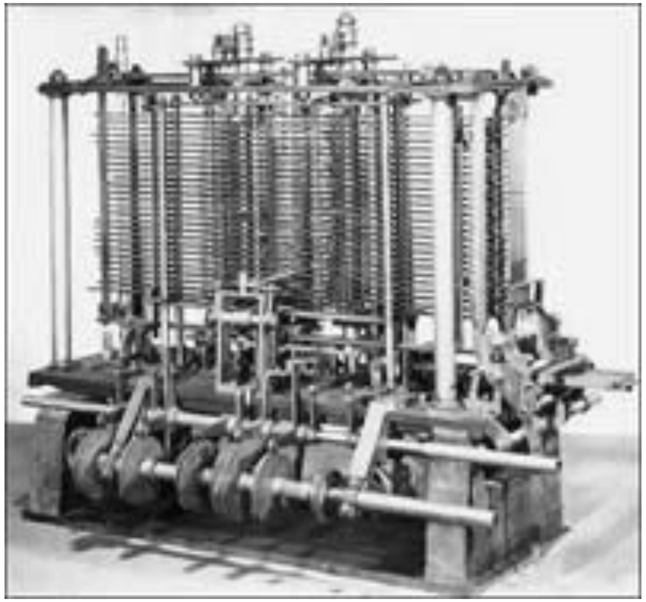
1. Academic interest in ML techniques
2. Wish to apply ML in the private sector
3. Wish to use ML for social good
4. Wish to undertake research

Slido.com
#54377

Computing Timeline

- 1837: Charles Babbage proposed the first general mechanical computer, the Analytical Engine.
- 1936: Alan Turing proposes the Turing Machine.
- 1943: Tommy Flowers develops Colossus to help British code breakers during WW2.
- 1946: ENIAC, first digital computer.
- 1953: IBM introduce the 701.
- 1975: IBM 5100 is the first portable computer.
- 1976: Apple introduce the Apple 1.

