Recap

Adrian Jackson, EPCC, The University of Edinburgh
a.jackson@epcc.ed.ac.uk

August 2025

www.archer2.ac.uk

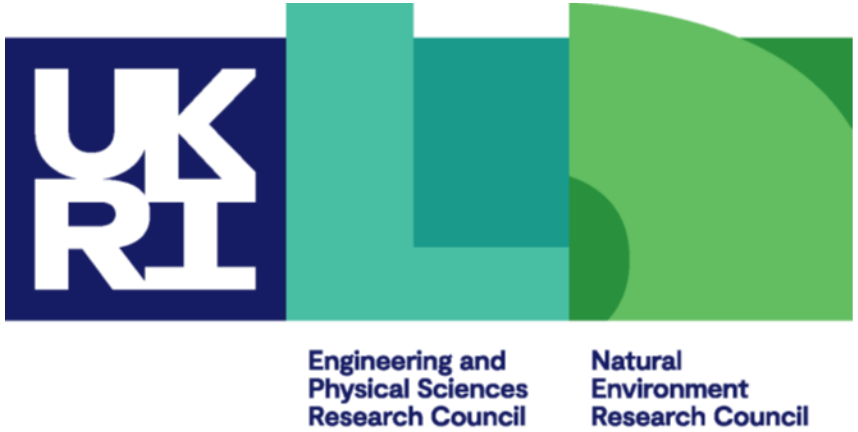
archer2

epcc

Reusing this material

Partners

epcc



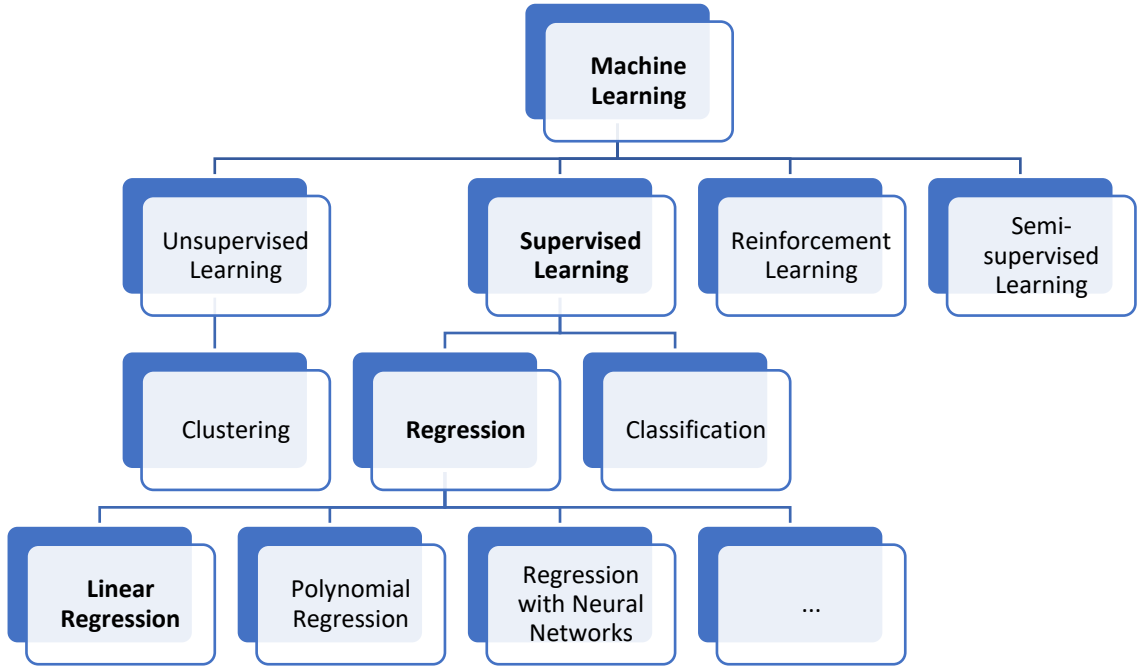
epcc

