Introduction to Machine Learning

Dr Maria Anastasiadi
(m.anastasiadi@cranfield.ac.uk)

14th January 2025

www.cranfield.ac.uk

Machine learning definition

Machine Learning (ML) is the field of data science interested in the development of computer **algorithms** to transform data into insights and action.

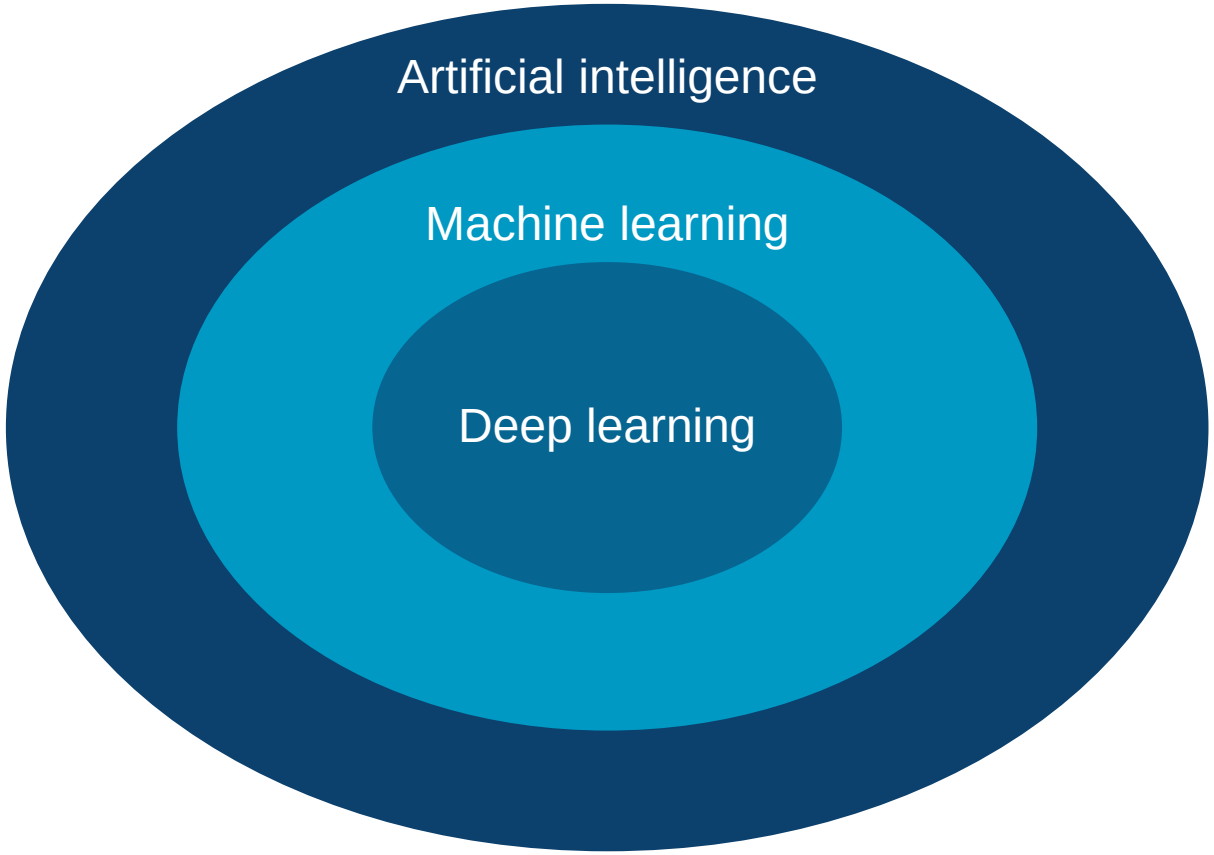
Machine learning is teaching computers how to use data to solve a problem.

ML driving force: Growth in **data** necessitated additional **computing power**, which in turn spurred the development of **statistical methods** to analyse large datasets.

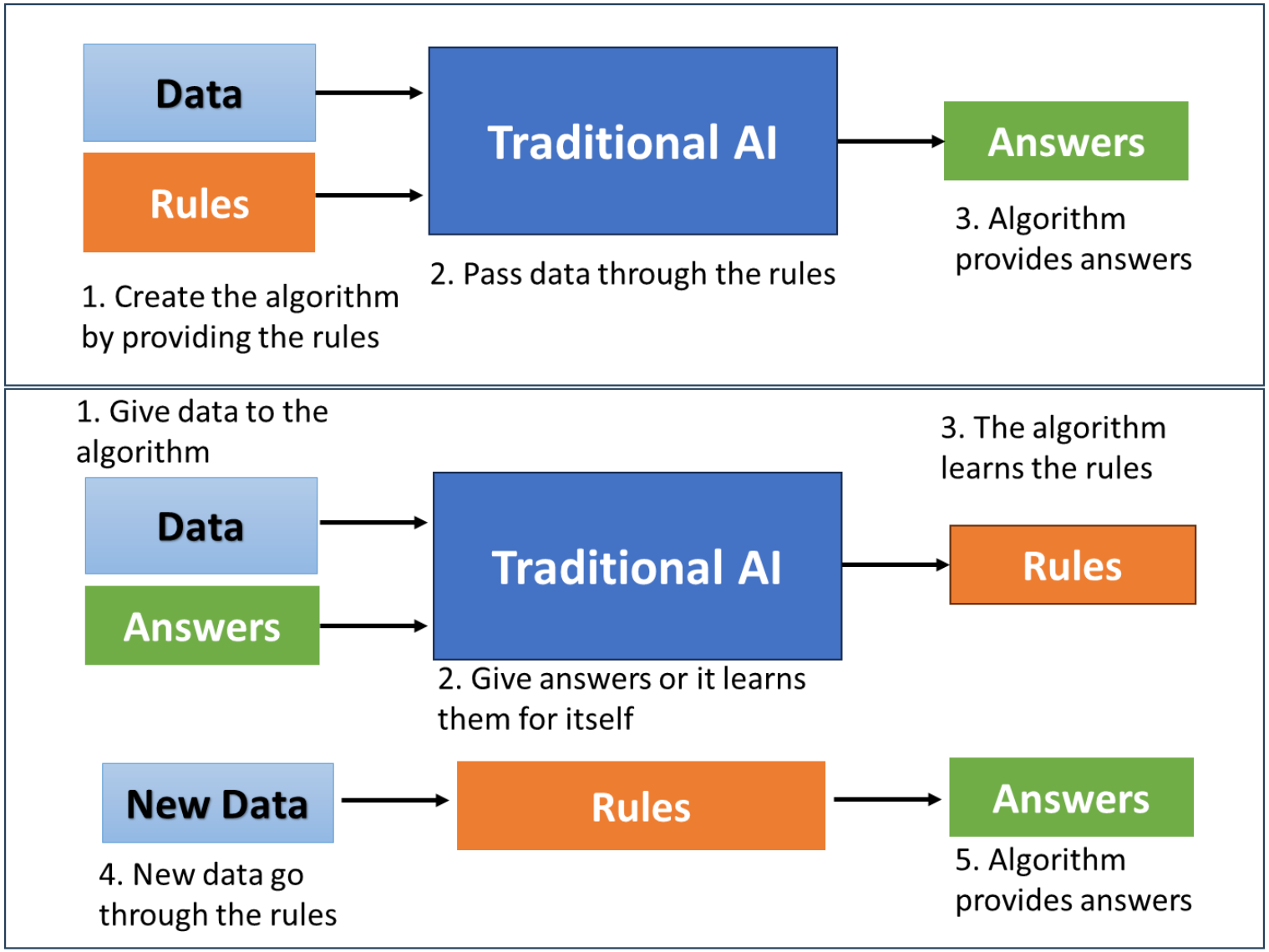
Machine learning definition

Machine Learning is a sub-field of artificial intelligence (AI).

