Imperial College London

Introduction to Machine Learning

Machine Learning
QSRI summer school – July 2022

Sarah Filippi

Department of Mathematics
Imperial College London

Outline

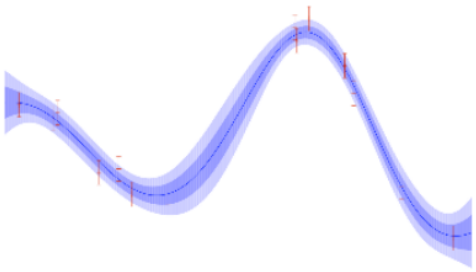
Structure

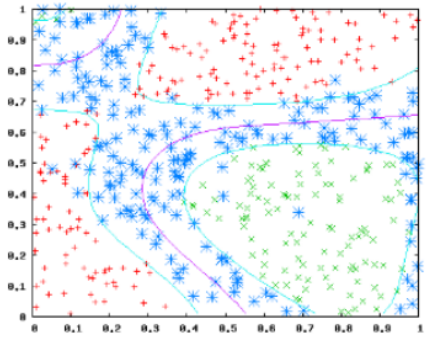
- Lectures
  - Tuesday 10am to 12pm and 1 to 2pm
  - Wednesday - 10am to 12pm and 1 to 2pm
  - Thursday - 10am to 12pm and 1 to 2pm
  - Friday - 10am to 12pm and 1 to 2pm
- Tutorials
  - Tuesday and Thursday
  - Group 1 - 2:15pm to 3:45pm
  - Group 2 - 3:45pm to 5:15pm
- Lecturers
  - Dr Sarah Filippi
  - Dr Xenia Miscouridou

Course Objectives

1. To provide an overview of machine leaning
2. To describe some common machine learning approaches for supervised and unsupervised learning techniques
3. To explain how these algorithms are trained given some data
4. To discuss the concept of generalisation
5. To learn how to apply supervised and unsupervised learning algorithms to real data in `R`