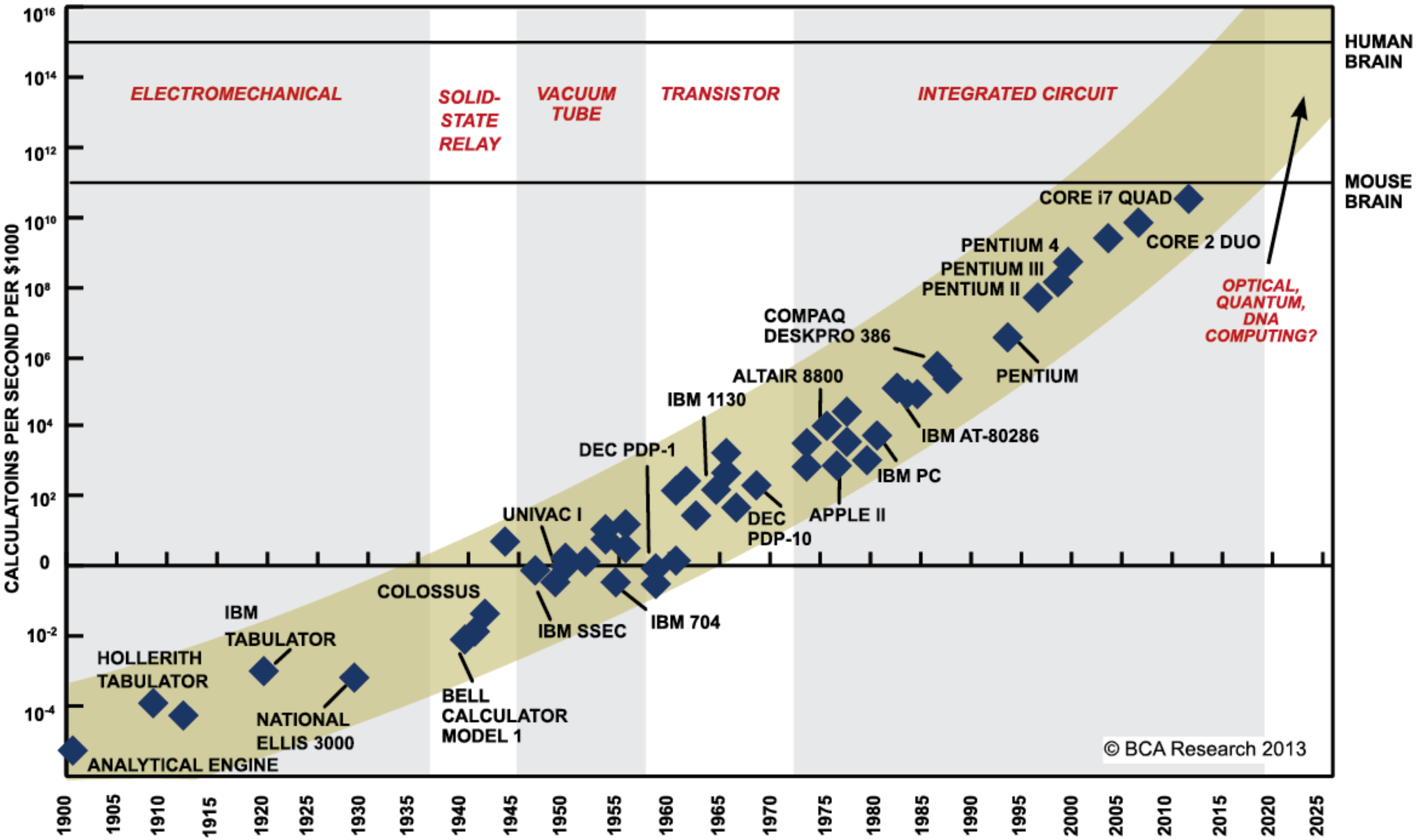
Evolution



1910

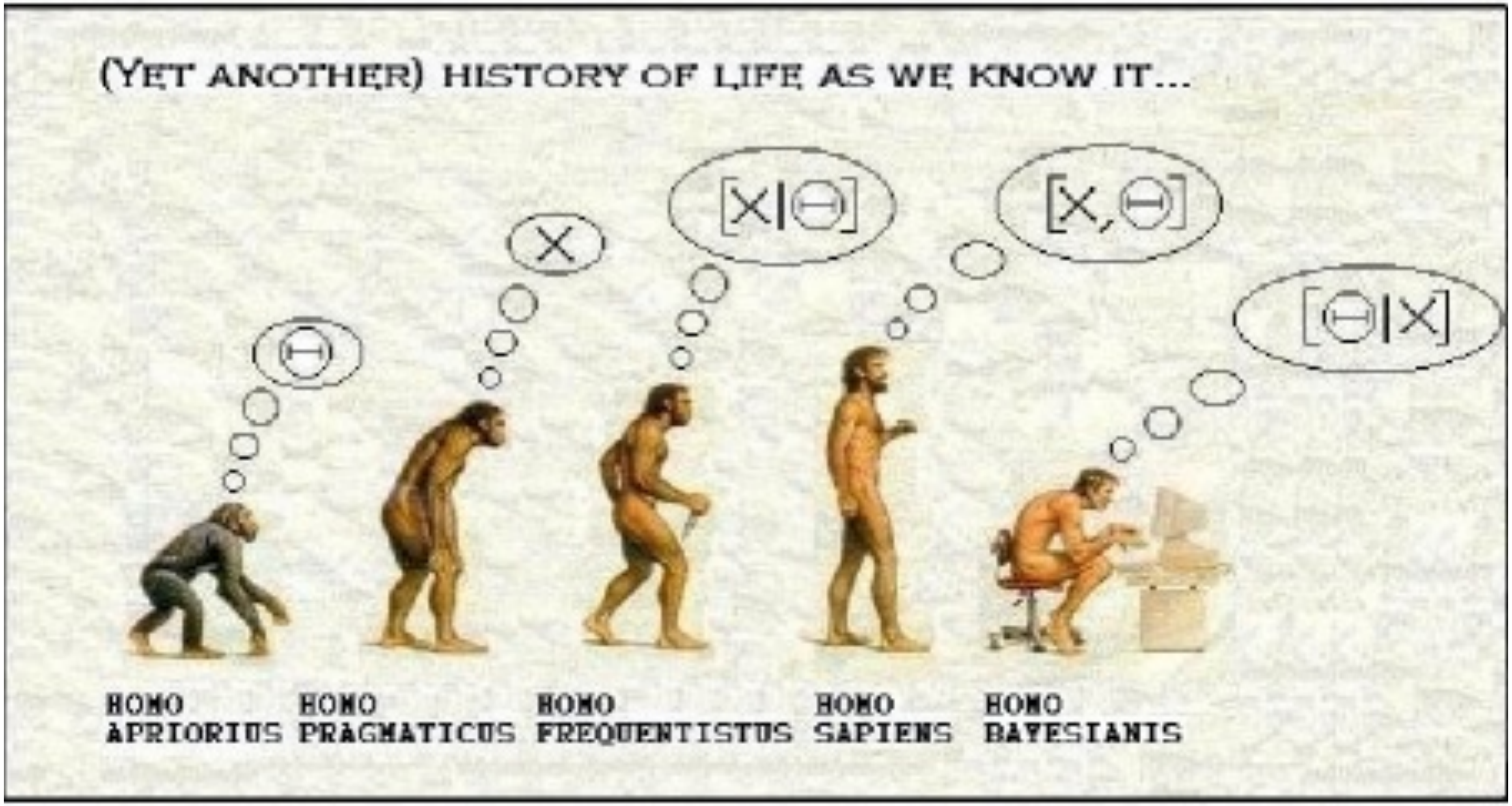
1975

Computers and Intelligence

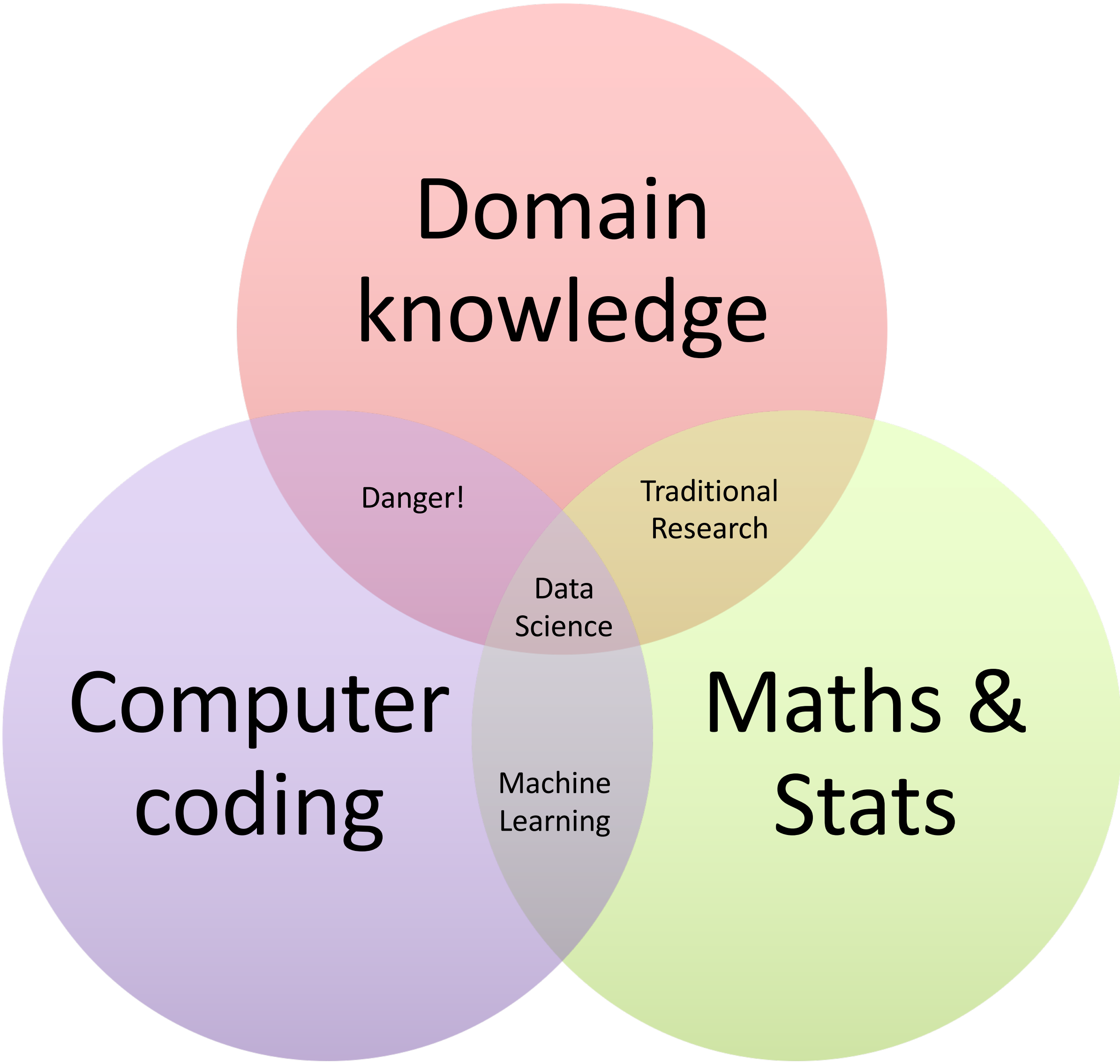


SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

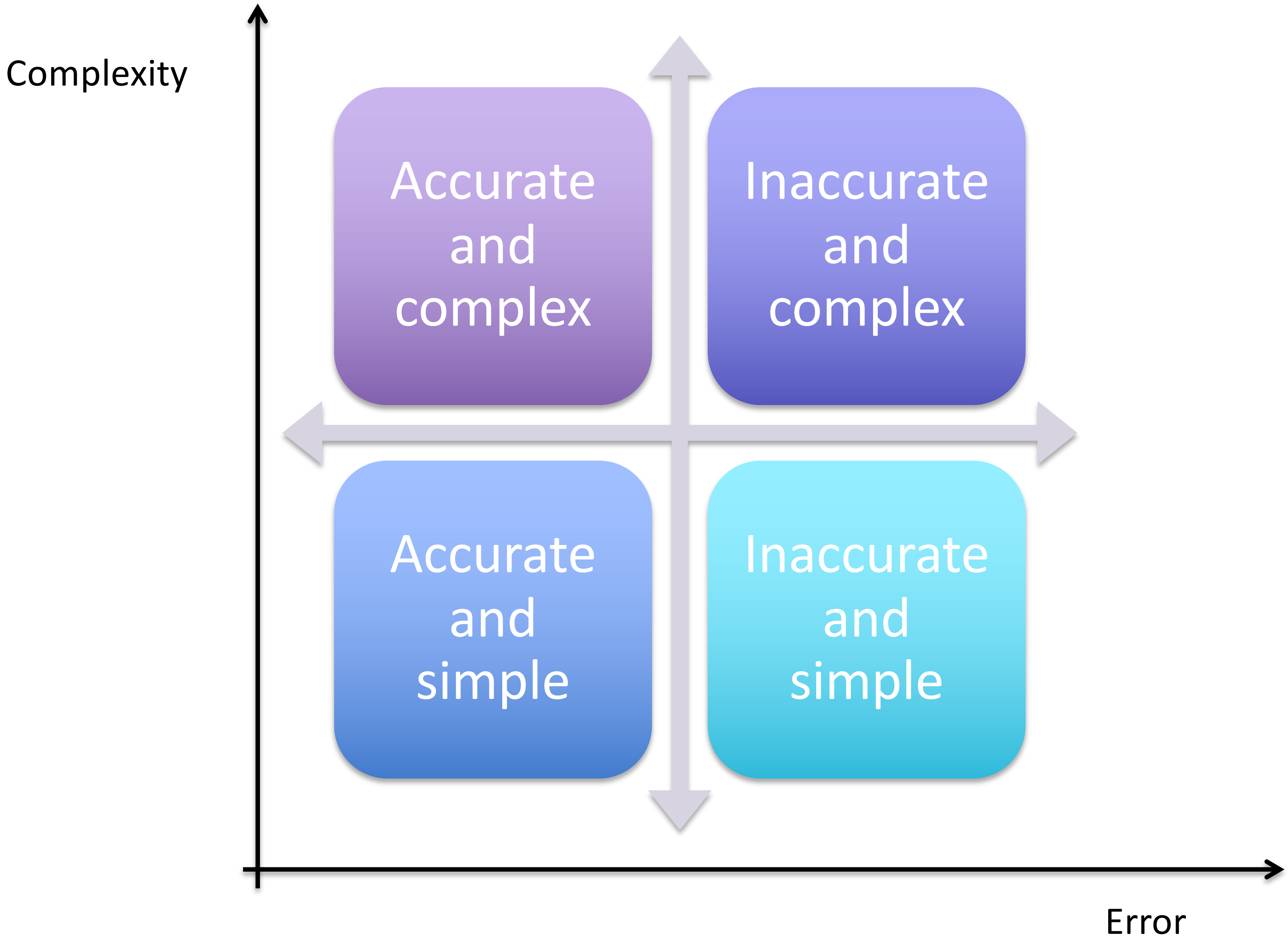
Where do we fit in?



Fusion of disciplines



Complexity versus Accuracy



Machine Learning Timeline

Year	Name	Description
1950	Alan Turing Test	A machine can actually learn, if when we communicate with it, we cannot distinguish it from another human.
1952	ELIZA	Arthur Samuel (IBM) wrote the first game-playing program, ELIZA, for checkers, to achieve sufficient skill to challenge a world champion.
1957	Neural Network	Frank Rosenblatt (Cornell University) invented the perceptron, a very simple linear classifier.
1990	Artificial Intelligence	Computer science and statistics combined to provide a data-driven approach to machine learning.
2010	Big Data	Exponential growth in the volume, velocity and variety of data available for analysis and research.
2014	Open Data & IoT	Infrastructure, protocols and standards for providing open access to data via APIs.

Machine Learning Definition

- Machine learning refers to the discovery of knowledge from data and experience.
- It involves the construction and study of systems that can learn from data and improve with experience.
- These techniques tend to be capable of identifying nonlinear relationships in large datasets.
- Initially popular in engineering and bioinformatics but now finding applications in a large number of areas.

Machine Learning Concepts

- **Representation**: ability to represent data instances and functions evaluated on these instances (model).

- **Generalisation**: the property that the system will perform well on unseen data instances (accuracy of application).