- The term ML was first used in 1959 to describe a form of AI that involved training an algorithm to ***learn*** to play the game of checkers.
- Traditional AI is programmatic.
- ML trains a computer to learn relationships in data to identify meaningful patterns or make predictions.



Traditional AI vs ML

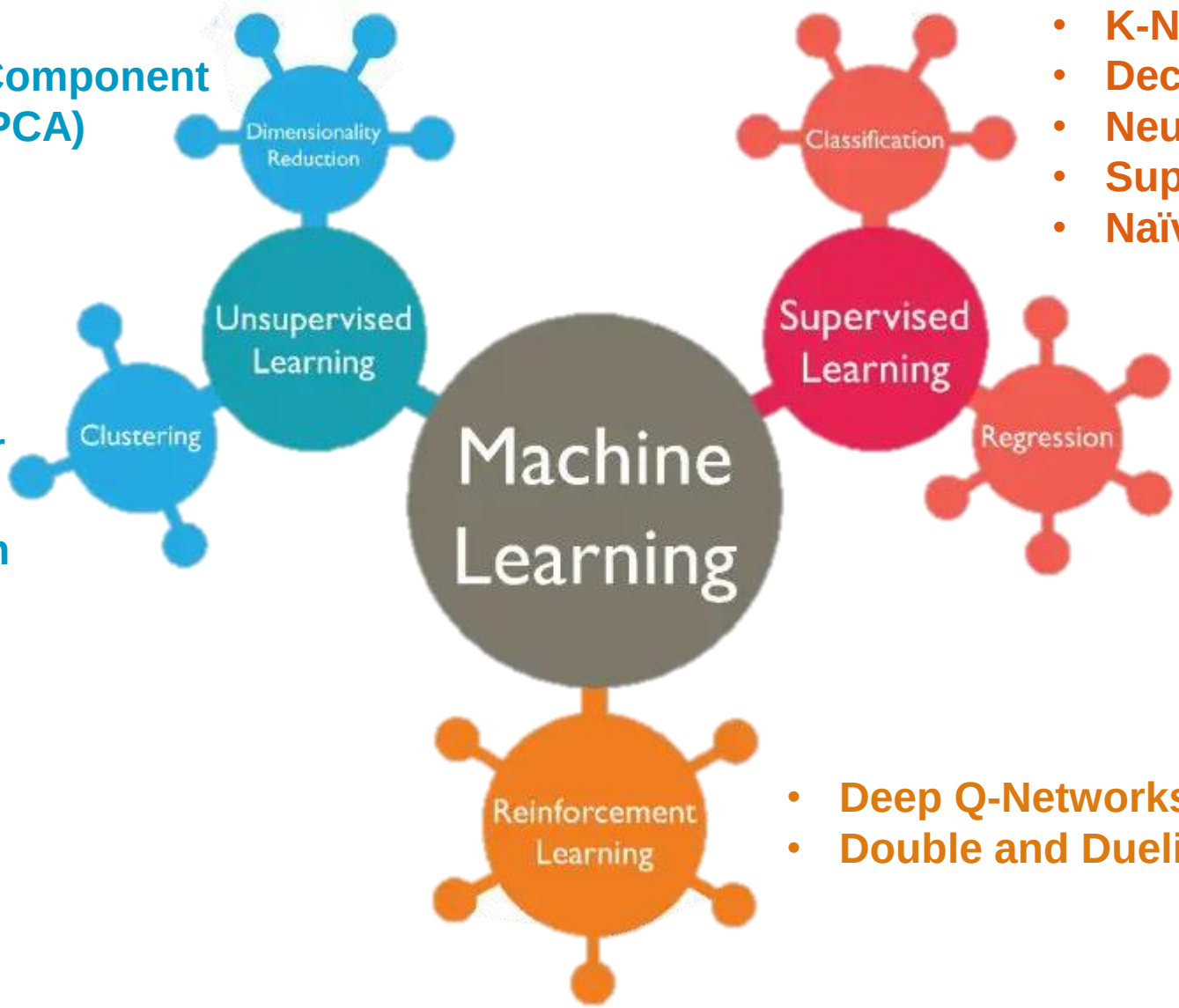


Adapted from: Hefin Rhys (2019). Machine Learning with R, tidyverse and mlr, v5, Manning.

ML applications

Image: https://www.javatpoint.com

Types of Machine Learning



Principal Component Analysis (PCA)

- **K-means**
- **Hierarchical Cluster Analysis (HCA)**
- **Fuzzy Classification**

- **K-Nearest Neighbours (kNN)**
- **Decision Trees**
- **Neural Networks**
- **Support Vector Machine (SVM)**
- **Naïve Bayes**

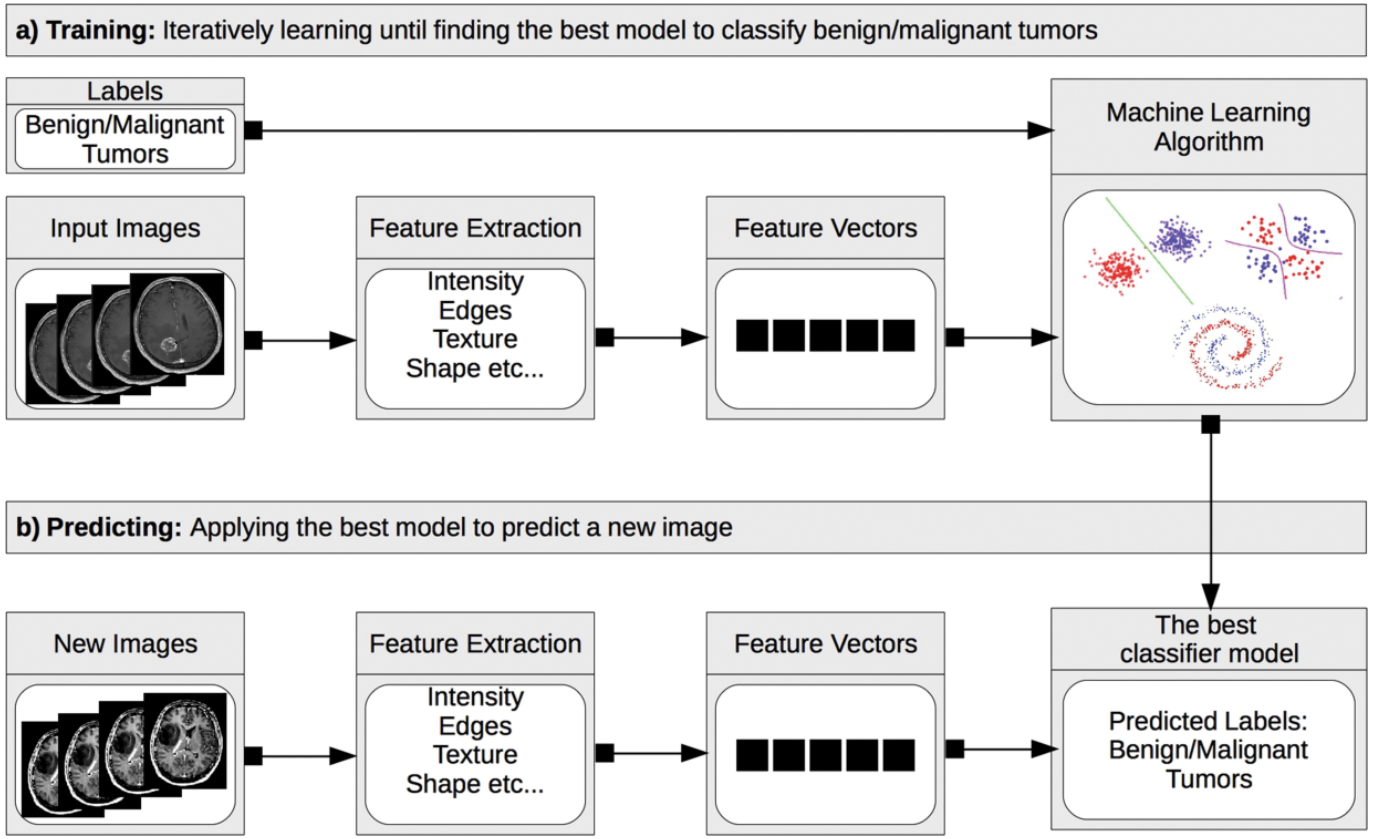
- **Generalised Linear Models**
- **Ridge Regression**
- **Elastic Net**
- **Partial Least Square (PLS)**

- **Deep Q-Networks (DQN)**
- **Double and Dueling DQN (DD-DQN)**

ML in healthcare

ML: a game-changer in healthcare

- Early disease identification and diagnosis:
  - Rapid cancer detection;
  - Heart disease;
  - Mental health disorders;
  - Genetic mutations.
- Personalised treatment.
- New drugs discovery.
- Predict health epidemics.



