

# COMS30035, Machine learning: Introduction

Rui Ponte Costa

Department of Computer Science, SCEEM  
University of Bristol

September 27, 2021

# Disclaimer

These are unusual years. COVID-19 means that we had to change our teaching processes. So please bear with us – we are trying our best.

# Agenda

- ▶ What is machine learning?
- ▶ Examples
- ▶ How does it relate to other units?
- ▶ What will we cover?
- ▶ Unit outline
- ▶ Assessments/coursework
- ▶ Labs

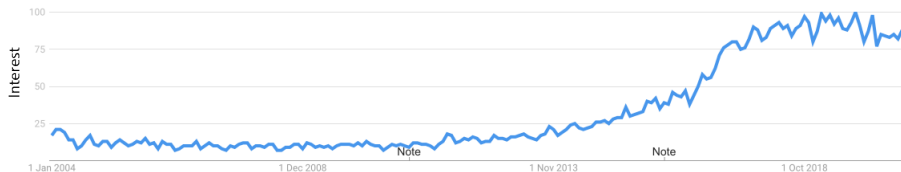
# What is machine learning?

- ▶ Machine Learning (ML) is the study of computer **algorithms** that **learn a model** of **data**. These models are then used to make **predictions** and better **understand** the problem at hand.
- ▶ It is typically grounded on statistical science<sup>1</sup> and seen as a subfield of Artificial Intelligence.

---

<sup>1</sup>Remember to review some key statistical concepts. See Iain Murray's crib-sheet with a revision of math commonly used in ML here.

# Machine learning interest [Google trends]



# Machine learning interest [Google trends]

In contrast to Computer Science, Machine Learning is increasing in popularity.

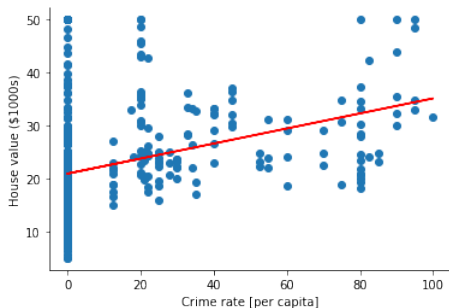
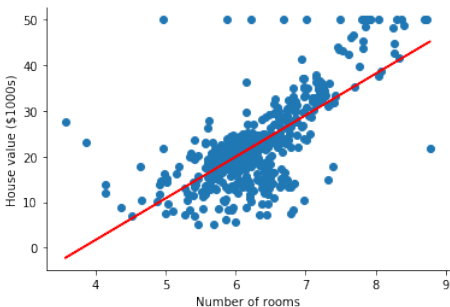


# Examples: Linear regression

When we regress one variable against another using a linear model.

**Pros:** Simple model **Cons:** Simple model

Example: Can we predict property prices in Boston?



# Examples: Classification

When we want to automatically separate data into classes.

**Pros:** A wide range of methods that work well (e.g. ANNs or SVMs).

**Cons:** Needs labelled data (i.e. sample X is an example of a dog).

Example: Is it a dog or a bagel?



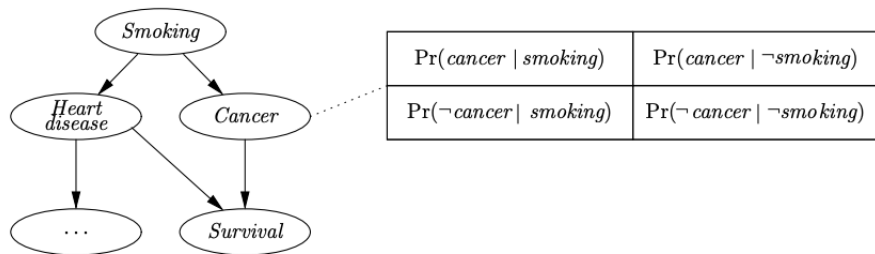


# Examples: Graphical models

When we use a graphical model to represent an explicit probabilistic model.

**Pros:** Easy to interpret and explicit notion of uncertainty. **Cons:** Can be hard to estimate.

Example: What are the changes of survival for a smoker?

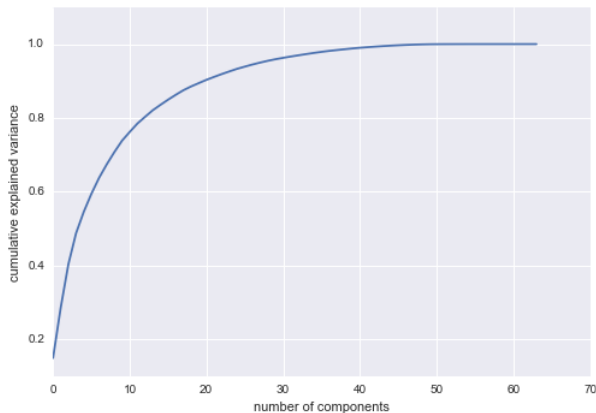


# Examples: Dimensionality reduction

Used to compress data to a much smaller number of dimensions.

**Pros:** Easy to compute (e.g. using PCA). **Cons:** Loss of features semantics.

Example: How many features/components are enough to explain 80% of the variance in a given dataset? [More here.](#)



# Examples: Generating text in natural language

**Pros:** Can give impressive results (e.g. using deep learning).

**Cons:** Models can be hard to interpret.

Example using a Transformer (a type of neural network):

**Input:** "Machine Learning is the study of computer algorithms that learn a model of data."

**Output/completion:** "Traditional AI, such as natural language processing, machine translation, image recognition, and machine translation relies heavily on well-defined data, and can perform well without much further manipulation of the data. Typically, a machine learning system recognizes patterns of data and then uses those patterns to predict outcomes of a given scenario. (see prediction)"

Try it out **[app.inferkit.com/demo](https://app.inferkit.com/demo)**!