Q&A

Applied Machine Learning

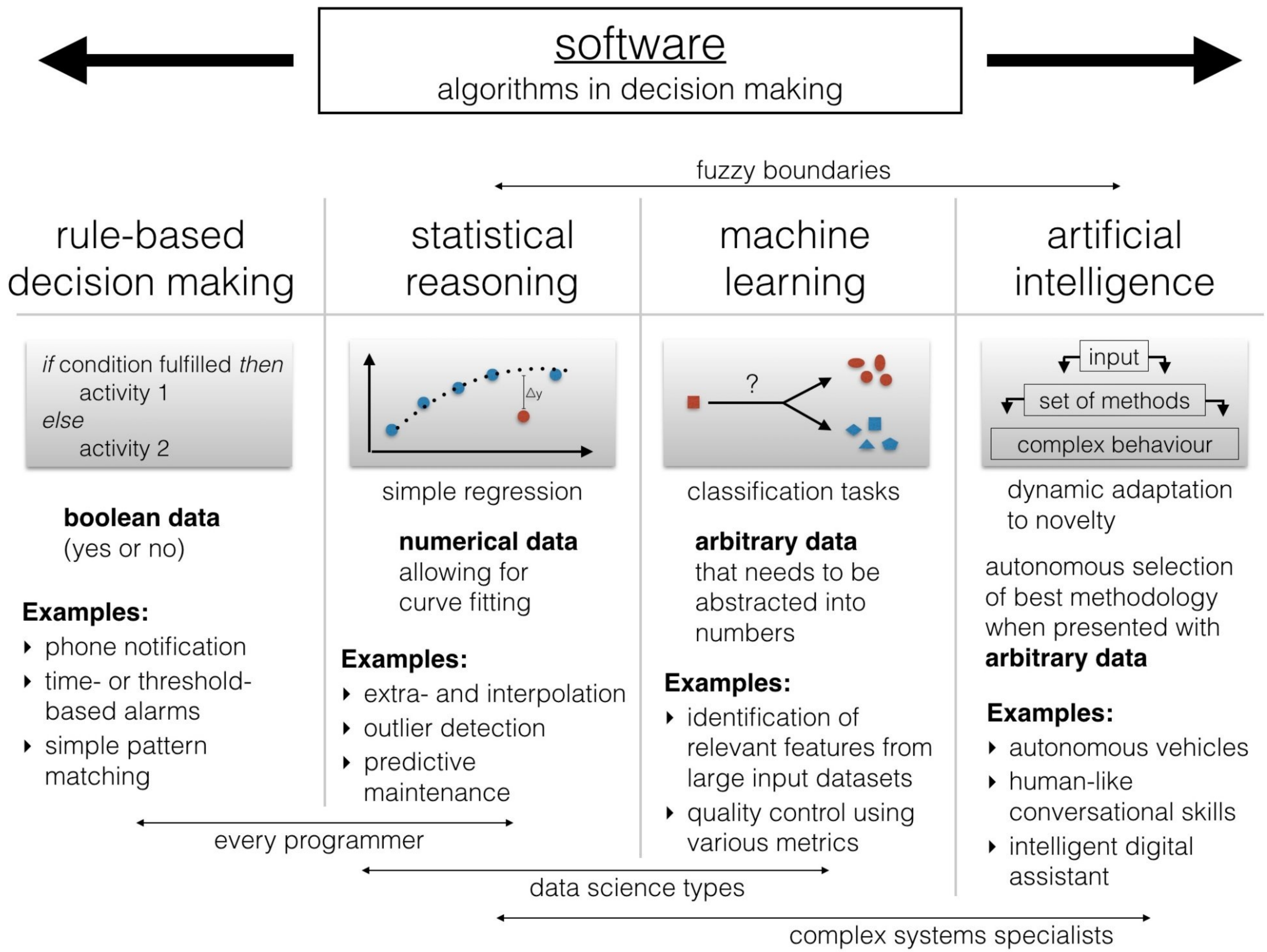
WEEK 7B

Review of last lecture

- Learning and decision-making
- Evolution of computing and intelligence
- History of machine learning
- Rule-based approaches
- Machine learning concepts
- Applications

Today's Lecture

No.	Activity	Description	Time
1	Challenge	Learning from data	10
2	Discussion	Sports prediction	10
3	Case study	Rugby	10
4	Analysis	Fitting models	20
5	Demo	Curve fitting	20
6	Q&A	Questions and feedback	10



Source: Lukose (2016), GCS Agile

Quiz

- Where is the Sharpe Ratio (return per unit of risk) not applicable?
  - Insurance
  - Financial trading
  - Health outcome prediction
  - Sports prediction

- **Slido.com**
  #78272

Case Study: Sport prediction

- How does one forecast the outcome of a game such as football, rugby, tennis or cricket?
- Support your favorite team
- Gut feeling
- Data-driven approach

Rugby

- Ireland playing Wales at home in Wales.
- How do we predict who will win?
- What is the probability of Ireland winning?

The statistics

- Ireland and Wales have played each other in 120 rugby games since 1882.
- A total of 120 matches have been played, with the following results:
  - Wales winning 65 matches
  - Ireland winning 49 matches and
  - Six matches drawn.
- Probability of Ireland winning = 49/120 = 0.41

Detailed statistics

Details	Played	Won by Ireland	Won by Wales	Drawn	Ireland points	Wales points
In Ireland	59	28	27	4	698	585
In Wales	56	19	35	2	537	735
Neutral Venue	5	2	3	0	69	81
Overall	120	49	65	6	1304	1401

Playing at home in Wales, Ireland has probability of win = 19/56 = 0.34

Poll

- What other variables would you consider to predict the outcome of a rugby match?

Slido.com
#78272

Other variables

- Managers: Joe Schmidt, from New Zealand, took over as manager of Ireland in 2013
- Composition of the teams
- Fitness of players
- Psychology
- Recent performance
- Weather conditions

Recent Statistics

Date	Venue	Score	Winner	Competition
14 March 2015	Millennium Stadium	–		2015 Six Nations
8 February 2014	Aviva Stadium	26 – 3	Ireland	2014 Six Nations
2 February 2013	Millennium Stadium	22 – 30	Ireland	2013 Six Nations
5 February 2012	Aviva Stadium	21 – 23	Wales	2012 Six Nations
8 October 2011	Regional Stadium	22 – 10	Wales	2011 Rugby World Cup
12 March 2011	Millennium Stadium	19 – 13	Wales	2011 Six Nations
13 March 2010	Croke Park	27 – 12	Ireland	2010 Six Nations
21 March 2009	Millennium Stadium	15 – 17	Ireland	2009 Six Nations
8 March 2008	Croke Park	12 – 16	Wales	2008 Six Nations
4 February 2007	Millennium Stadium	9 – 19	Ireland	2007 Six Nations
26 February 2006	Lansdowne Road	31 – 5	Ireland	2006 Six Nations
19 March 2005	Millennium Stadium	32 – 20	Wales	2005 Six Nations
22 February 2004	Lansdowne Road	36 – 15	Ireland	2004 Six Nations
16 August 2003	Lansdowne Road	35 – 12	Ireland	2003 Rugby World Cup preparatory game
22 March 2003	Millennium Stadium	24 – 25	Ireland	2003 Six Nations
3 February 2002	Lansdowne Road	54 – 10	Ireland	2002 Six Nations
13 October 2001	Millennium Stadium	6 – 36	Ireland	2001 Six Nations

Six Nations Table

POS		TEAM	PL	W	D	L	PF	PA	DIFF	TF	TA	PTS
1	⬆ (2)	IRELAND	3	3	0	0	63	23	40	3	1	6
2	⬇ (1)	ENGLAND	3	2	0	1	77	52	25	8	5	4
3	⬆ (4)	WALES	3	2	0	1	62	57	5	4	5	4
4	⬇ (3)	FRANCE	3	1	0	2	39	46	-7	2	2	2
5	⬆ (6)	ITALY	3	1	0	2	42	92	-50	6	9	2
6	⬇ (5)	SCOTLAND	3	0	0	3	50	63	-13	4	5	0