Syllabus

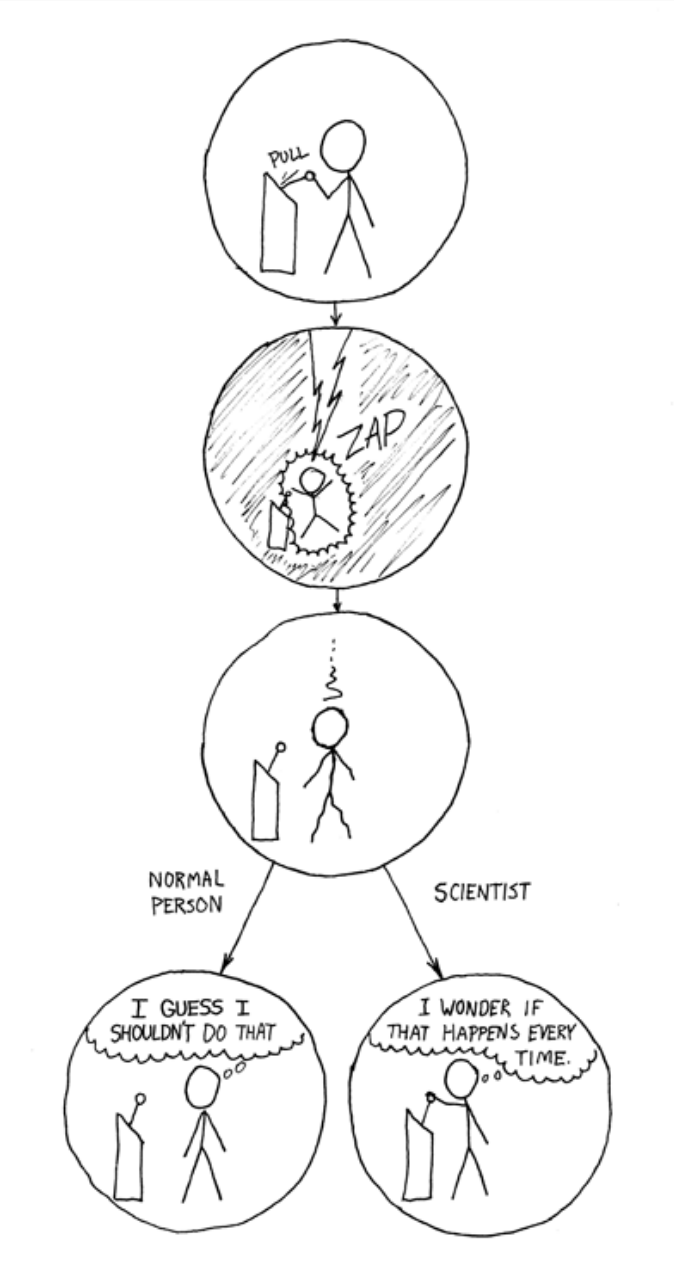




- Supervised vs unsupervised learning
- K-nearest neighbour classifier
- Regression using parametric models
- Generalisation and evaluating learning methods
- Dimensionality reduction
- Clustering
- Classification using logistic regression
- Neural network
- Decision trees and random forest
- Ethic in Machine Learning

Outline

What is Machine Learning?



Humans are life long learners.

As with humans, machine learning algorithms "learn" from examples/experience how to carry out specialised tasks.

Learning is not learning by heart - the difficulty lies in generalising the behaviour to novel situations.

What is Machine Learning?

Field of study that gives computers the ability to learn without being explicitly programmed. Arthur Samuel, 1959

Any computer program that improves its performance at some task through experience. Tom Mitchell, 1997

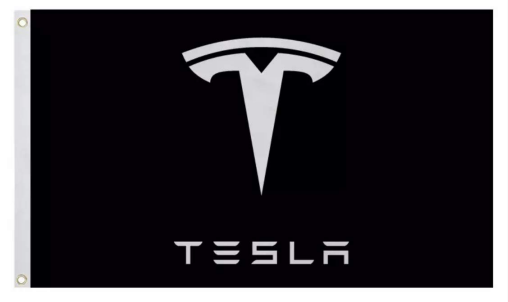
To develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest. Kevin Murphy, 2012

Some areas of application

Spam filtering

Recommending system

Fraud detection



Self driving cars

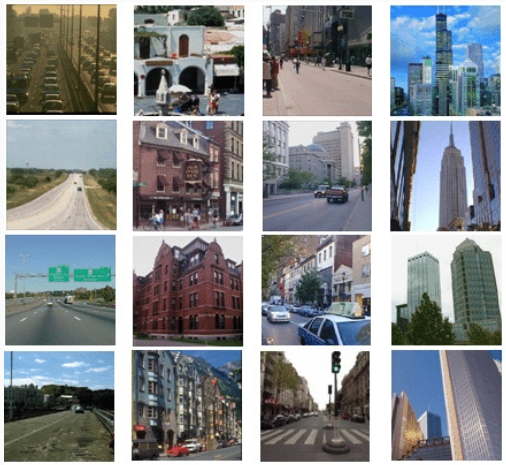
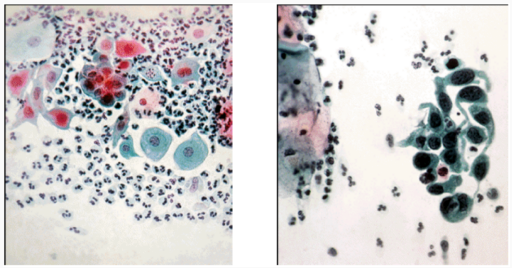


Image recognition



Medical diagnosis

Applications of Learning

Learning to predict real-valued measurements

- House or stock prices given some economic variables
- Medical diagnoses given some clinical/genetic variables

Learning to assign objects to one of many predefined classes

- Classify webpages given their text and image content
- Recognises faces in photos and videos
- Speech and handwriting recognition

Learning to discover novel or unusual patterns

- Discover previously unknown tumour subtypes
- Detecting credit card fraud

Different Types of Learning

Supervised Learning algorithms are provided with a training set of features and associated labels.