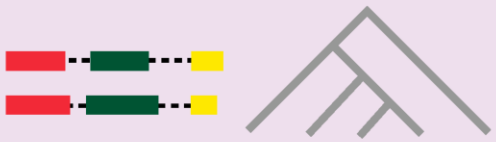
Erickson et al., 2017, *RadioGraphics*

ML applications in bioinformatics

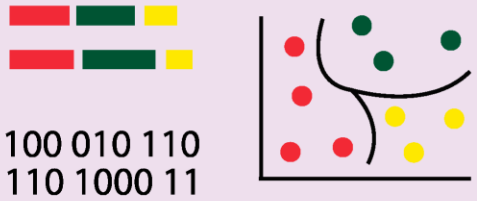
Machine and deep learning integration with bioinformatics

Molecular evolution

Phylogenetic inference



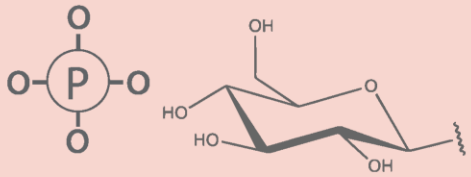
Alignment-free sequence classification



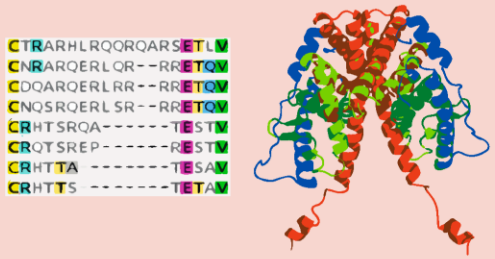
- *Inference of tree topology*
- *Sequence classification*
- *Viral sequence identification*
- *functional annotation*

Protein structure Analysis

Post translational modification



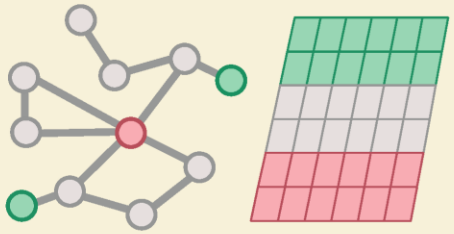
Folding and structure



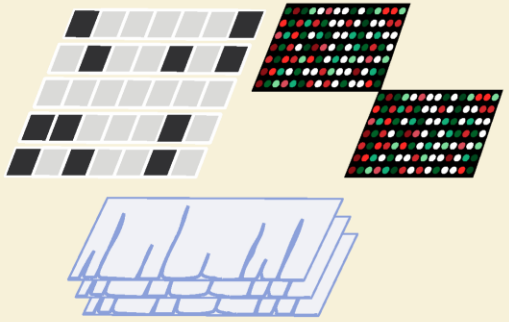
- *Phosphorylation site prediction*
- *Protein glycosylation prediction*
- *Protein contact maps prediction*
- *Structural homology prediction*

Systems biology

Biological Networks



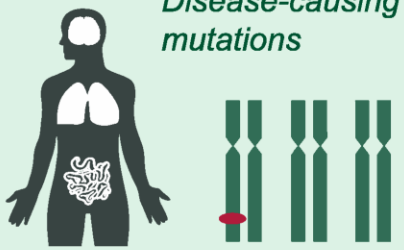
Multi-Omics integration



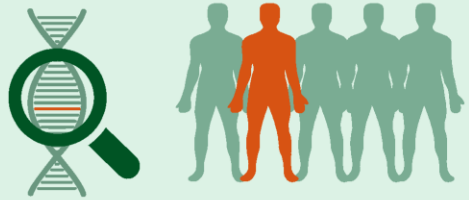
- *Biological networks construction*
- *Biological interactions prediction*
- *Pathway dynamics prediction*
- *Platform integration frameworks*

Genomics for Disease Research

Disease-causing mutations



Biomarkers discovery



- *Diesease associated genes and mutations*
- *Biomarkers*
- *Precision medicine applications*

Auslander N., et al. *Int. J. Mol. Sci.* 2021; https://doi.org/10.3390/ijms22062903

Applications of ML in Food Science and Technology

Nutrition related:
- amount of calorie
- level of nutrients

Agricultural:
-seed choice
-plant
-harvest

Categorizing:
-freshness
-variety and origin
-other parameters

Formulation:
-developing new products
-R&D

Waste:
-reusing
-management

Food Supply Chain Management:
- forecasting to manage pricing and inventory

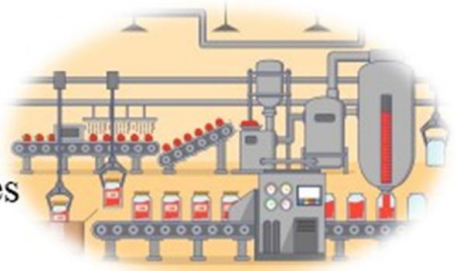
aspects of assess

Sensory Evaluation:
-flavor
-odor
-texture
-appearance

Quality:
-microbial/ chemical/ physical/ textural
-freshness
-detection un-related compound

Food Safety:
-ripeness
-detect adulteration, pesticides, additives

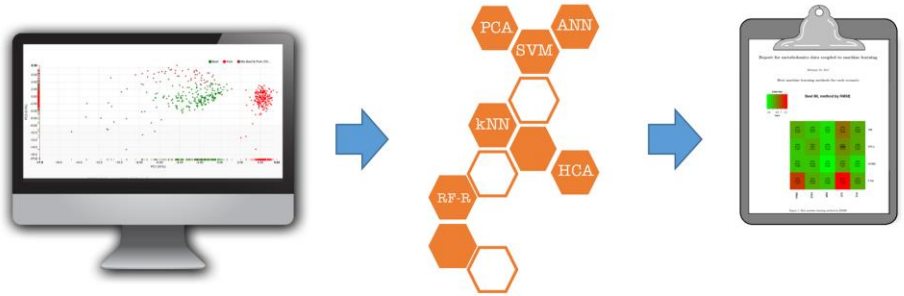
Process:
-drying
-sterilization
-other processes