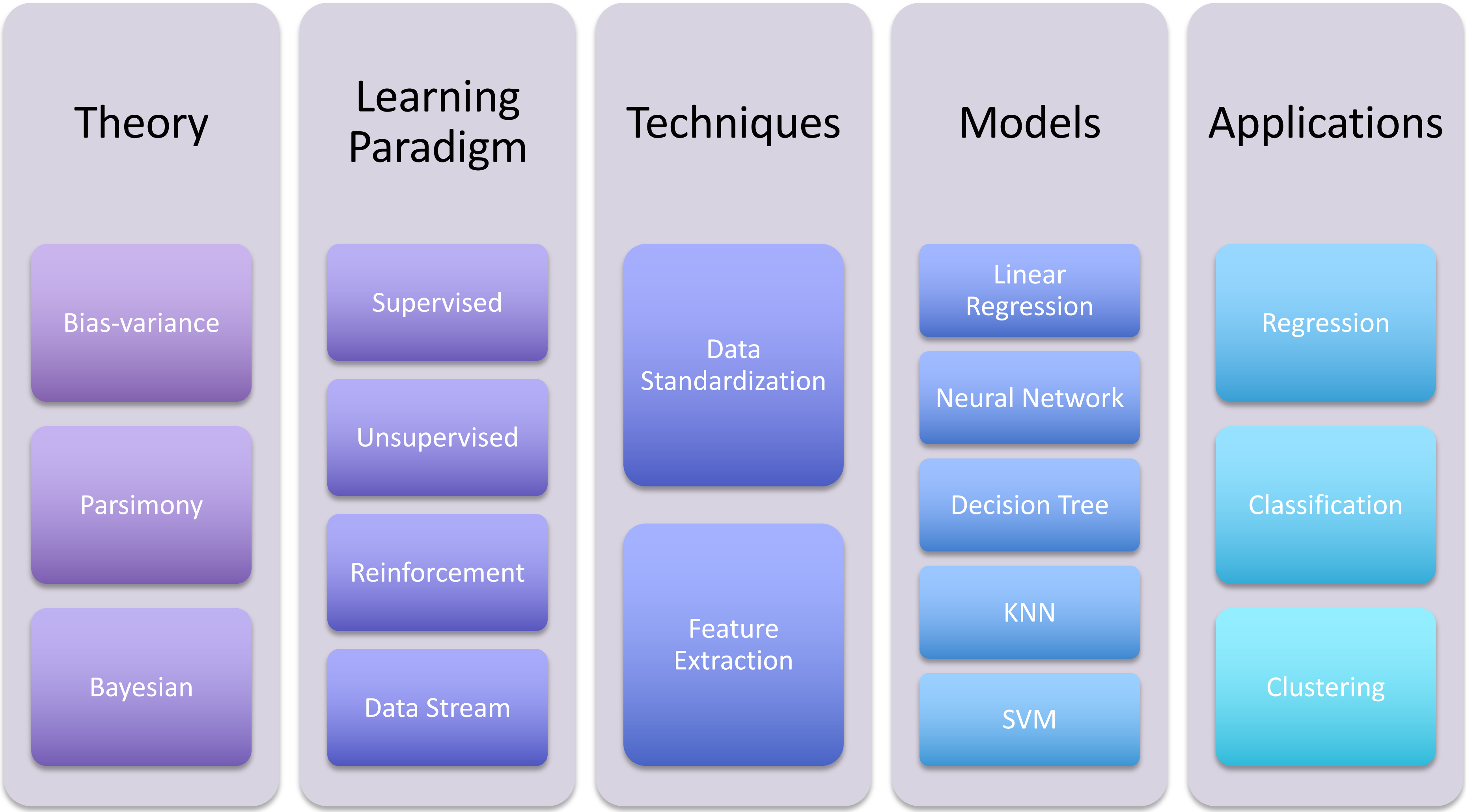
Odds on the game

Wales v Ireland 6 Nations Odds

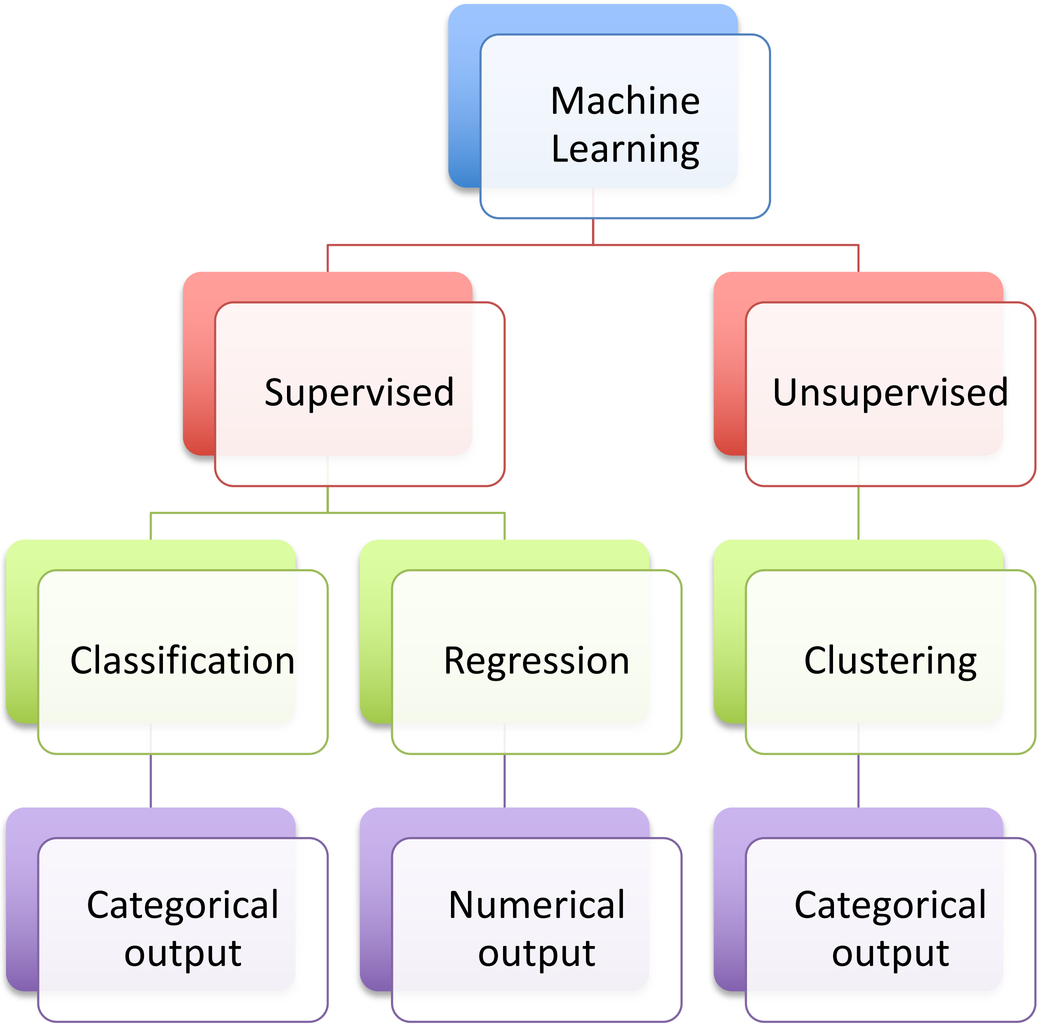
Match Betting:

	PADDYPOWER	William HILL	Ladbrokes	BETFRED	bet365	betfair	BETVICTOR	sky BET
Ireland	4/6	4/5	5/6	4/5	4/5	5/6	4/5	8/11
Wales	6/5	11/10	1	1	1	11/10	1	6/5
Draw	18	18	18	20	20	14	18	20

Machine Learning Roadmap



Machine Learning Techniques



Machine Learning Algorithms

	Unsupervised	**Supervised**
Continuous	Clustering • PCA • K-means	Regression Decision Trees Random Forests
Categorical	Association Analysis • Apriori Hidden Markov Models	Classification • Logistic Regression • KNN • SVM

Poll

- Logistic regression is an appropriate model for predicting:
- (a) tomorrow’s temperature
- (b) closing price of S&P500 next week
- (c) probability of South Africa winnning the next Rugby World Cup
- (d) severity of Parkinson’s disease for a particular patient

Learning from Data

- Most machine learning problems start with identifying a dependent variable of interest, y_n and collecting n=1,..,N measurements
- Based on intuition or availability of data, it is usually possible to collect a series of M features (explanatory variables),

  X_{nm}, m=1,…,M.
- Each row of the feature matrix X_{nm}, corresponds to an instance y_n

Function fitting

- Given the feature matrix, X_{nm}, and corresponding instances of the dependent variable, $y_{n,}$ the challenge becomes one of function fitting.
- We wish to identify a function, F, that serves as a model and maps from

  X_{nm} to y_n:

$$y_n = F(X_{nm}) + \varepsilon_n$$

Estimating a Model

- In practice, the model F depends on some parameter values, a, which need to be estimated.
- Therefore we estimate model parameters, *a*, denoted by *a**, such that we have the approximation:

$$Y_n = F(X_{nm}, a^*) + \varepsilon_n$$

- In summary, *F* and *a** are selected to ensure that the residuals ε_n have desirable properties.