Unsupervised Learning algorithms are provided with a training set of data features, and are tasked with finding useful representations or structure in the data.

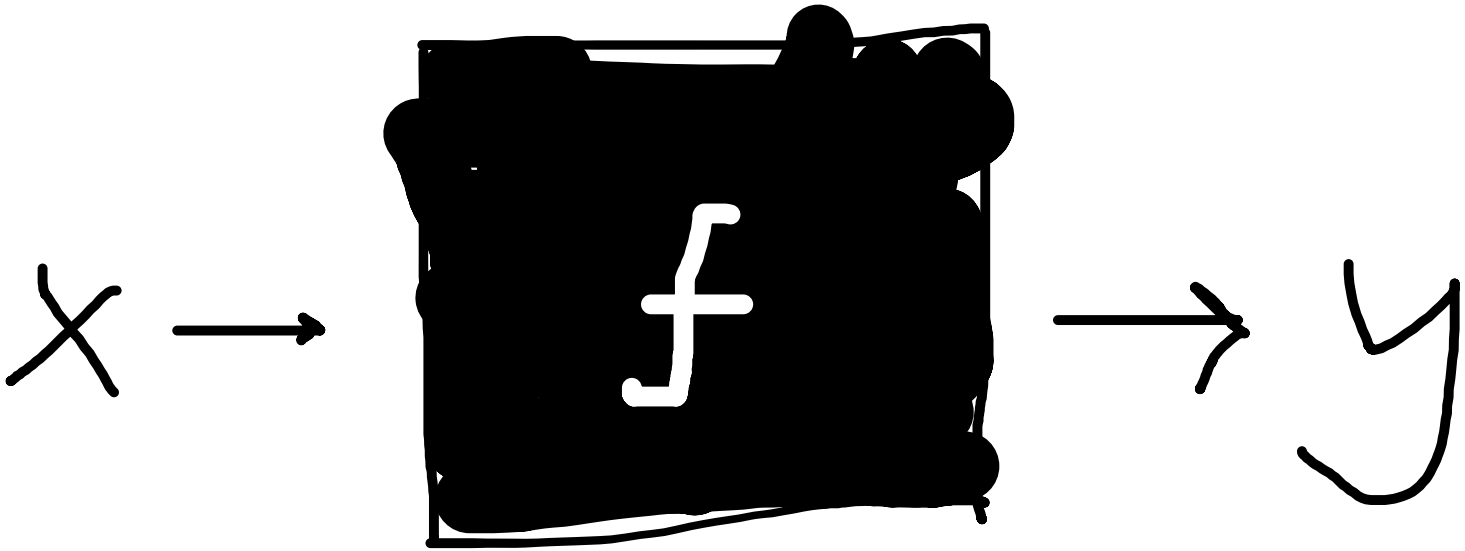
Reinforcement Learning algorithms interact with an environment, and are provided with features of the environment, as well as reward signal. The algorithm uses this feedback from the environment to learn the optimal way to interact with the environment.

Supervised learning

Given

- Inputs, also known as: independent variables, predictors, covariates, patterns, x, $\boldsymbol{X}, \ldots$
- Outputs, also known as: dependent variables, responses, labels, y, $Y, \ldots$

Learn a function (algorithm, black box, decision rule, classifier, probability distribution)