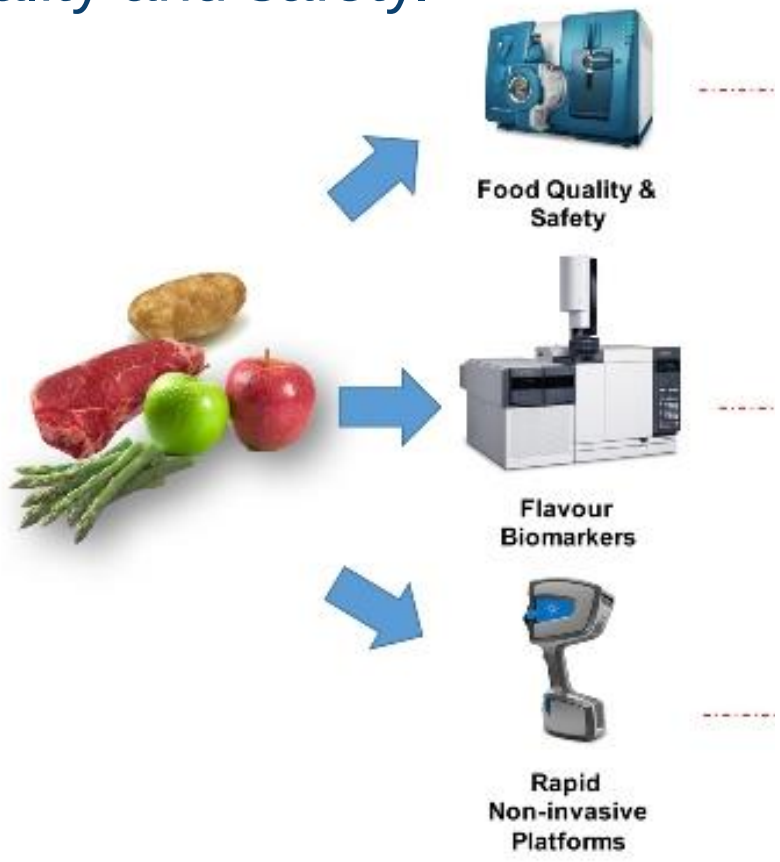


https://doi.org/10.1016/j.tifs.2023.104286

Bioinformatics Team ML applications

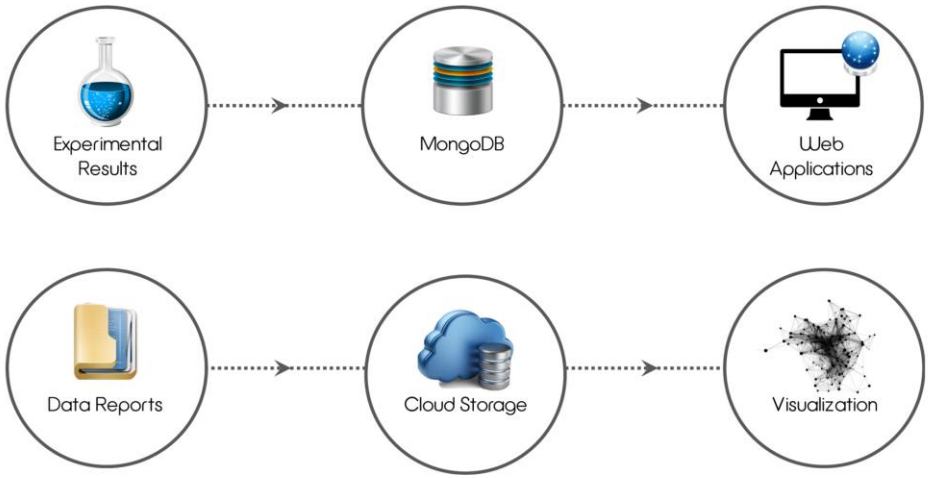
Applications in Food Diagnostics for monitoring of food quality and safety.



