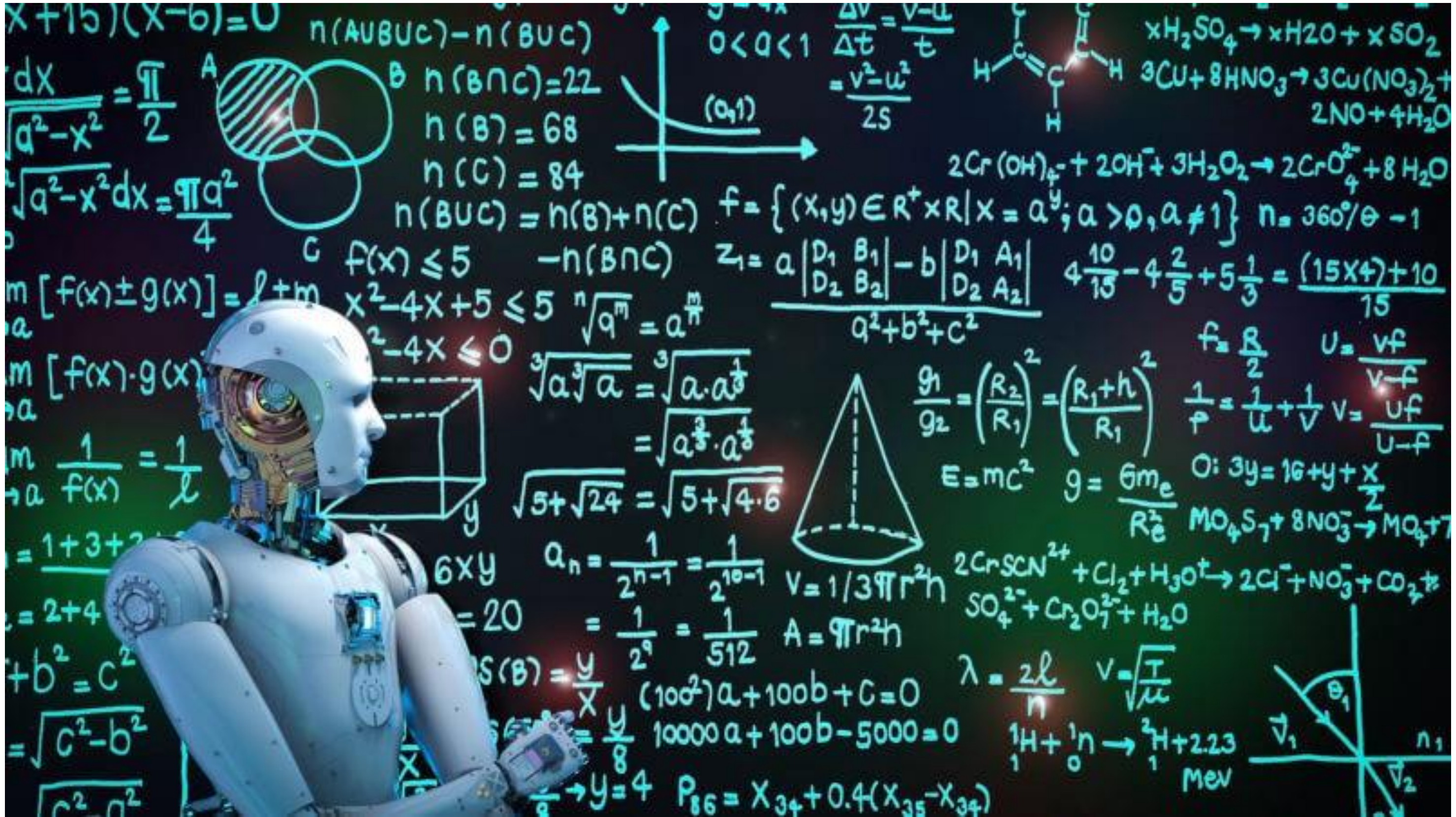
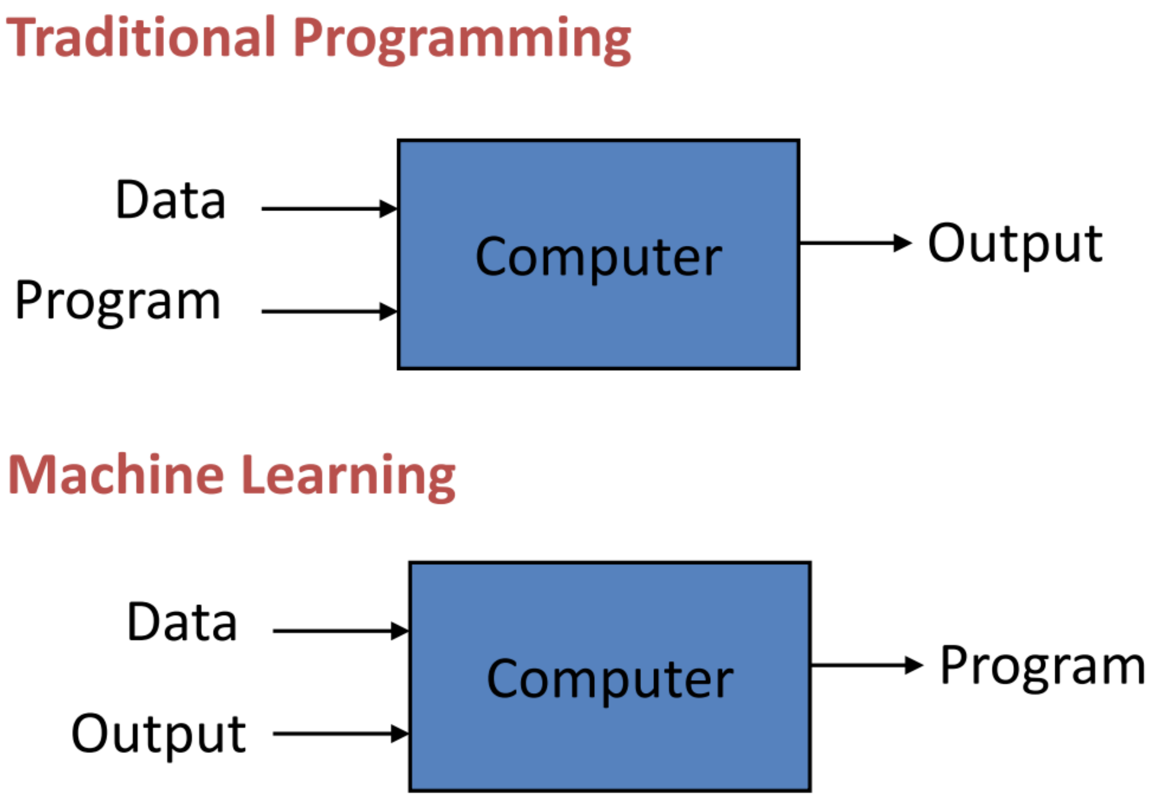
Linear Algebra for Machine Learning – An Introduction