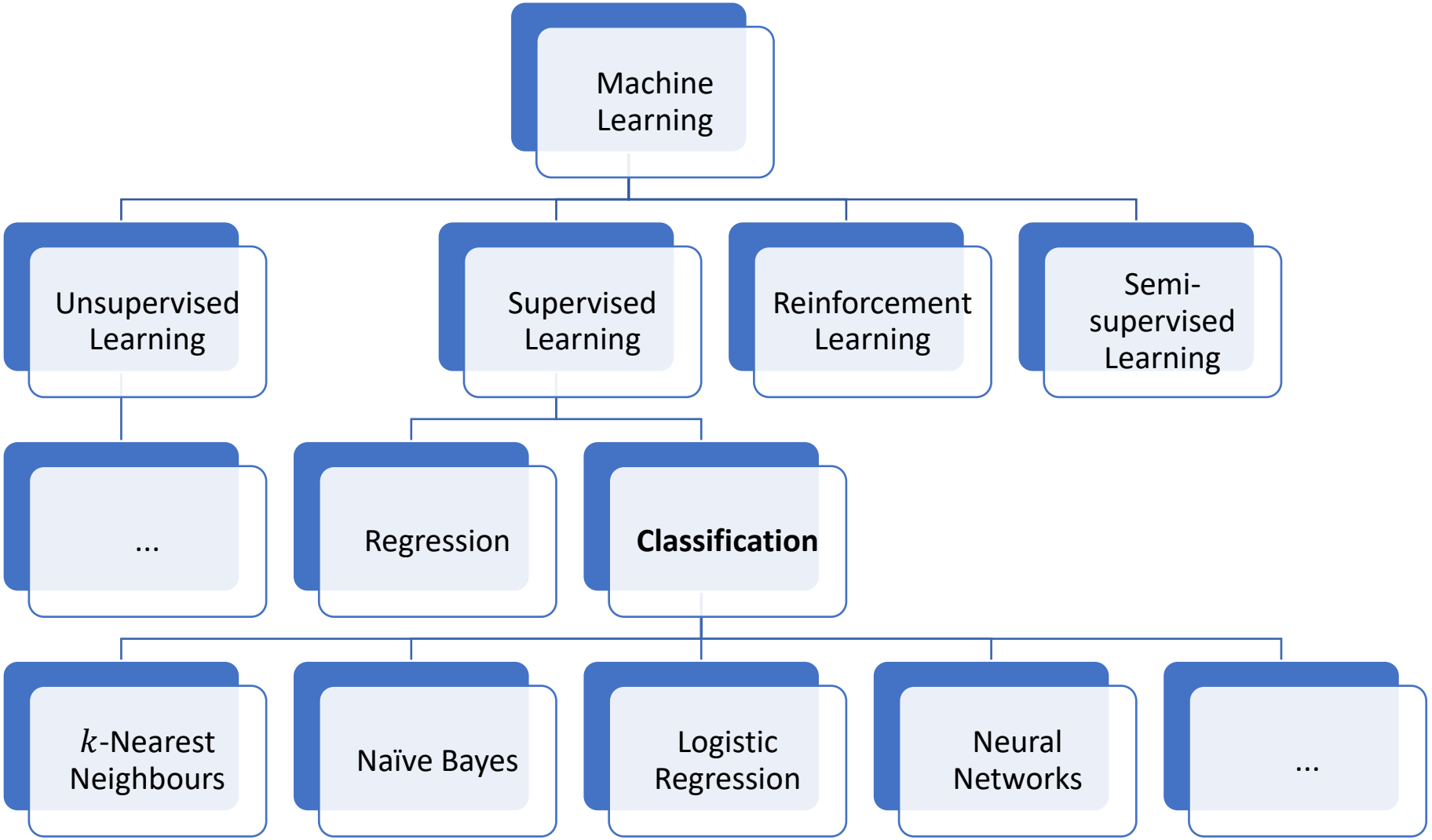
THE UNIVERSITY of EDINBURGH

Hewlett Packard Enterprise

Linear Regression

- Linear Regression predicts a continuous variable from one or more variables
- One of the simplest predictive models
- Existing data is used to create a **linear model**