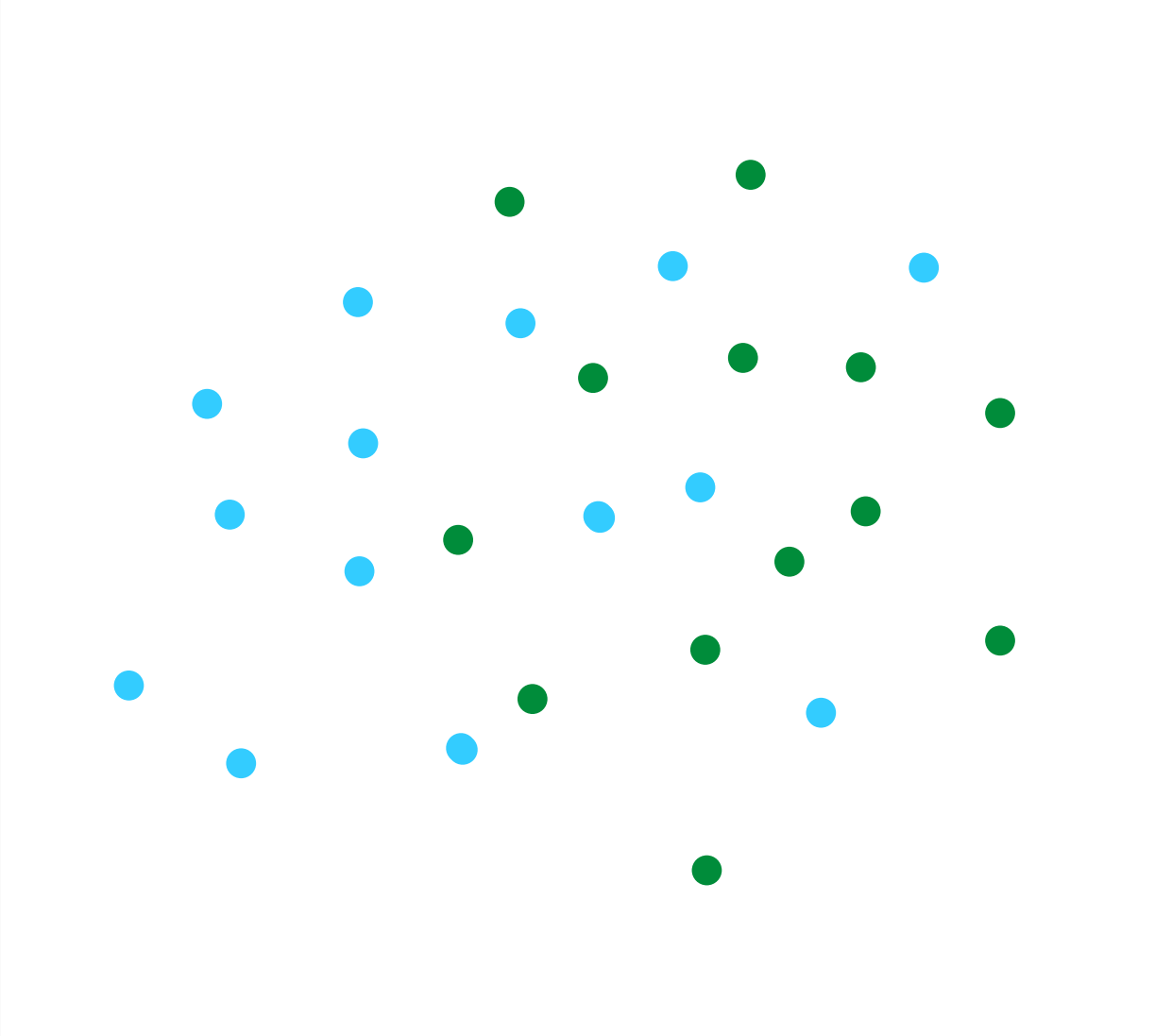


Supervised learning

Regression - supervised learning in which the labels are continuous values.

Classification - supervised learning in which the labels are discrete values.