Bahavathy Kathirgamanathan
University College Dublin

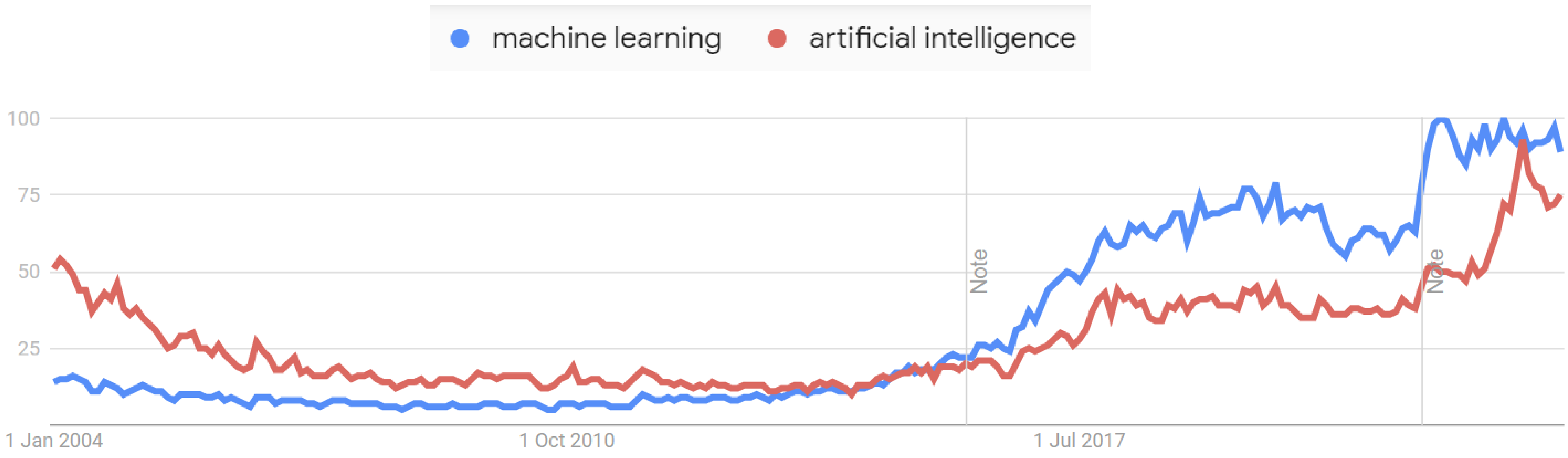


What is Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel (1959)