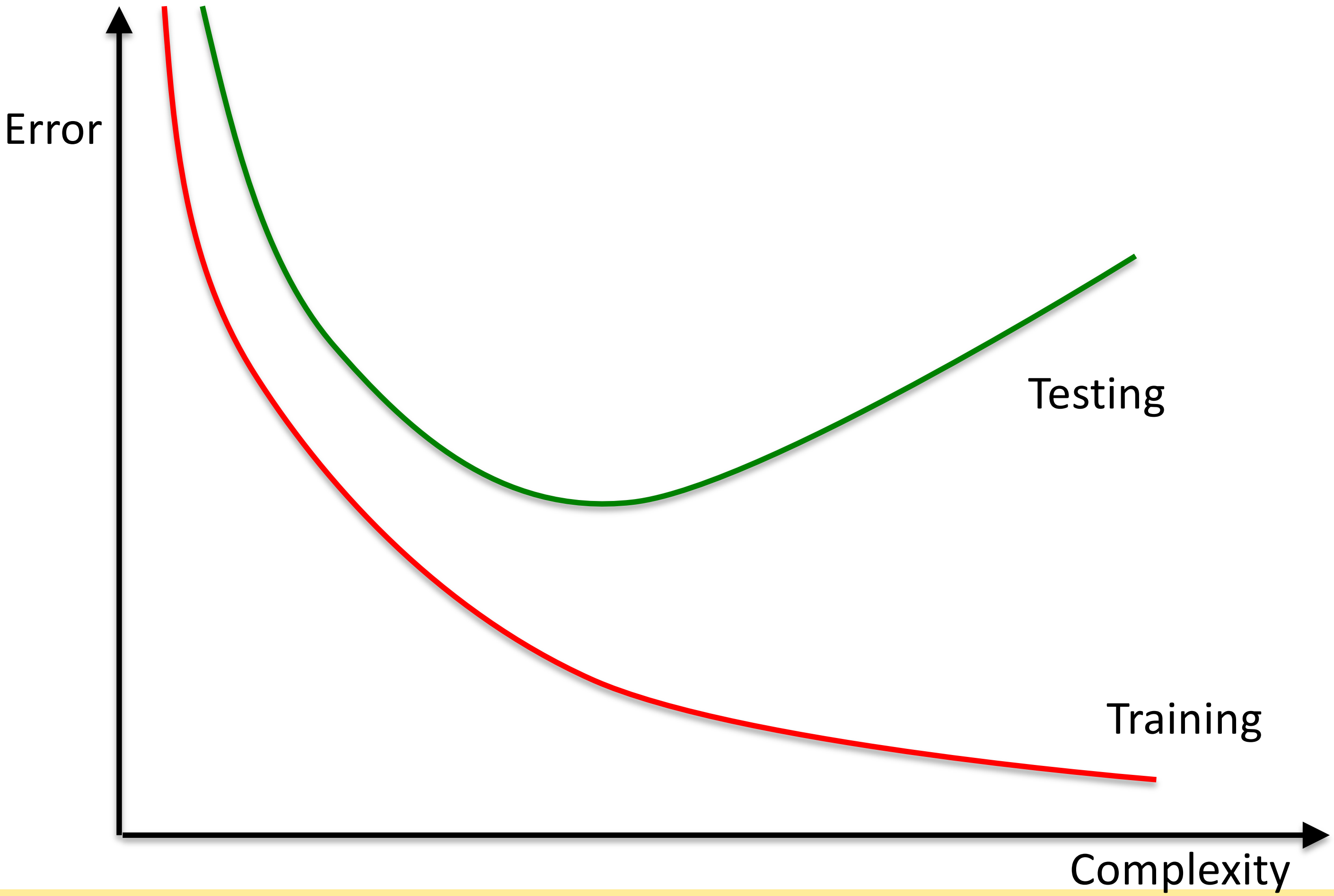
Curve fitting

- Consider fitting a curve using a polynomial model structure:

  $$y_n = a_0 + a_1x_n + a_2x_n^2 + \ldots + a_px_n^p$$

- Polynomial of order p requires estimation of a total of p+1 parameters.
- Complexity of the polynomial model can be quantified by the k=p+1 parameters.
- What should we expect when we use an order p polynomial to fit N observations?
- That is we fit the model to the N pairs (x_n,y_n).