Supervised learning: k-nearest neighbors (k-NN)



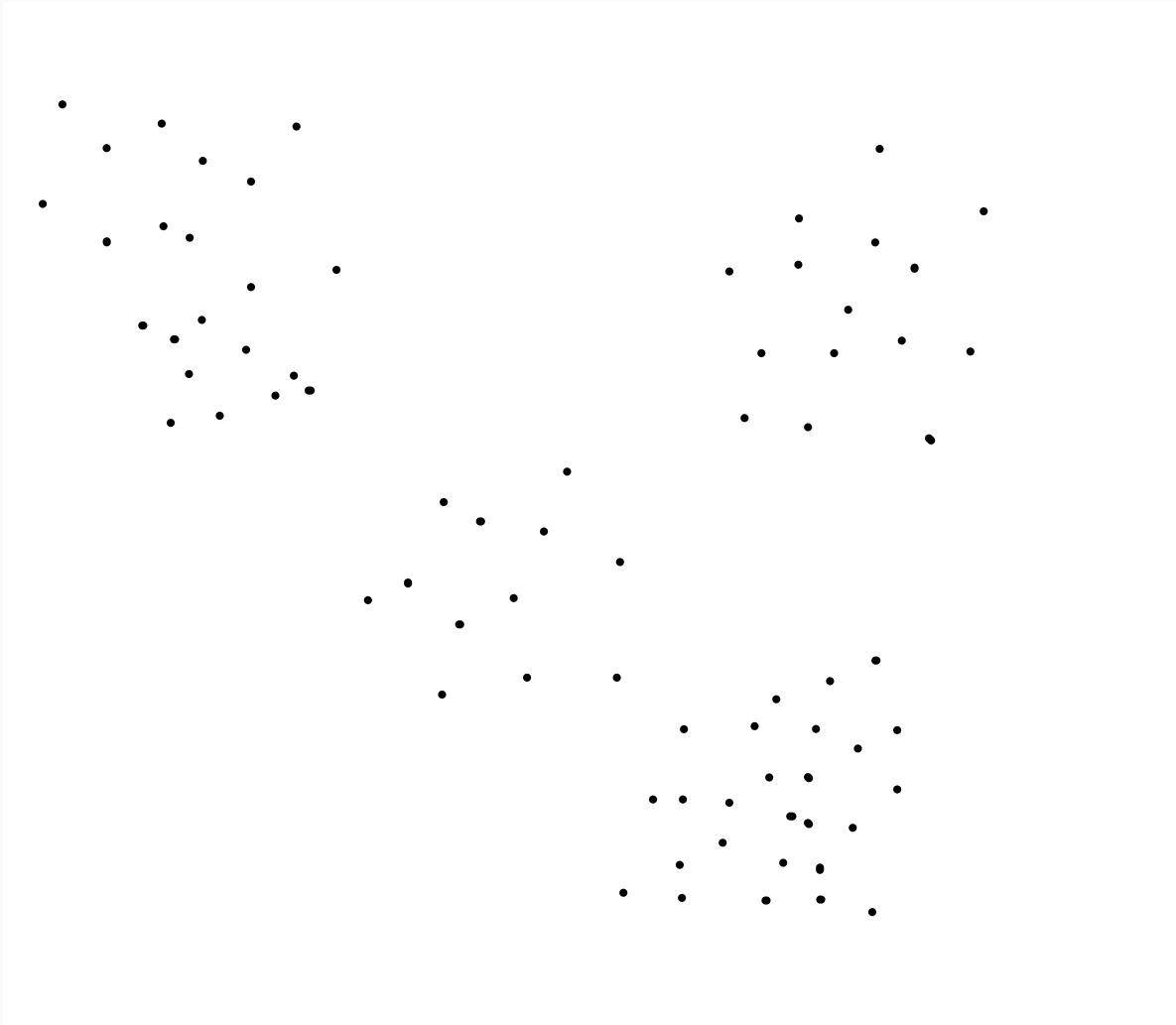
How would you classify the new point? (www.menti.com with code 9018 6653)

Note: k-NN can also be used for regression

Unsupervised learning

Unsupervised learning tasks: clustering, data mining, dimensionality reduction, ...

Illustration of clustering:



Learning Machine Learning

Machine learning is a huge and incredibly useful topic.

Fortunately, there are basic underlying issues and principles that arise again and again, many of them statistical!

We will focus on simpler models that illustrate the main principles, that are also pertinent for more complex models.

Lots to cover - so let's begin!