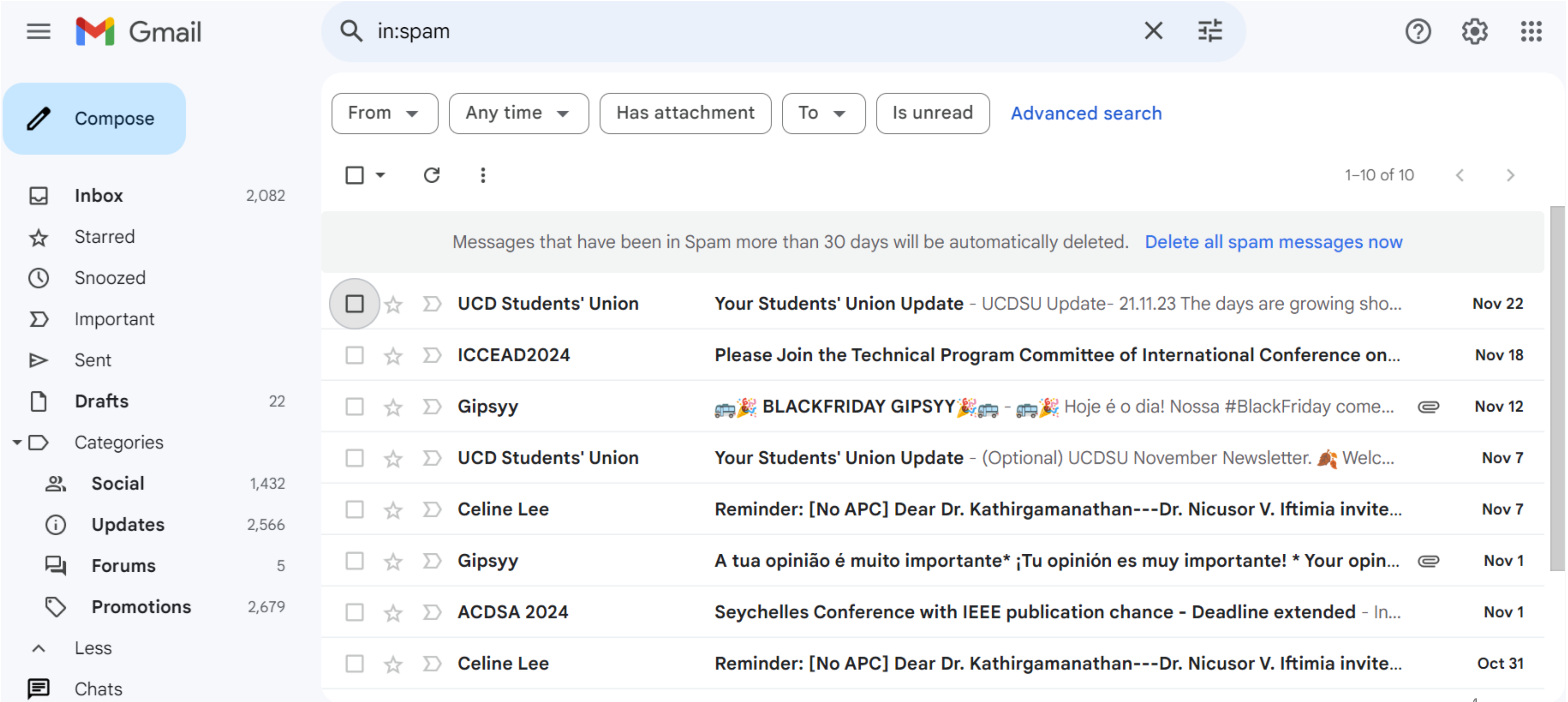


Machine Learning and Artificial Intelligence

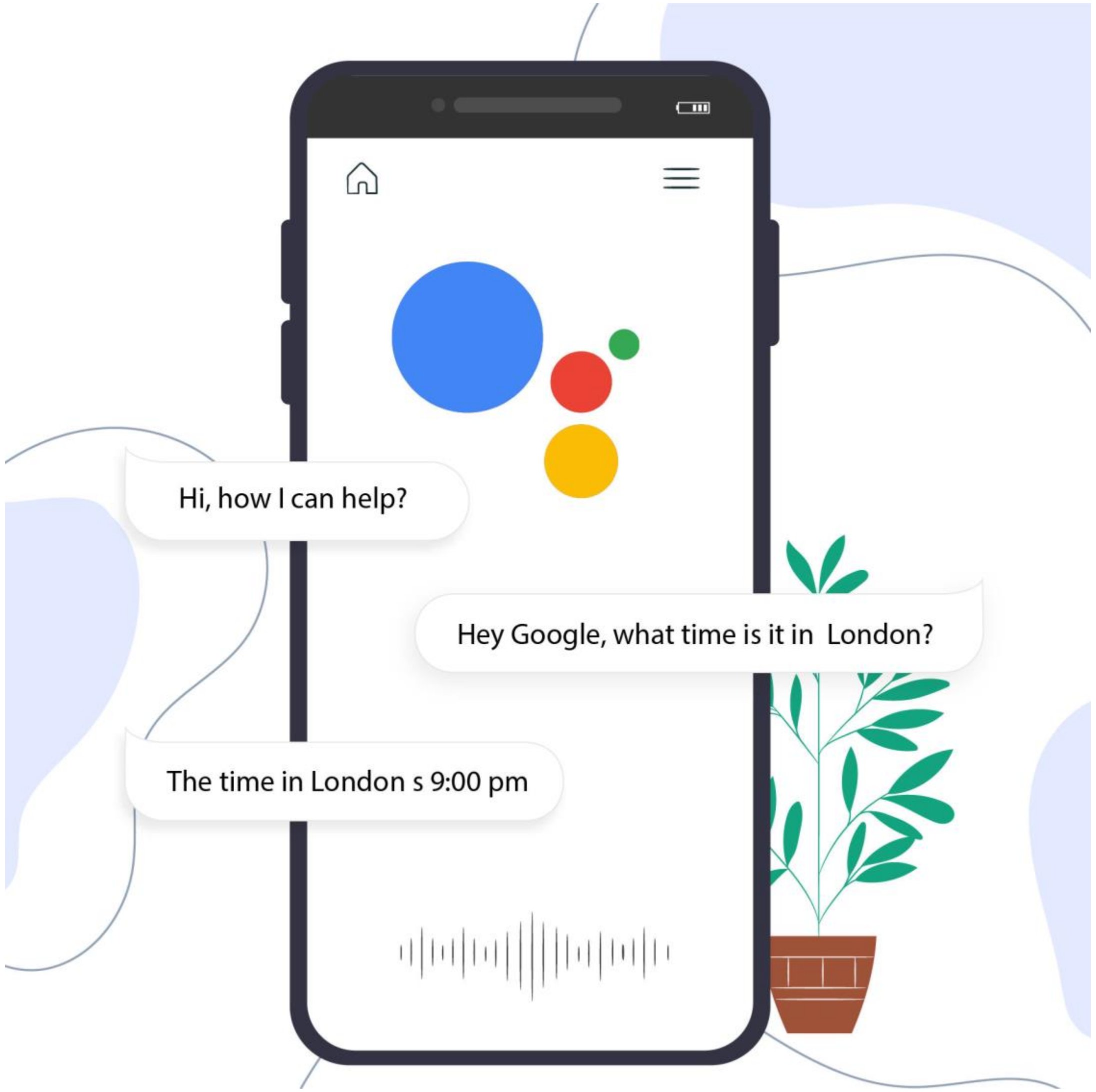


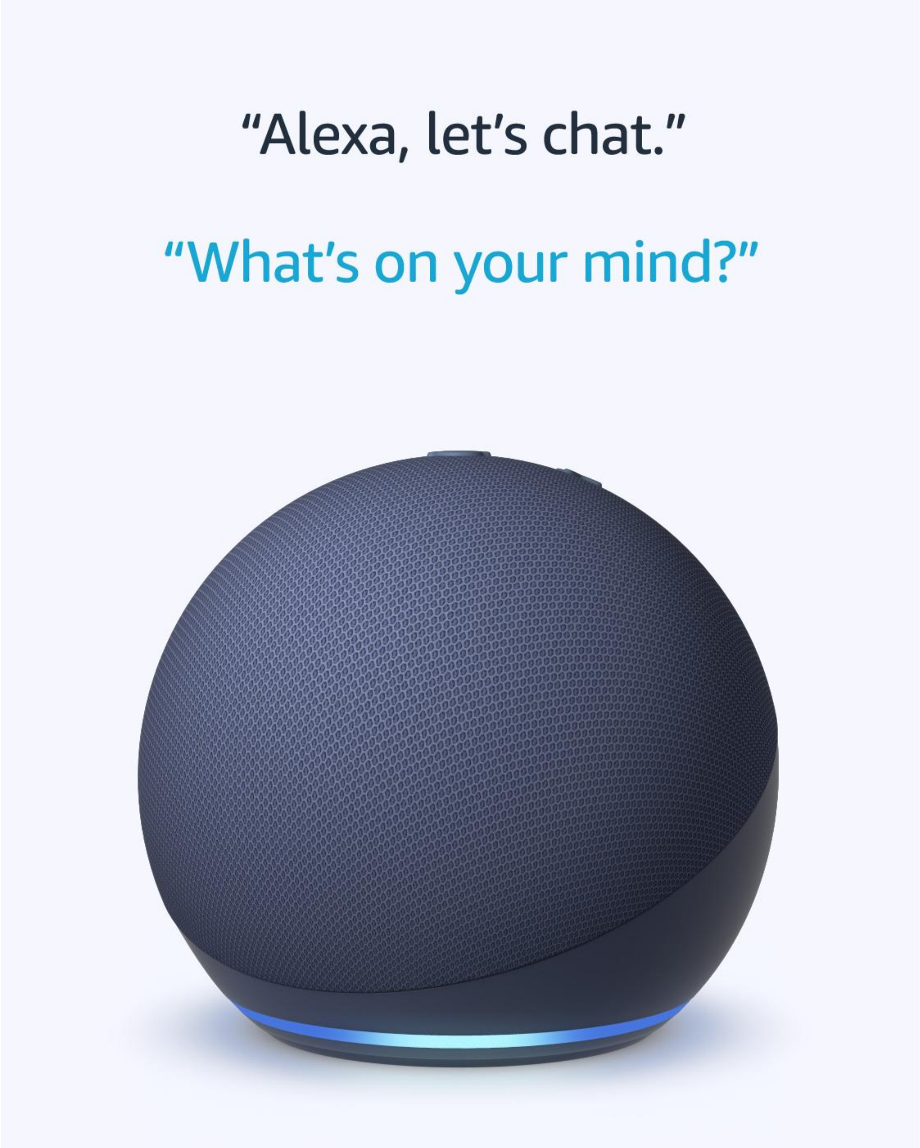
Generated using Google trends
https://trends.google.com/trends/explore?date=all&q=machine%20learning,artificial%20intelligence&hl=en-GB