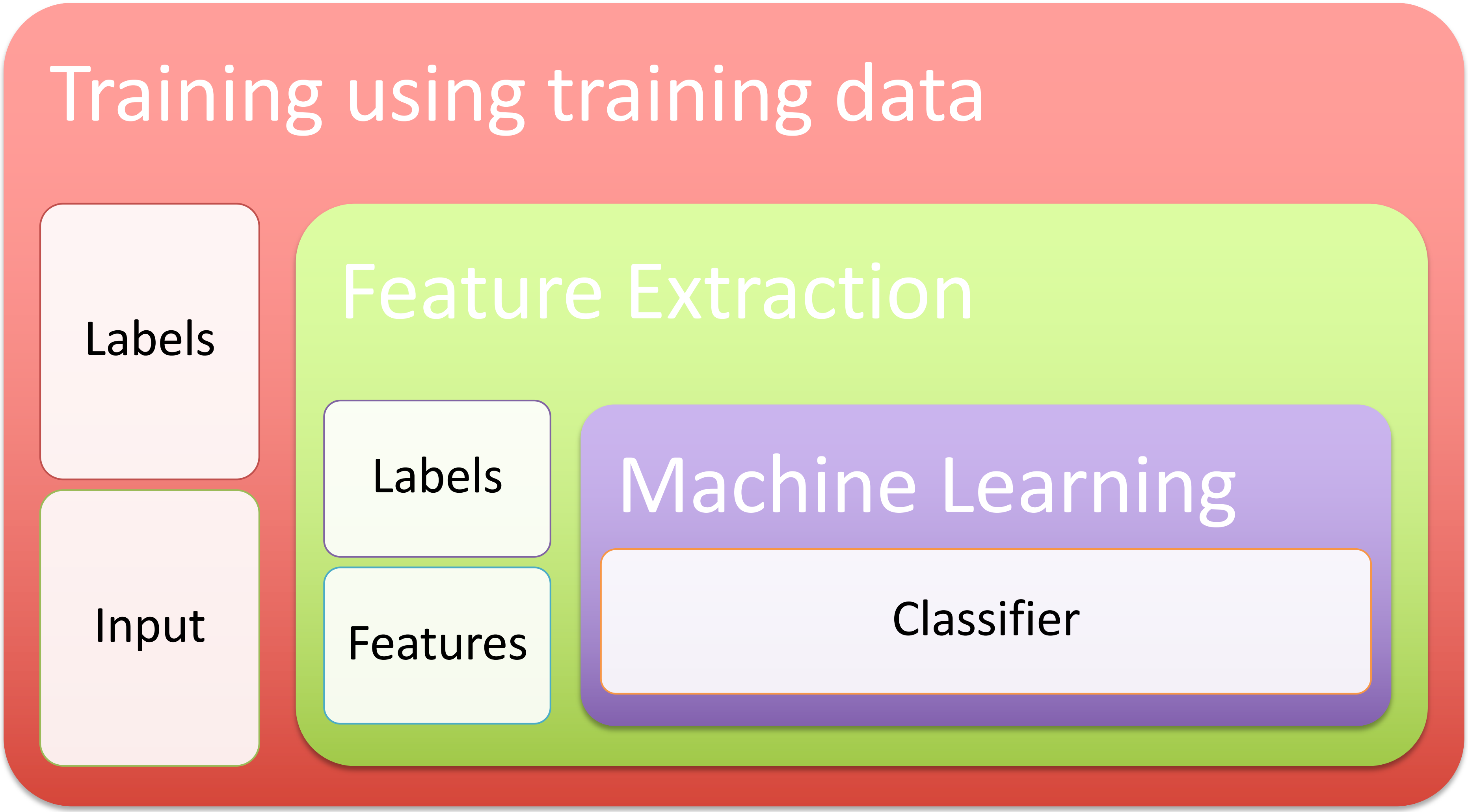
In-sample and out-of-sample



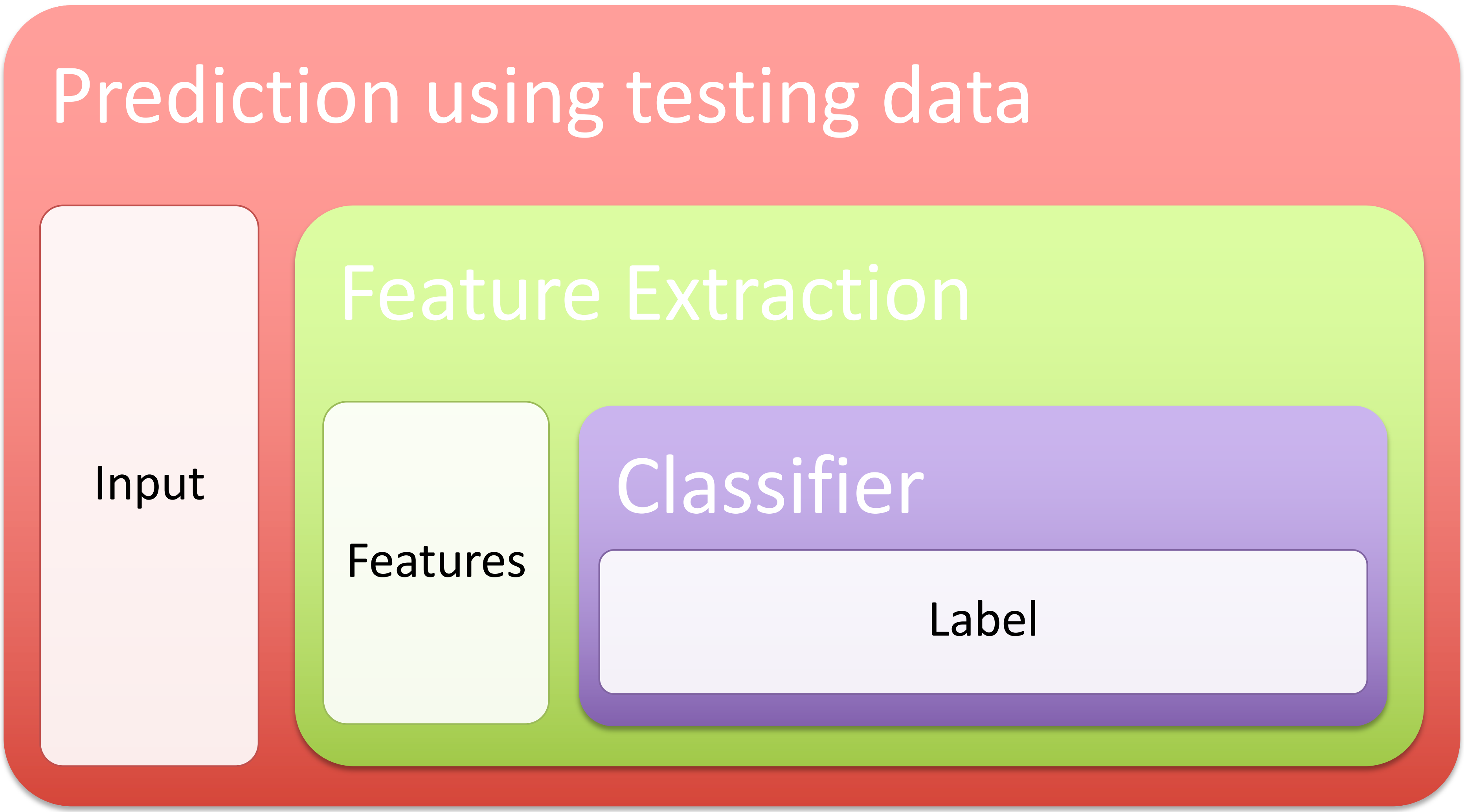
- Model over-fitting: performance is better on training (in-sample) data than testing (out-of-sample) data.