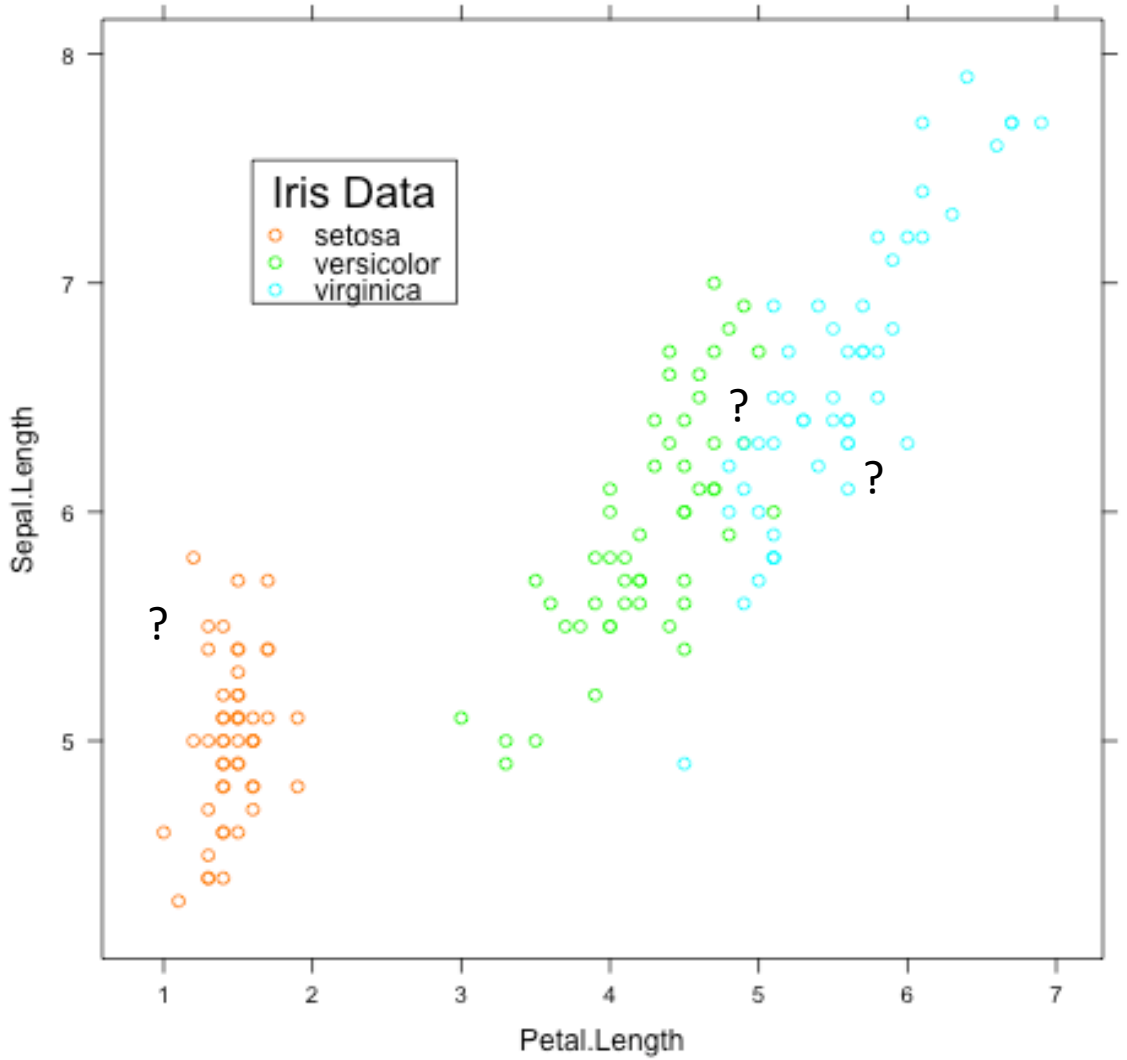


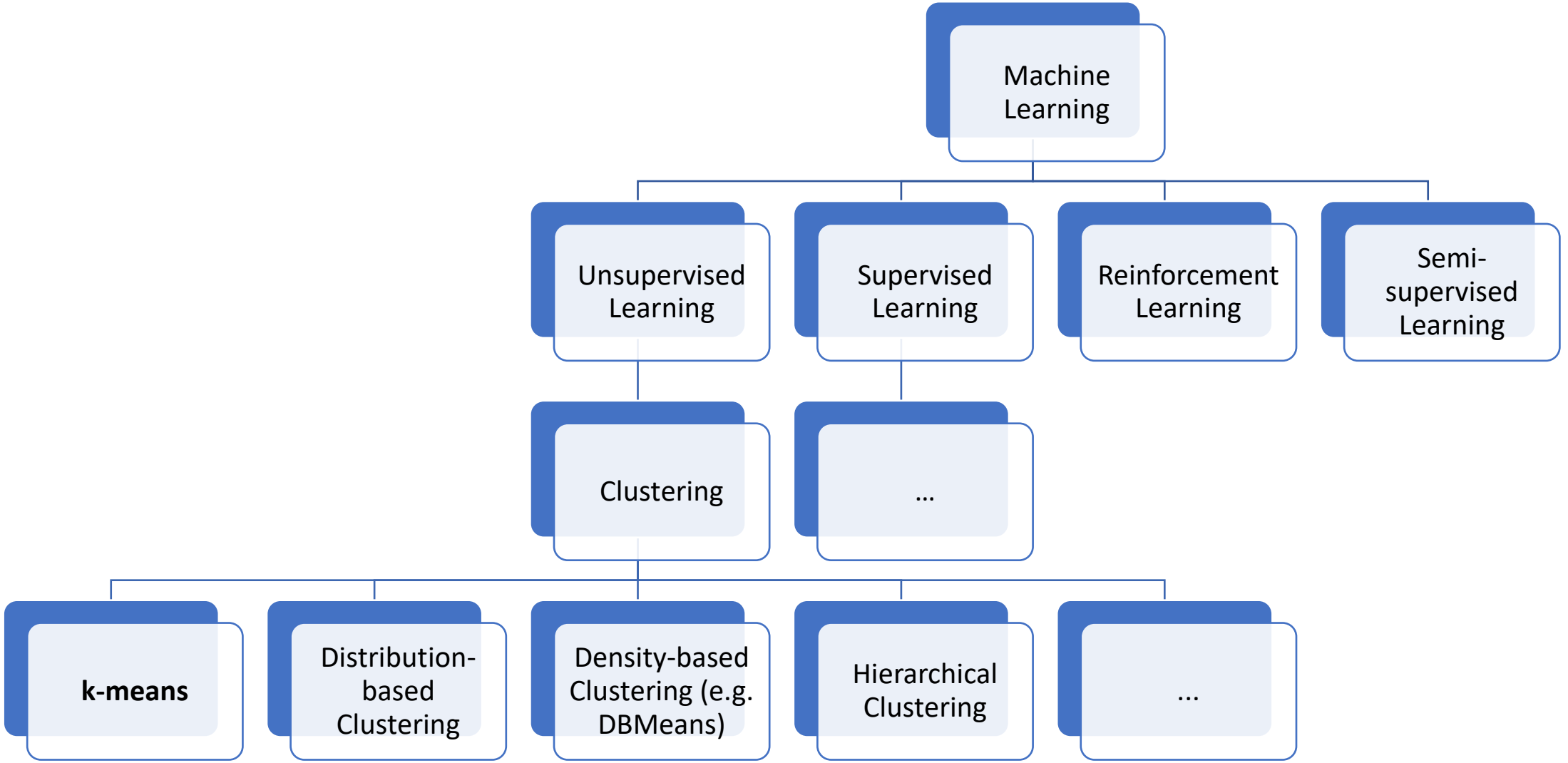

Machine Learning
Unsupervised Learning
Supervised Learning
Reinforcement Learning
Semi-supervised Learning
...
Regression
Classification
k-Nearest Neighbours
Naïve Bayes
Logistic Regression
Neural Networks
...