Application: Spam Classification

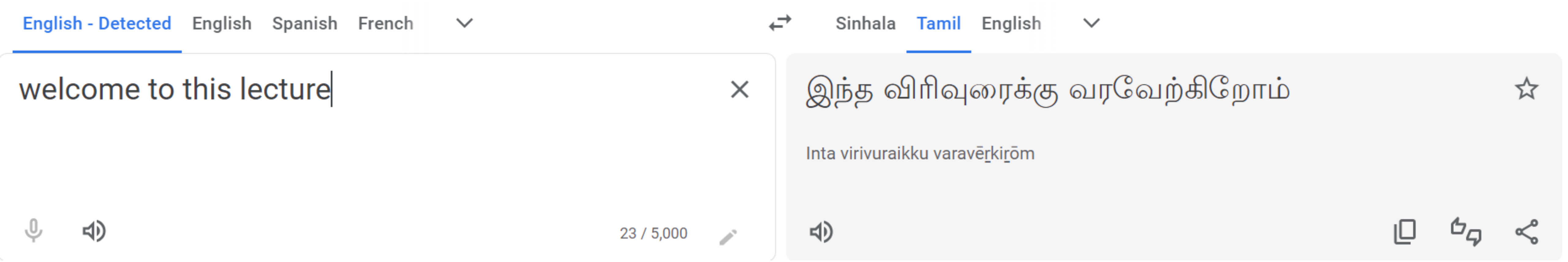


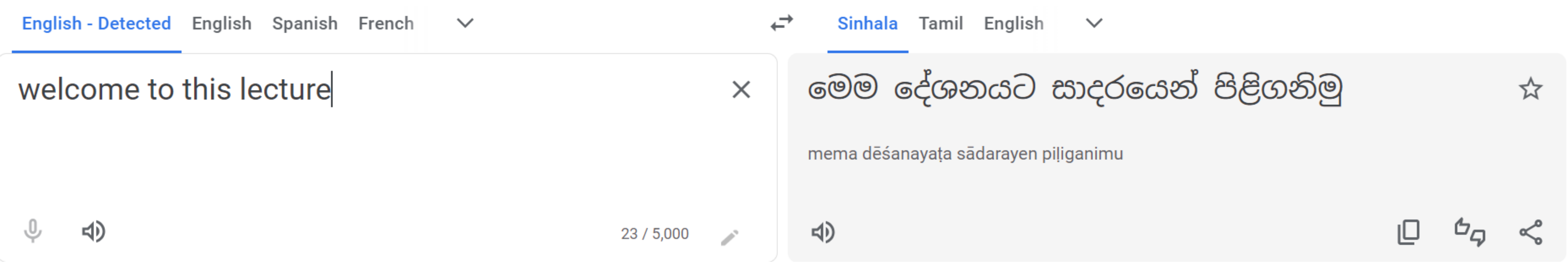
Application: Chatbots





Application: Machine Translation





Application: Self-Driving Cars

Supervised vs Unsupervised learning

Supervised Learning:

An algorithm that learns a function from examples of inputs and outputs. This type of learning required manually labelled data in a training set which is then used to train a model to predict on “unseen” data.

i.e: Classification, Regression

Unsupervised Learning:

An algorithm that finds structure in unlabelled data. These algorithms are more focused on data exploration and knowledge discovery.

i.e: Clustering

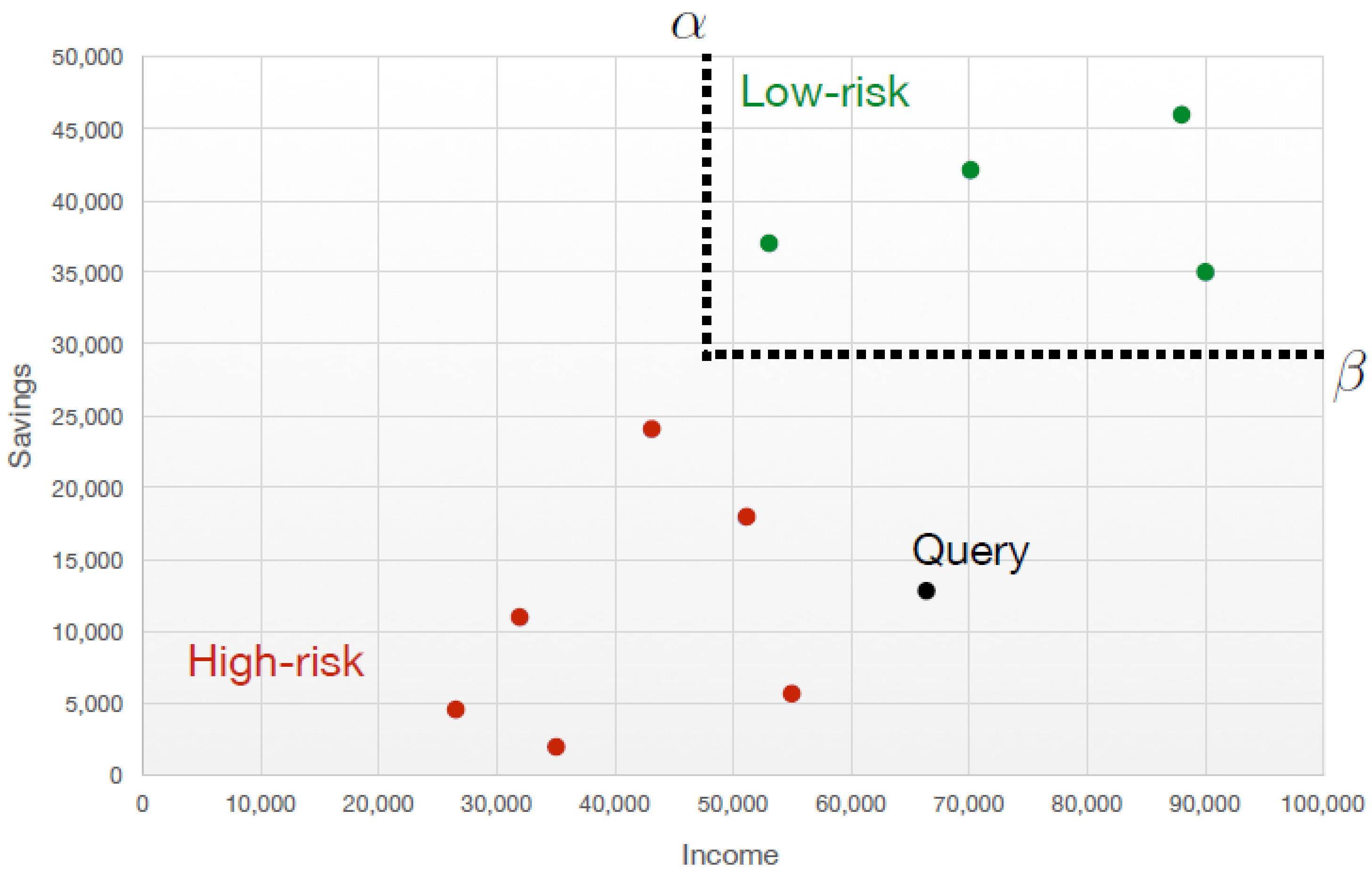
Typical Supervised Learning Process

A Decision Process: Based on the input data, the algorithm produces an estimate of the pattern in the data

An Error Function: A method which evaluates how good the prediction made is

A Model Optimisation Procedure: Model is adjusted to try to fit the data better and the error is re-evaluated. This step is repeated until the threshold accuracy is met.