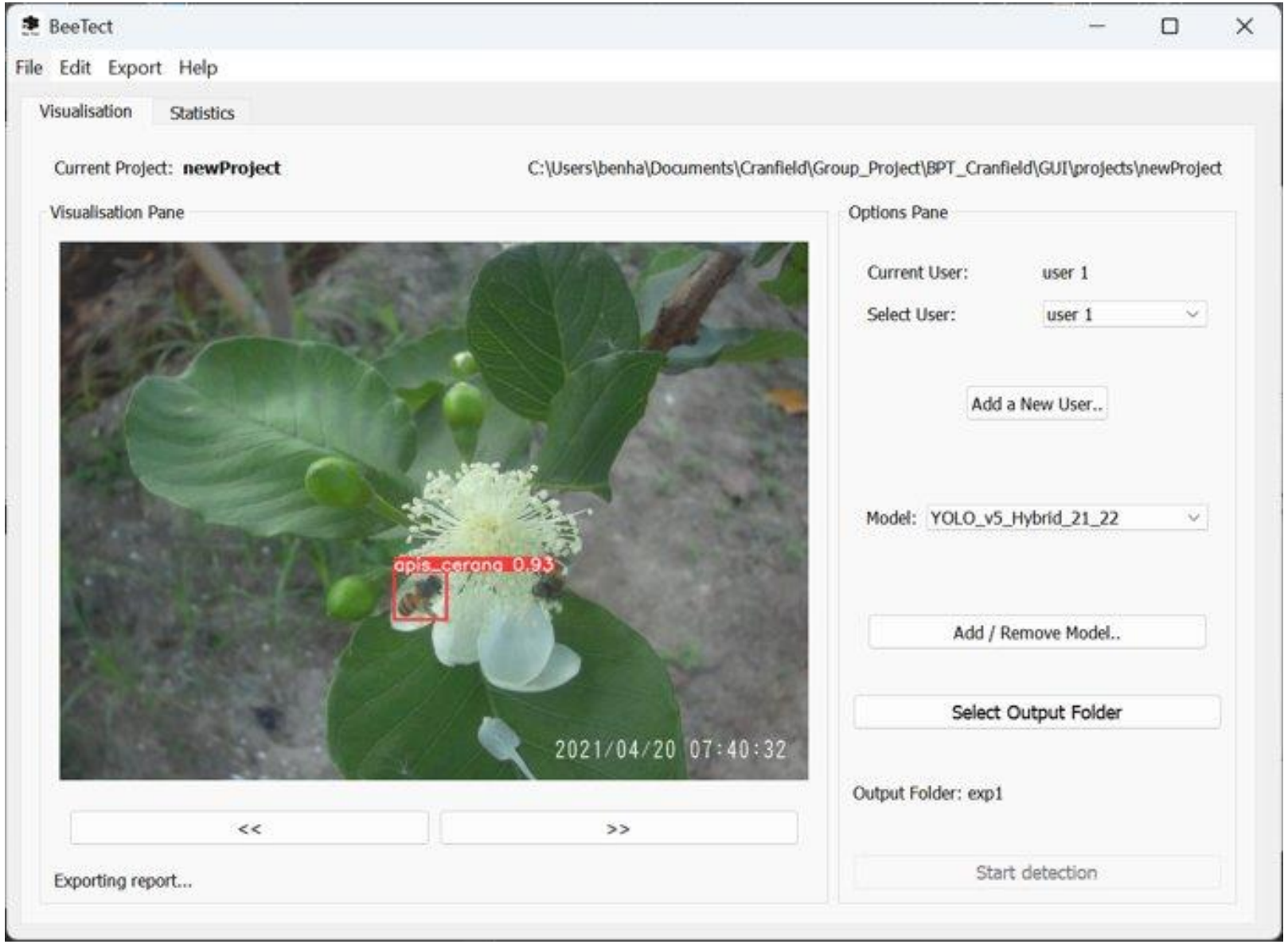


sorfML

The machine learning classification & regression analysis ranking system

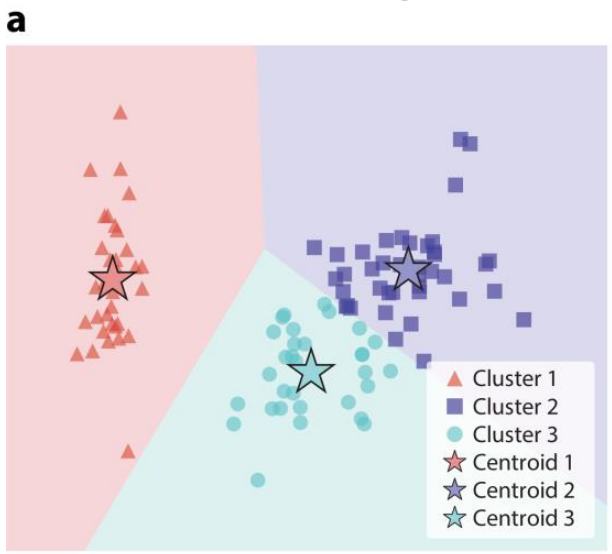


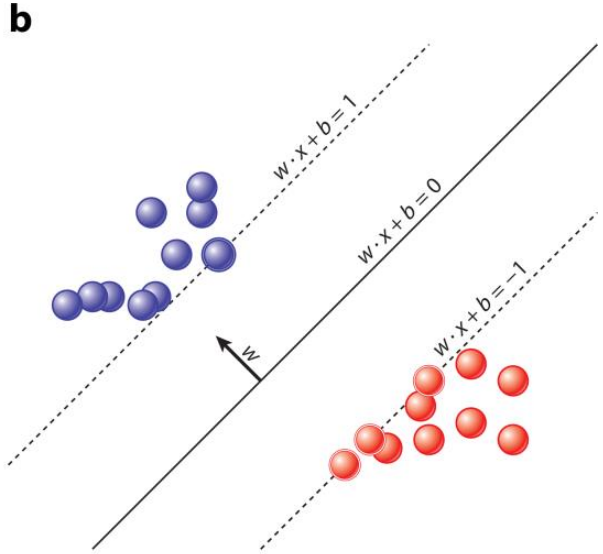
Bioinformatics Team ML applications



Examples of ML algorithms

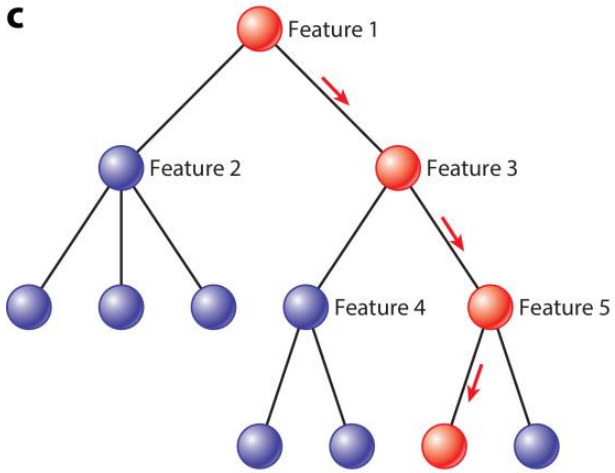
k-means clustering

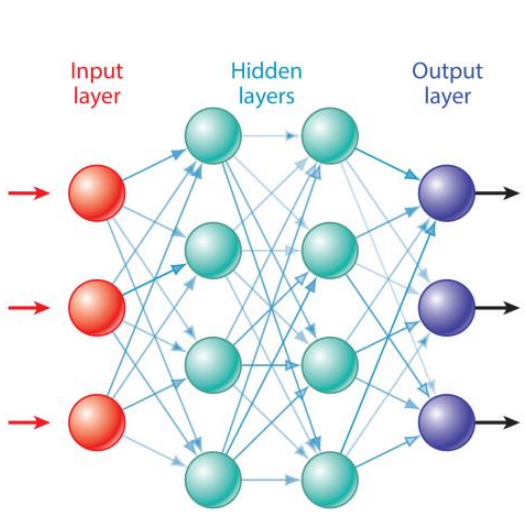




Support vector machine

Decision trees





Neural Networks

Deng X, et al. 2021
Annu. Rev. Food Sci. Technol. 12:513–38

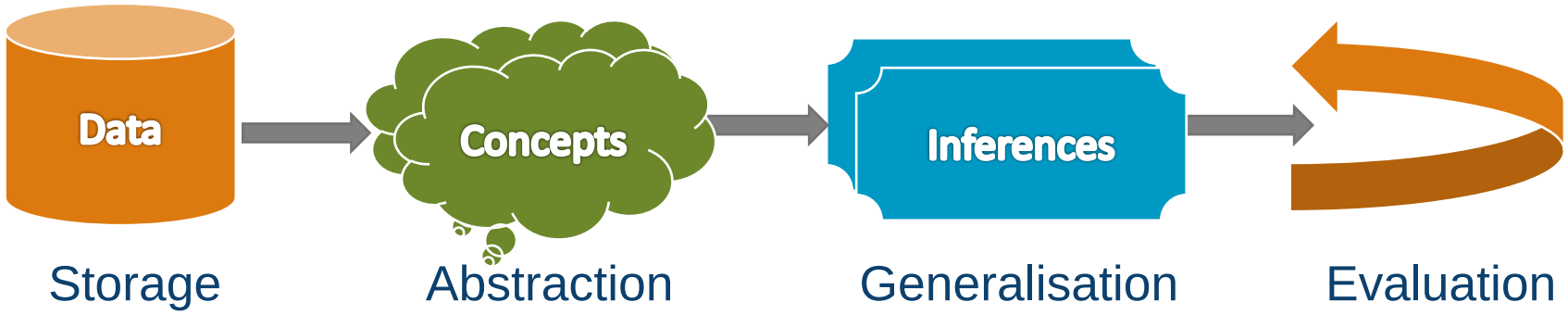
ML limitations

- Machine learning is not in any way a substitute for a human brain…..(yet).
- Machine learning is only as good as the data it learns from.
- A computer still needs a human to motivate the analysis and turn the result into meaningful action.
- It has very little flexibility to extrapolate outside of the strict parameters it learned and knows no common sense.
- Machine learning is most successful when it augments rather than replaces the specialised knowledge of a subject-matter expert.

How machines learn

Learning Process:

1. **Data collection/storage:** all learning must begin with data.
2. **Abstraction**: translates data into broader representations and concepts.
3. **Generalisation**: uses abstracted data to create knowledge and inferences that drive action in new contexts.
4. **Evaluation**: provides a feedback mechanism to measure the utility of learned knowledge and inform potential improvements.



Data abstraction

A computer summarises stored raw data using a **model.**

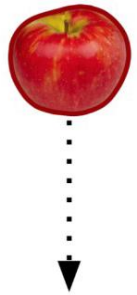
Model fitting = Model training = Data abstraction

Some model types

- Mathematical equations;
- Relational diagrams such as graphs;
- Logical if / else rules;
- Groupings of data known as clusters.

A model can be useful for the discovery of previously unseen relationships among data.