Classification: problem overview

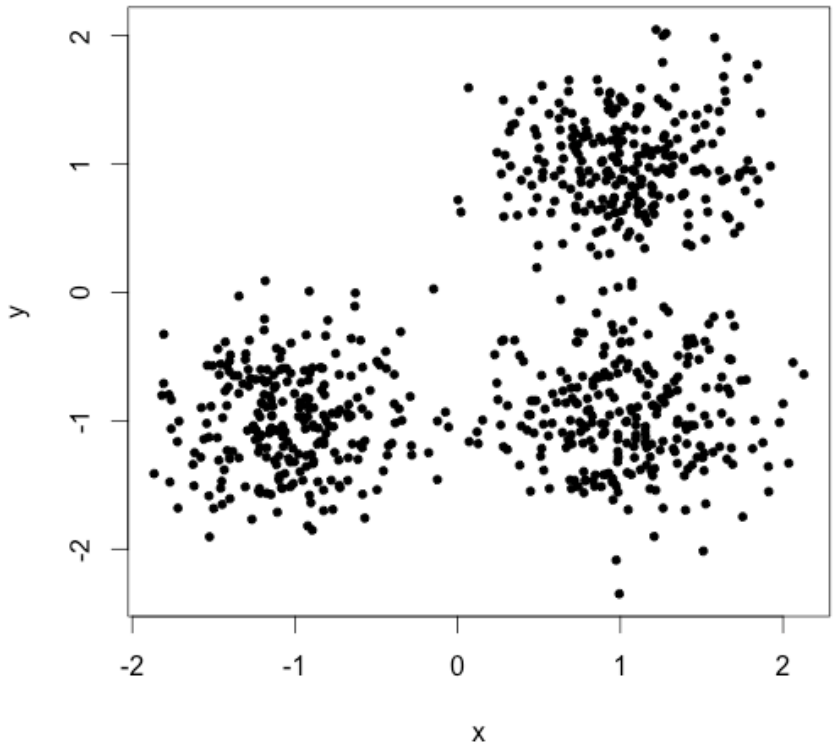
- Input is one or more features
  - numerical or categorical
- Output is a category (or class)

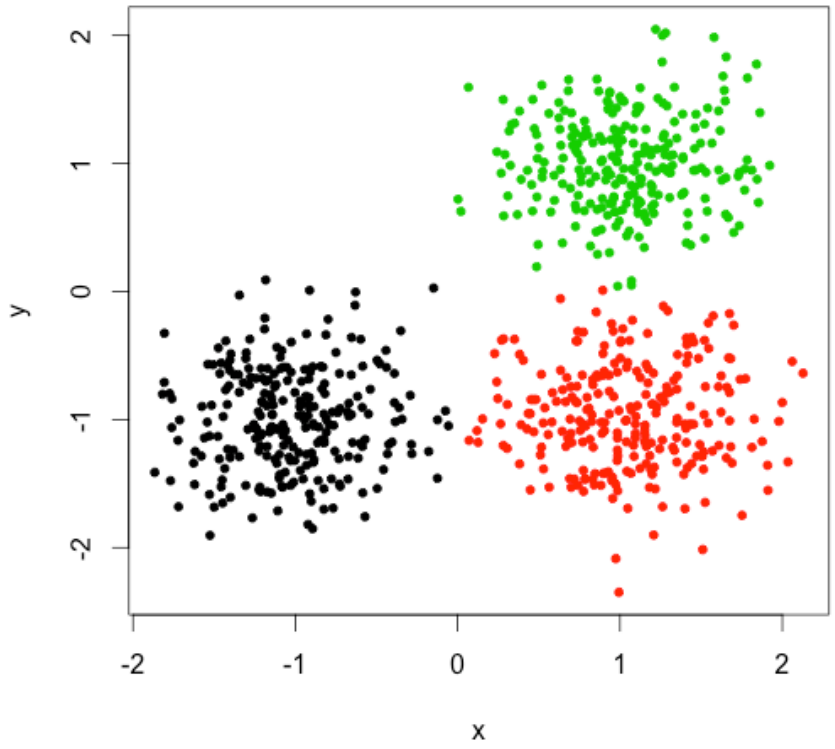
Iris image courtesy of David Berger under a CC-BY licence




Machine Learning
Unsupervised Learning
Supervised Learning
Reinforcement Learning
Semi-supervised Learning
Clustering
...
k-means
Distribution-based Clustering
Density-based Clustering (e.g. DBMeans)
Hierarchical Clustering
...

Clustering – the problem





Data science and machine learning

- Lots more to data science than model building
  - Data collection, cleaning, analysis, visualisation, etc…
- Model choice and training/building requires understanding your probably
  - Might be simple approaches that work well
  - With complex problems or large datasets machine learning approaches can be useful