Typical Classification Task



Aim: Can we train an algorithm to learn how to classify a new customer as high risk or low risk?

i.e: Can we learn α and β – the decision boundaries

Typical Classification Task

Example	Income	Savings	Married	Gender	Age	Class
1	35,000	2,000	Y	M	32	High-risk
2	51,000	18,000	N	M	34	High-risk
3	70,000	42,000	Y	F	41	Low-risk
4	26,500	4,500	N	M	22	High-risk
5	32,000	11,000	N	F	25	High-risk
6	53,000	37,000	N	F	39	Low-risk
7	88,000	46,000	Y	M	48	Low-risk
8	55,000	5,700	N	M	55	High-risk
9	90,000	35,000	Y	F	61	Low-risk
10	43,000	24,000	Y	M	33	High-risk

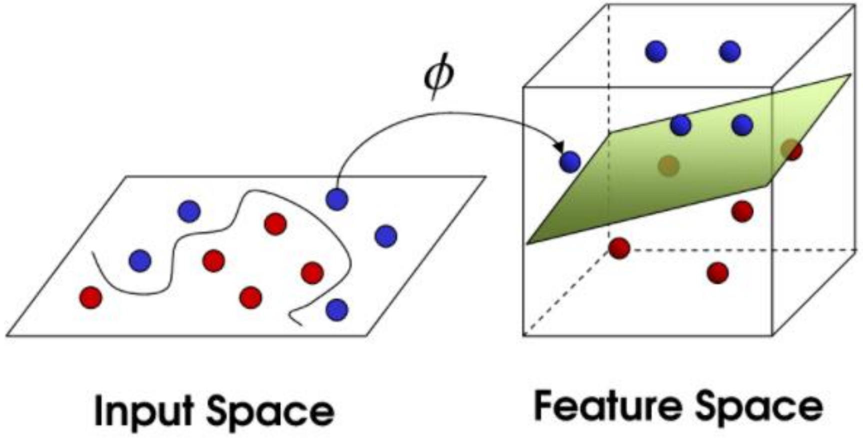
Example	Income	Savings	Married	Gender	Age	Class
X	66,000	13,000	Y	M	44	???

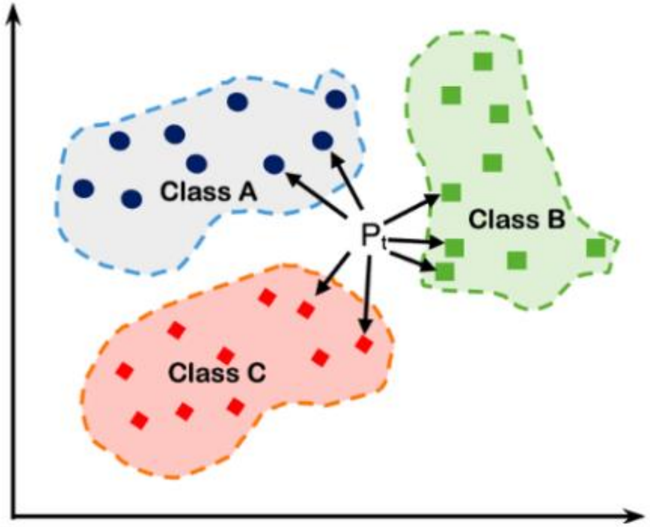
Classification Algorithms

- There are many different learning algorithms for classification
- Popular algorithms include: k-Nearest Neighbour, Decision Trees, Neural Network, Support Vector Machine)
- The choice of classification algorithm depends on the data. There is no one model that performs best for every dataset

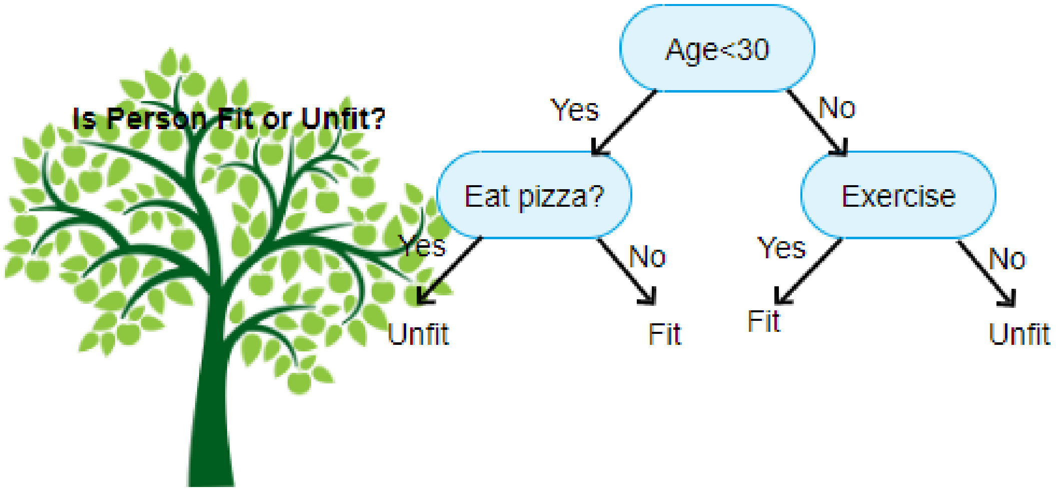
Classification Algorithms

Support Vector Machines





K-Nearest Neighbour



Decision Tree

Typical Regression Task

Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, can we learn a function f(x) to predict y given x?