Observations → **Data** → **Model**



Distance	Time
4.9m	1s
19.6m	2s
44.1m	3s
78.5m	4s

Gravitational force

$g = 9.8 m/s^2$

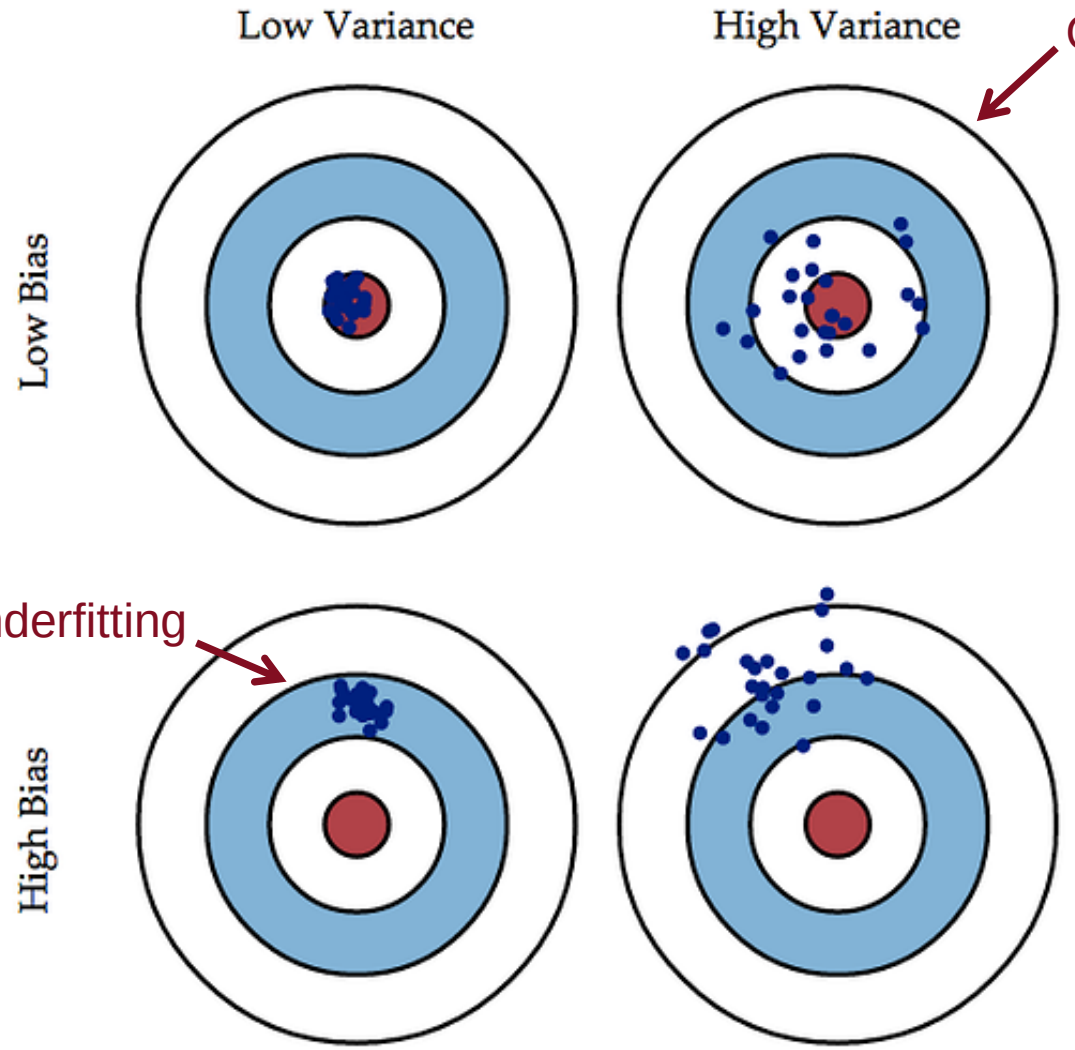
$t=\sqrt{(2d/g)}$

Generalisation

- **In generalisation, the learner tries to limit the patterns it discovers to only those most relevant to its future tasks.**
- Machine learning algorithms generally employ shortcuts that reduce the search space more quickly.
- the algorithm employs **heuristics,** which are educated guesses about where to find the most useful inferences.
- The heuristics employed by machine learning algorithms can introduce a bias.

Generalisation: Bias vs Variance

Bias is an error from erroneous assumptions in the algorithm.

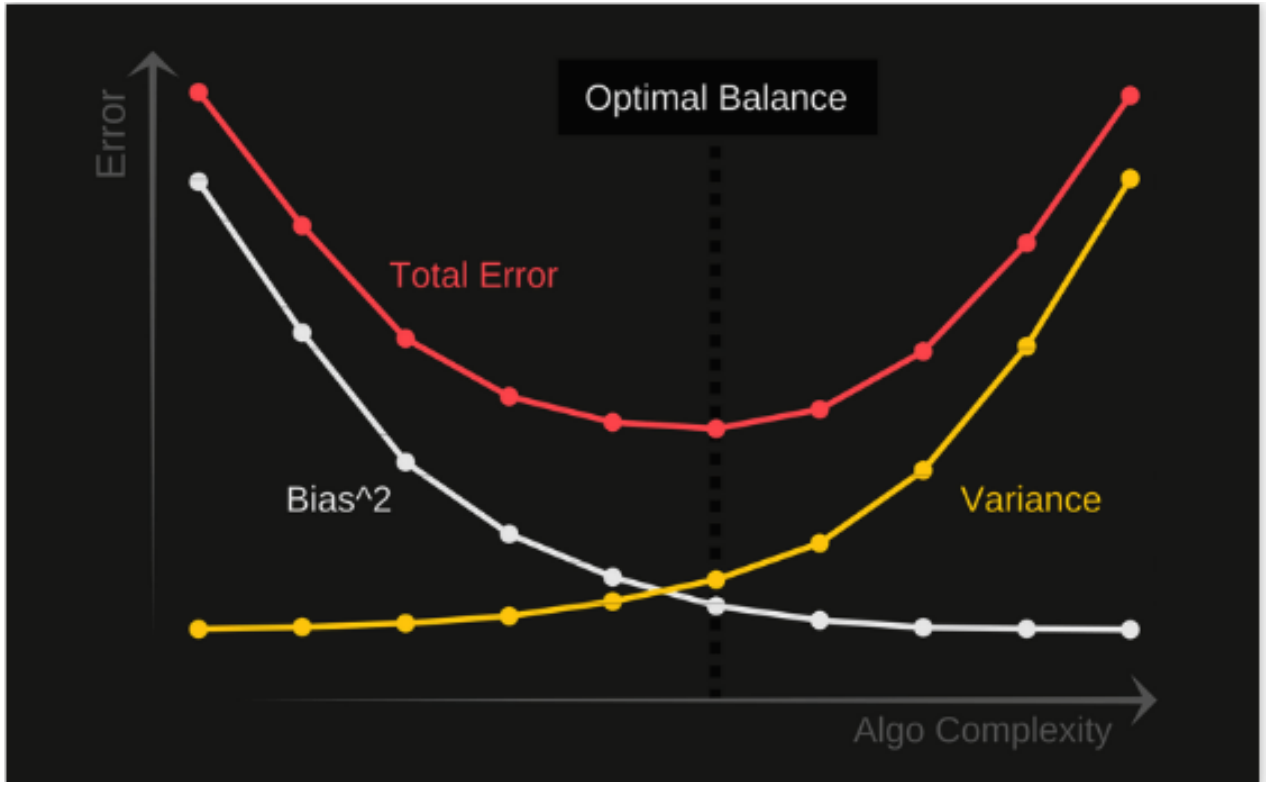


Variance is an error from sensitivity to small fluctuations in the training set

Generalisation: Bias vs Variance

A good model minimizes the total error by finding a balance between bias and variance.

Total Error = Bias^2 + Variance + noise

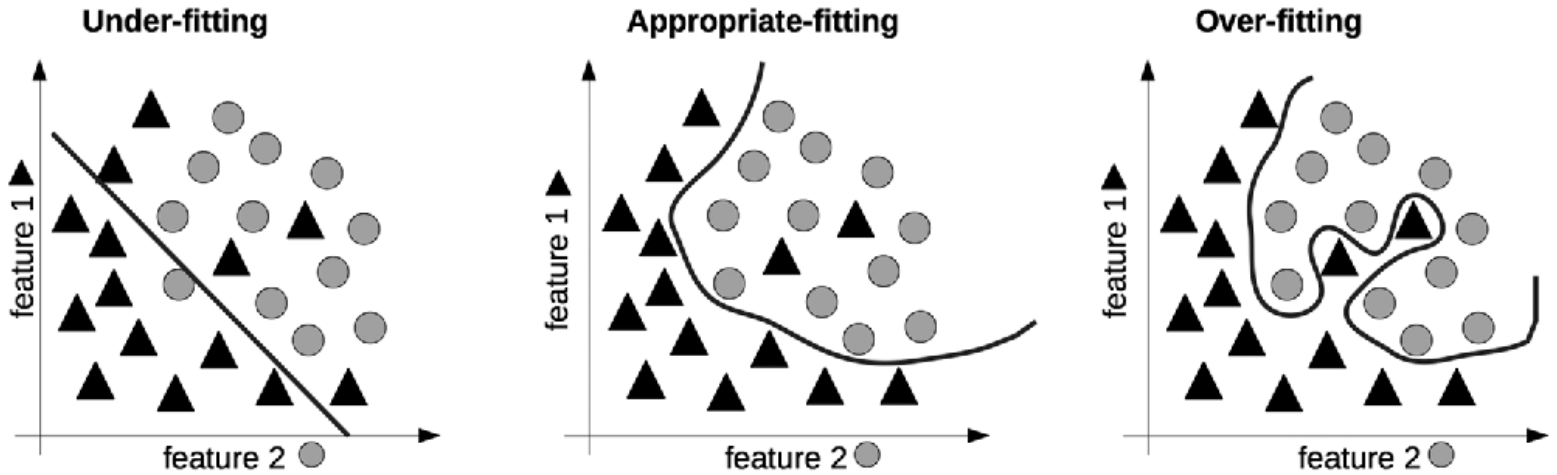


Evaluation

Evaluation: occurs after a model has been trained on an initial training dataset. Then, the model is evaluated on a new test dataset in order to judge how well its characterisation of the training data generalises to new, unseen data.