# How does it relate to other units

- ▶ This unit builds directly on previous units
  - ▶ Data Science for CS (previously known as SPS) [\[introduction to data science in Python\]](#)
  - ▶ Algorithms I and II [\[design/analyse algorithms, and data structures\]](#)
  - ▶ Math for CS I and II [\[algebra and statistics\]](#)
- ▶ And is a building block for more advanced units (4th year):
  - ▶ Advanced ML
  - ▶ Applied Deep Learning (neural networks)
  - ▶ Information Processing & the Brain (ML and neuroscience)
  - ▶ Cloud Computing and Big Data (ML on the cloud)
  - ▶ Advanced Computer Vision (ML for vision)
  - ▶ Applied Data Science (ML and data management)

# What we will cover

We will focus on the following textbook

- ▶ **Bishop, C. M., Pattern recognition and machine learning (2006). Available for free [here](#).**

With some sections from:

- ▶ Murphy, K., Machine learning a probabilistic perspective (2012). The book is also freely available [here](#).

# What we will cover

1. Part 1: Overview, ML concepts, revisiting regression and neural networks [» Rui Ponte Costa, unit director]
2. Part 2: SVMs, Graphical models, mixture models, EM, Continuous latent models (PCA, ICA) [» James Cussens]
3. Part 3: Sequential data (RNNs, HMMs), model ensembles [» Edwin Simpson]



# Unit Outline

On the unit BB page; and [uob-coms30035.github.io](https://uob-coms30035.github.io)

Weeks	First lecture	Second lecture	Labs	Q&A class
1	Introduction	Machine learning concepts	L1: Revision of Jupiter Notebook and ML libraries	General questions about the unit
2	Linear models and neural networks	Kernel machines	L2: Linear models, nnets and SVMs	ML concepts, linear models and nnets
3	Introduction to graphical models	Bayesian ML using graphical models	L3: Probabilistic graphical models	Probabilistic graphical models
4	k-means and mixtures of Gaussians	The EM algorithm	L4: k-means and EM	The k-means and EM
5	PCA	kernel PCA and ICA	L5: PCA and ICA	PCA and ICA
6	Sequential data and hidden Markov models	Sequential data and linear dynamical systems	L6: Hidden Markov Models	Modelling sequential data
7	Ensemble methods	Probabilistic model combination	L7: Ensemble methods	Combining models using ensembles and probabilistic methods
Coursework weeks				
11	Review/QA part I (Rui)	Review/QA part II (James)	-	Review/QA Part III (Edwin)
12	Review week			

# Online/in person teaching

Every week you will get two teaching components:

## 1. **Asynchronous/during your own time:**

- ▶ 2 recorded lectures (available on BB>Recordings)
- ▶ Pointers to textbook (Bishop book)

## 2. **Synchronous/during the time set in your timetable:**

- ▶ 1 live lecture (QA/problem class) [1 hour; **Tue 9-10am**]
- ▶ 1 lab based on the lectures taught on that week [three hours; **in person** (or online if you cannot be in person)]



# Microsoft Teams

All our synchronous activities will be run on Microsoft Teams:  
**COMS30035: Machine Learning (Teaching Unit)**

## How to join?

1. Install Microsoft Teams.
2. You should be already added to our Team.
3. Enter Team and the appropriate channel.

[Run BB/Teams demo]

# Assessments

- ▶ **Option 1, 100% Coursework<sup>2</sup>:** A small project involving a ML problem [fulltime in weeks 8-10]  
Coursework released at beginning of week 8.  
Discussion with others is encouraged, but submissions need to be unique [**plagiarism is taken seriously!**]
- ▶ **Option 2, 100% Exam** [multiple choice]
- ▶ Labs are only formative, but we highly encourage you to finish all of them, as it will massively help you preparing for either option.

---

<sup>2</sup>It is possible to change options until end of Week 2, but remember that you can only do 2 units via Option 1.

# Labs

- ▶ Thursday 9am - 12pm [as in timetable]: In person (or online in bubbles)
- ▶ Lab Environment [Jupyter + Python] <sup>3</sup>



- ▶ **Lab Work:**
  - ▶ Work proactively offline and/or online.

---

<sup>3</sup>For those doing the CS undergrad, we will follow a similar setup to Data Science for CS.

# Labs: Important!!

- ▶ Main source of 1:1 support will be from the TAs in the labs!
- ▶ Labs are in person **unless**:
  - ▶ for a COVID related have to do them online, in that case let the lecturer know
- ▶ Labs are essential for a good understanding of ML!

# Your TAs

That you will interact with through Teams:

- ▶ Stefan Radic Webster
- ▶ Will Greedy
- ▶ Dabal Pedamonti
- ▶ Alexander Quessy
- ▶ Amirhossein Dadashzadeh
- ▶ Tamsin Huggins

# Tasks

- ▶ Live lecture (Tue 9-10): Bring questions about the structure of the unit
  - ▶ You can use the Teams QA>Ask Question system to ask question before hand.
- ▶ Next Lab (Week 1): Jupyter Notebook and intro to Python-Sklearn
  1. Go to MVB computer labs (2.11) / Or join your online Teams bubble
  2. See link to lab 1 on BB
  3. Ask TAs for help!