If y is real-valued, the task is regression

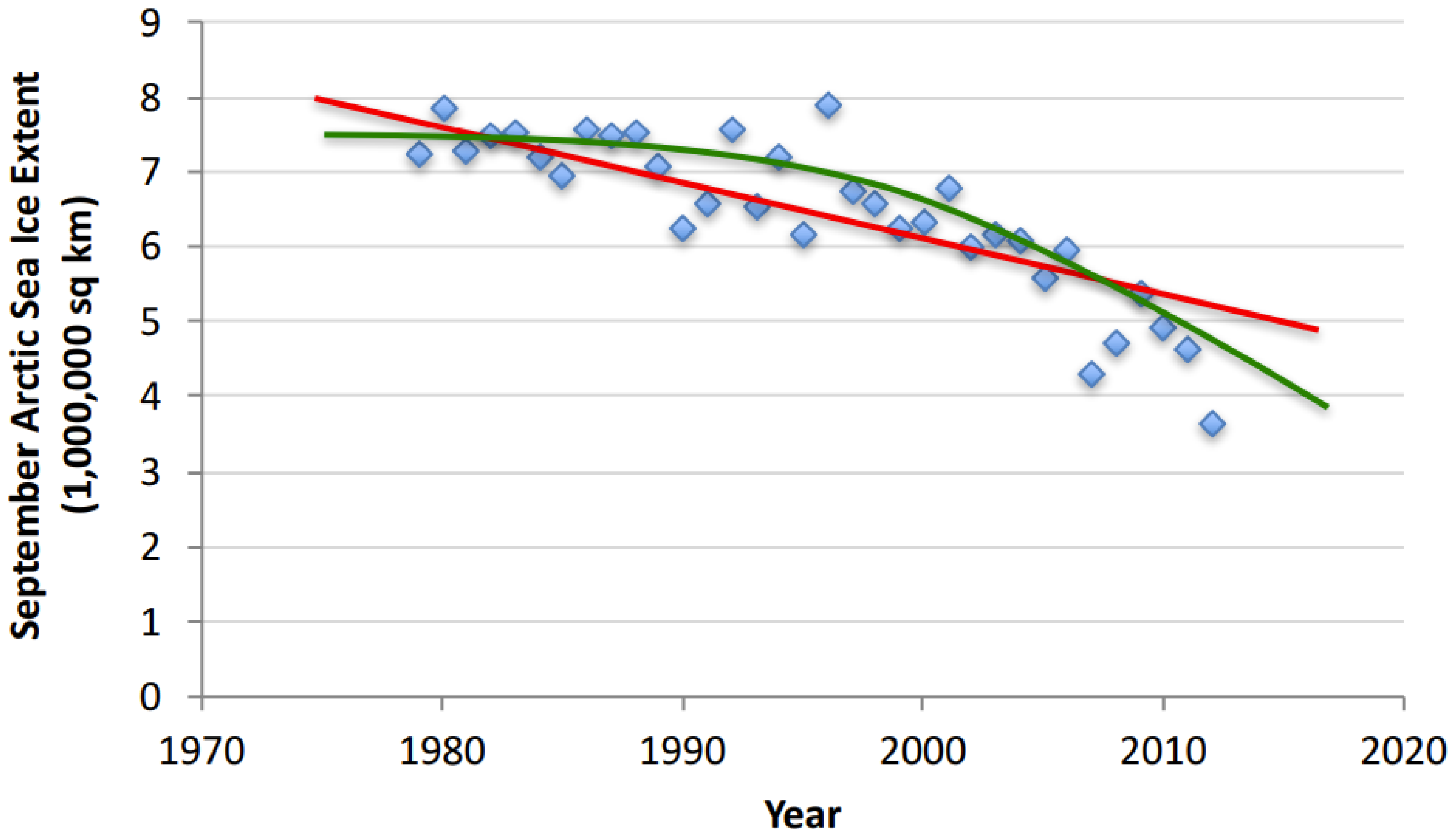


Image based on slide by Eric Eaton, University of Pennsylvania

Linear algebra and Machine Learning

- Linear algebra is the **mathematics of arrays**…

Many tasks in Machine Learning are performed using arrays:

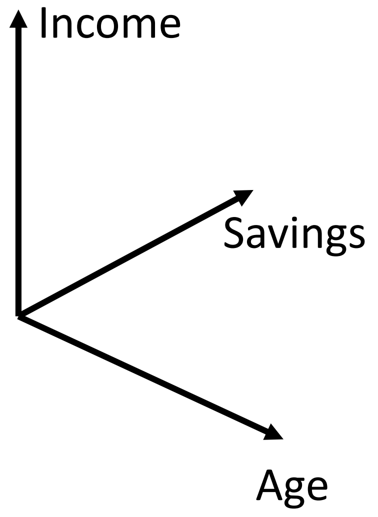
- Representing data (inputs and outputs)
- Some models are represented using arrays
- Some internal computations use arrays

Representing data as vectors…

A feature vector is a vector containing the "features" of some instance

$$C = \begin{array}{cl} \begin{bmatrix} 35{,}000 \\ 2000 \\ 32 \end{bmatrix} & \begin{array}{c} \mathit{Income} \\ \mathit{Savings} \\ \mathit{Age} \end{array} \end{array}$$

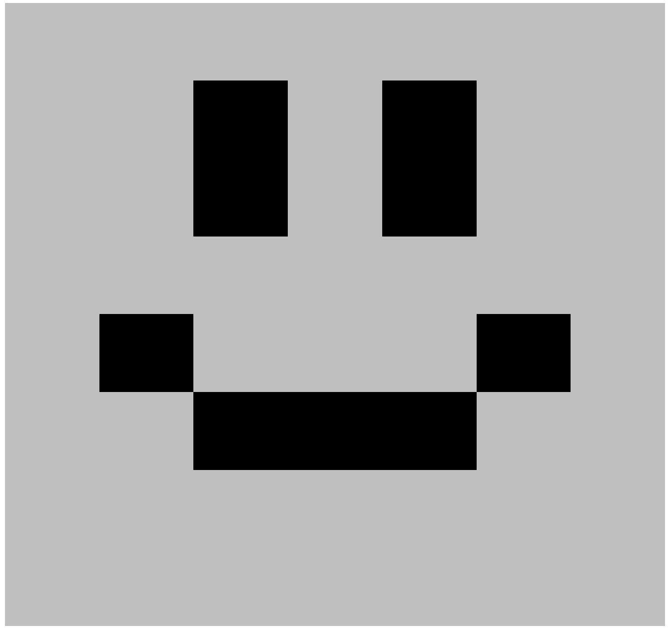
Where C represents the feature space for the credit risk for one person



In Machine Learning, this vector space is called the feature space

Representing data as vectors...

Images



$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Representing data as arrays…

Words and Documents

Given a collection of Wikipedia articles, create a vector for every word where the i^{th} entry is the number of times that word appears in the i^{th} document.

$$\text{Dog} = \left[\begin{array}{c} 0 \\ 1 \\ 0 \\ 10 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 55 \\ 0 \\ 0 \\ 1 \end{array}\right] \begin{array}{l} Wiki\ doc\ 1 \\ Wiki\ doc\ 2 \\ Wiki\ doc\ 3 \\ Wiki\ doc\ 4 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ Wiki\ doc\ 11 \\ Wiki\ doc\ 12 \\ Wiki\ doc\ 13 \end{array}$$

Has potential to end up as a very large vector space

Representing data as arrays...