In parts, models fail to perfectly generalise due to the problem of **noise.**

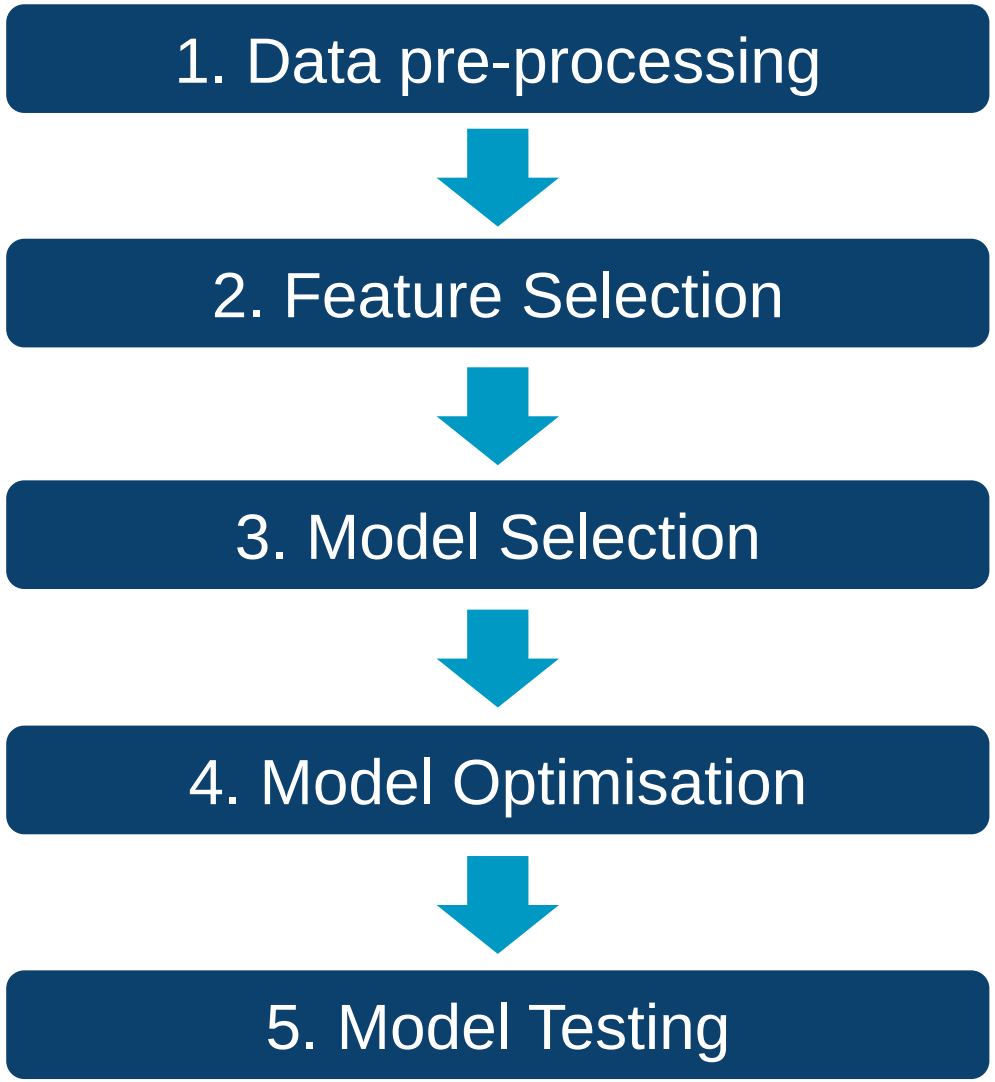
Trying to model noise causes **overfitting**.



High bias

High Variance

Typical ML Workflow



Applications of ML algorithms

a) Predictive models: discover and model the relationship between the **target** feature and the other features → **supervised learning.**

b) Descriptive models: aiming at gaining insights from summarising data in new and interesting ways → **unsupervised learning.**

c) Meta-learning analysis: algorithms focused on how to learn more effectively.

Meta-learners use the result of some learning to inform additional learning.