Source: McSharry (2005), The danger of wishing for chaos

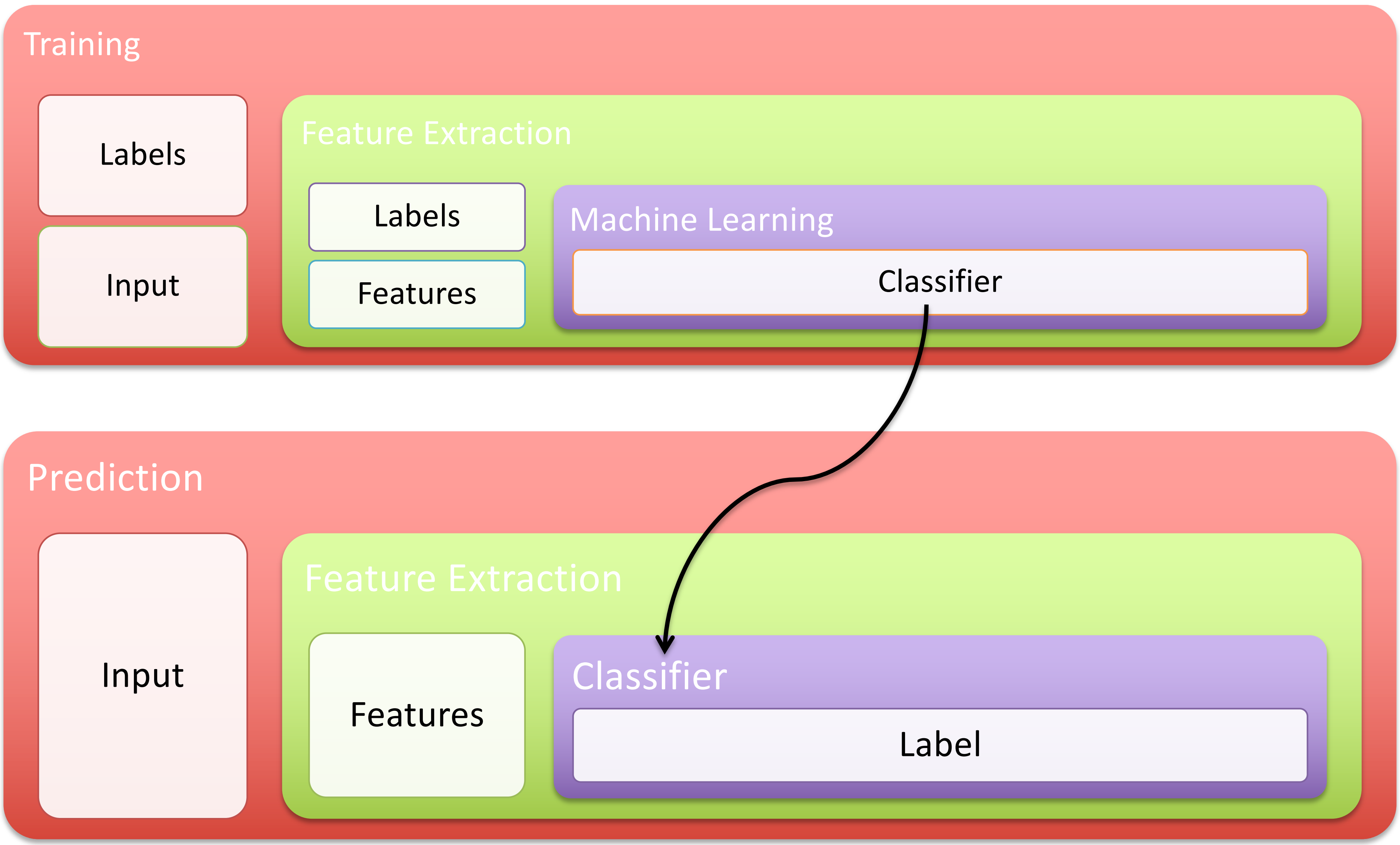
Training: estimation



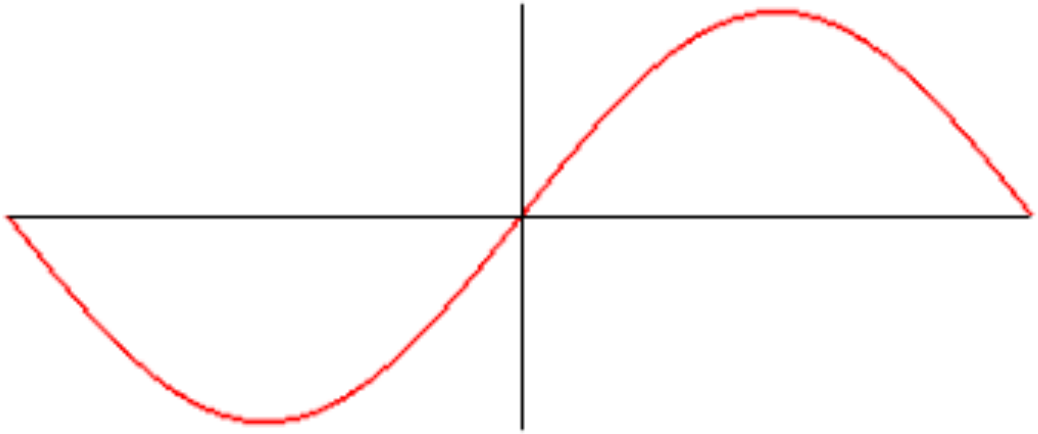
Prediction: evaluation



Training & Prediction

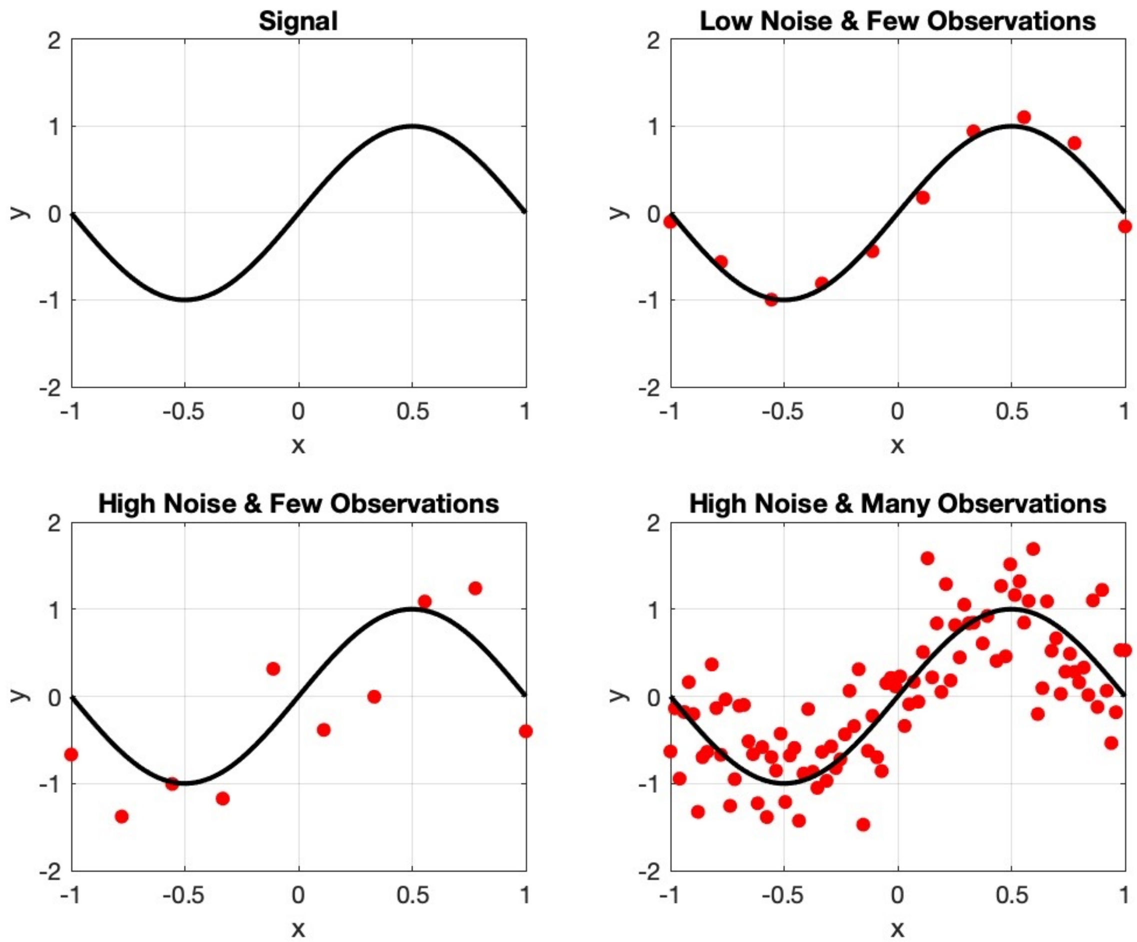


Poll 3



- Observations from one period of a sinusoid can be adequately modelled using a polynomial of degree:
  - (a) one
  - (b) two
  - (c) three
  - (d) four

Noisy Observations



Practical Exercise

- We explore fitting a polynomial model structure to a sinusoid observed with noisy measurements (DIAMLtutorial7b.m):
  - sin
  - randn
  - polyfit
  - polyval
- Study the effect of changing:
  - (1) the degree of the polynomial (model complexity);
  - (2) the number of observations; and
  - (3) the amount of noise.

Q&A