How about non-numerical data (e.g. words)

One hot encoding

One hot encoding means, each object (or each word) is assigned a vector with a single one and zeros elsewhere. As an example if we have a dictionary with four words.

Apple = $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ Mango = $\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ Orange = $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ Banana = $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

Challenges in representing data as vectors…

- Data can be very large – hence large vectors/matrices

- Vectors can be very **sparse** meaning a lot of zeros.

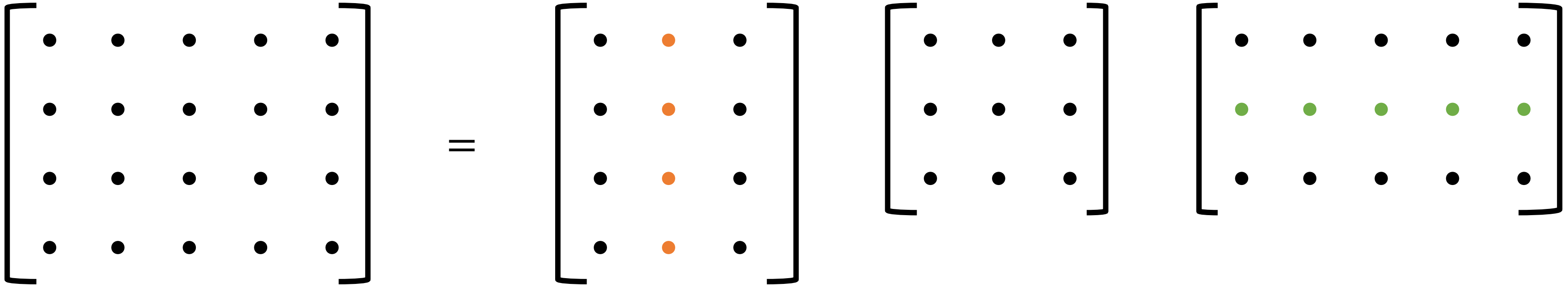
Singular Value Decomposition (SVD)

- SVD of a matrix means the factorization of that matrix into three smaller matrices

Columns are Orthonormal

Diagonal Matrix

Rows are Orthonormal



M

n x m

U

n x k

D

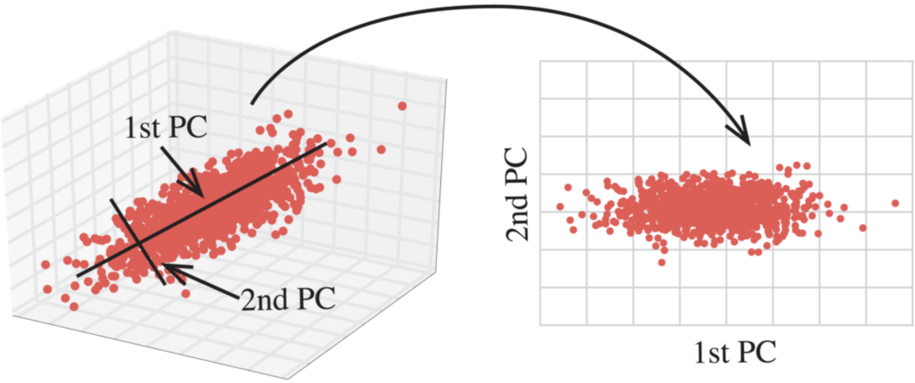
k x k

k = rank of M

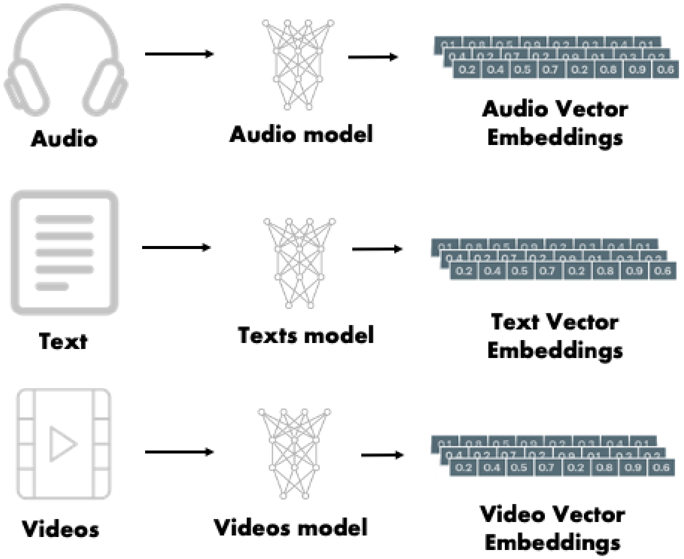
V

k x m

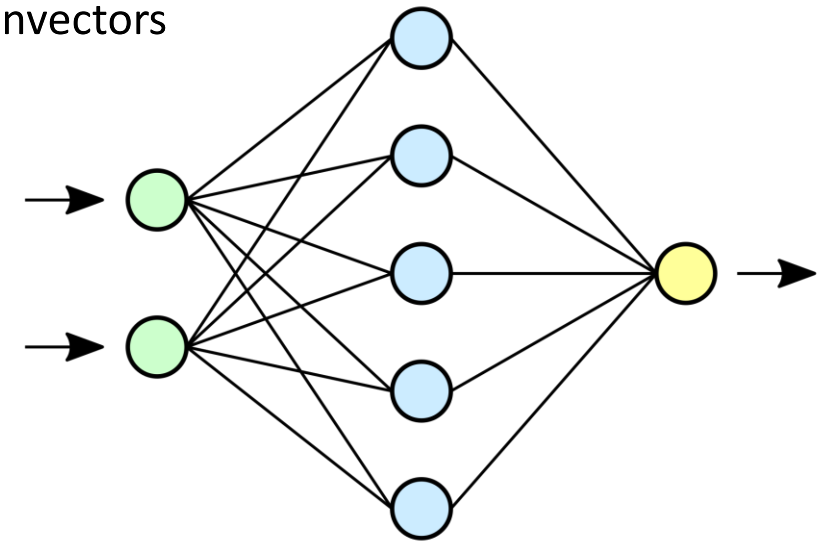
There is much more…



Dimensionality Reduction
Reduce large dimensional data using linear maps and their eigenvalues and eigenvectors



Vector Embeddings
Convert data into meaningful embeddings



Math inside ML Models
Much of the math inside machine learning models involve linear algebra techniques

What do I do?