Matching ML algorithms to learning tasks

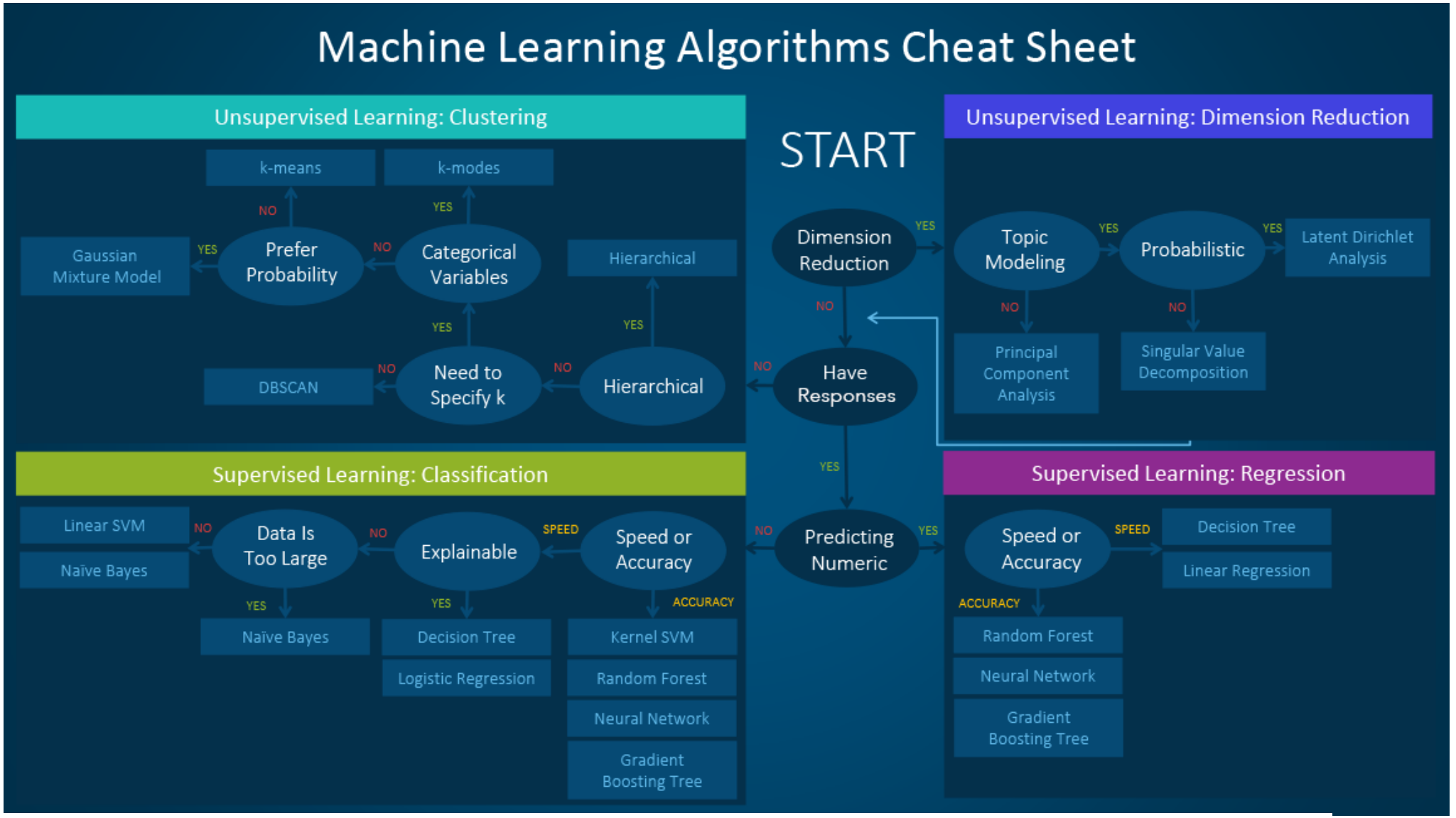
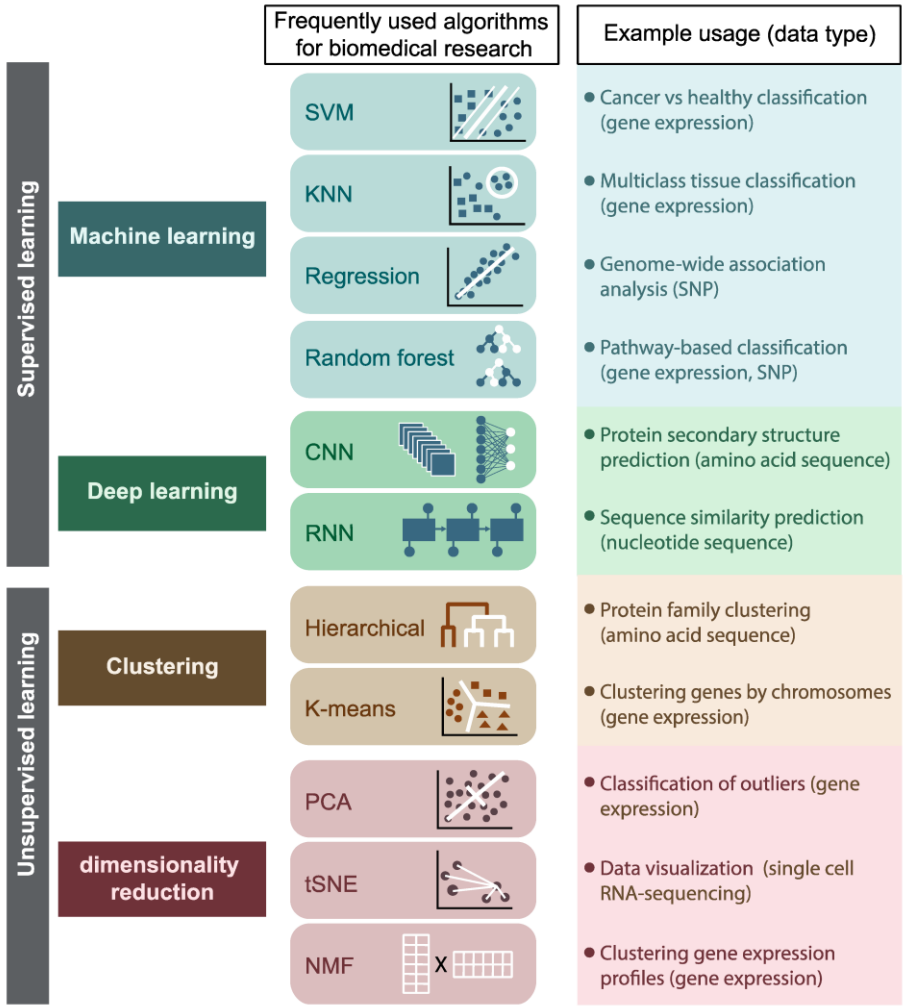
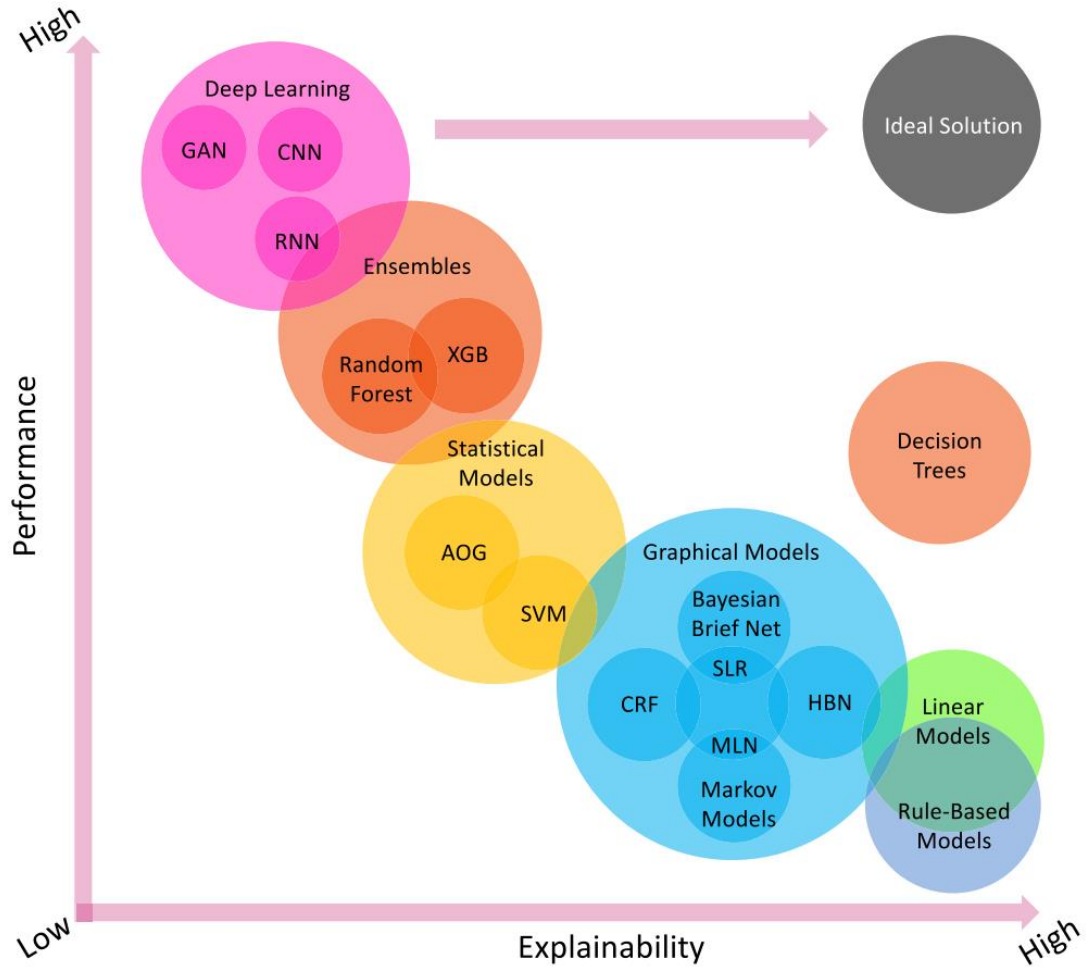


Image source: https://blogs.sas.com/content/subconsciousmusings/2020/12/09/machine-learning-algorithm-use/

Matching ML algorithms to learning tasks





Schema source: Auslander N., et al. *Int. J. Mol. Sci.* 2021

Schema source: Yang G., et al. *Information Fusion.* 2022

Cranfield
Environment
and Agrifood

www.cranfield.ac.uk

T: +44 (0)1234 750111

@cranfielduni

@cranfielduni

/cranfielduni