Understand Performance

Diagnose Injury

Monitor Recovery

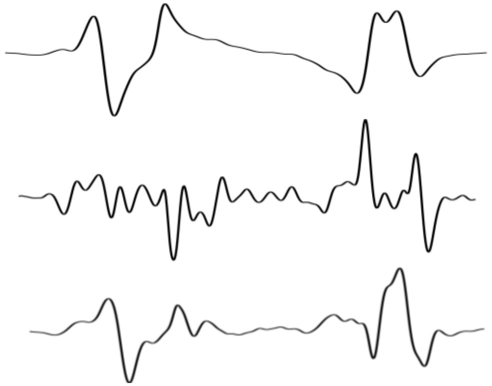
- PhD Researcher at University College Dublin in Ireland
- Applying Machine Learning methods to Sports Analytics applications
- Particularly, using time series techniques on wearable sensor data

My Research

Fatigue ⬆

Risk of Injury ⬆

Inertial Measurement Units (IMU's)



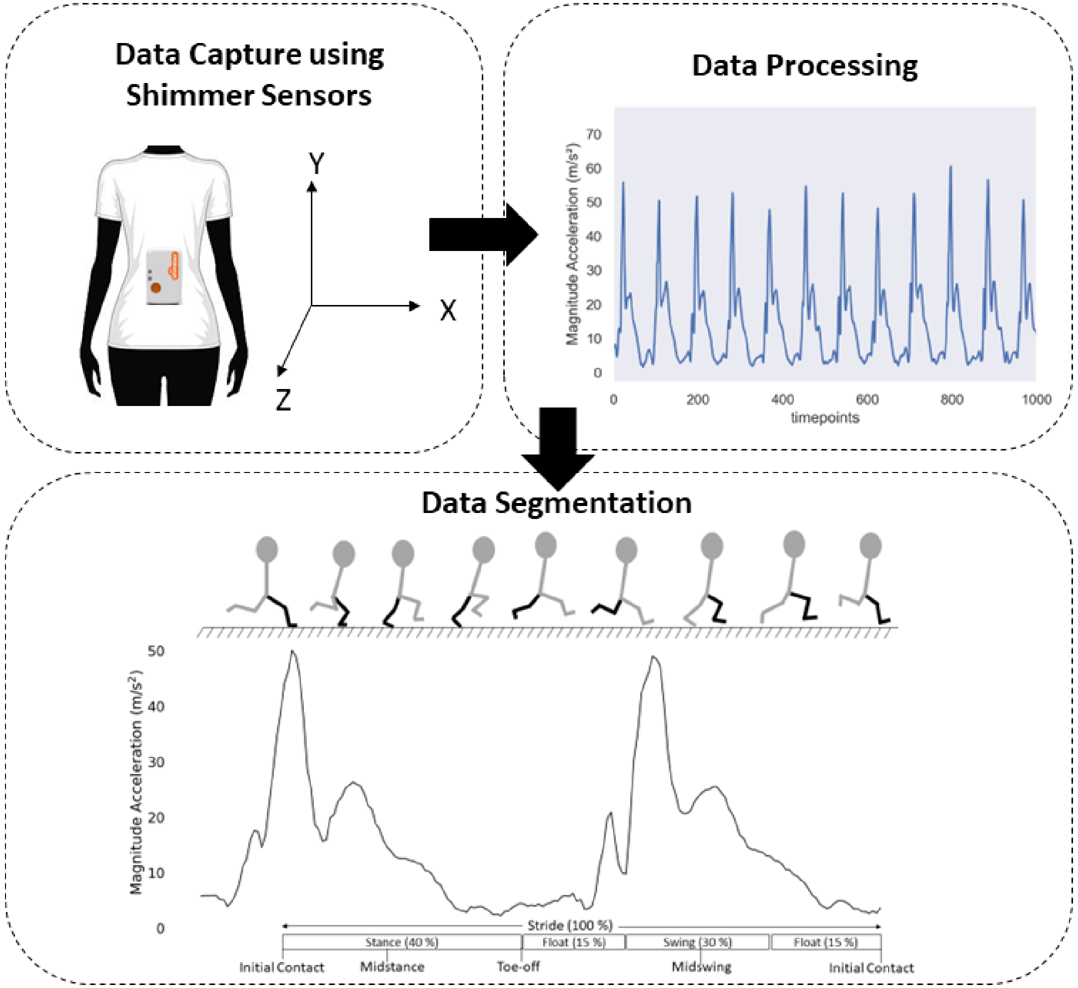
Time Series Signals from sensors

➧ *Lack of data*

➧ *Subtle differences between classes*

➧ *uninformative*

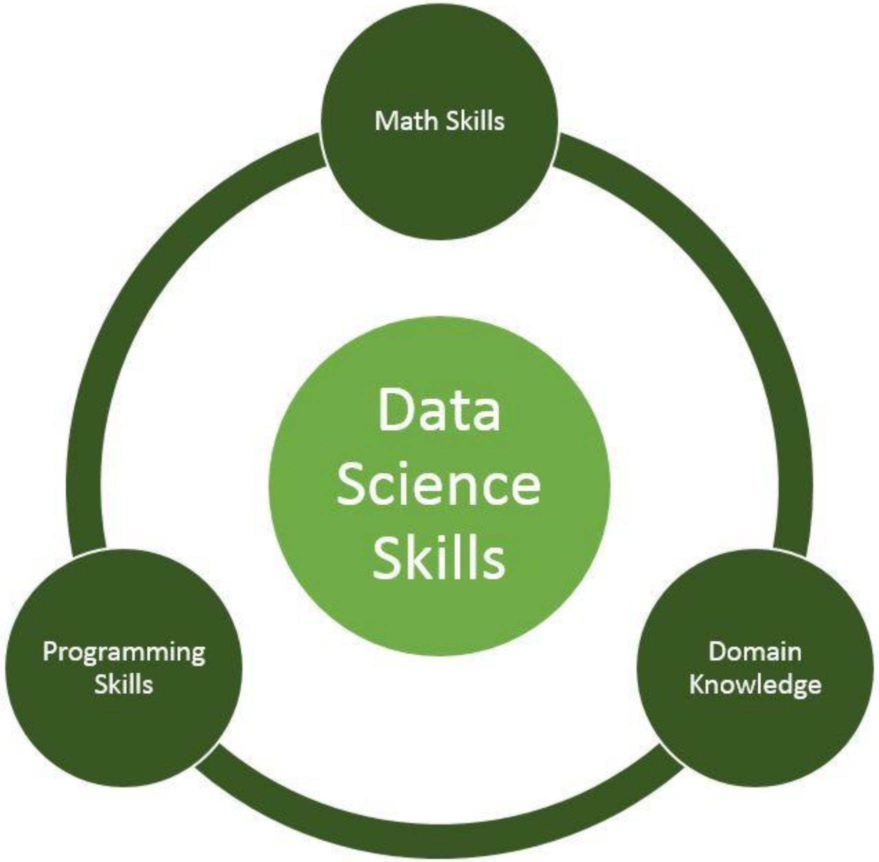
My Research



- Data is in a time series format.
- Features are represented as feature matrices instead of feature vectors
- 3D representation -> number of instances x number of timepoints x number of variable
- Hence a lot of matrix manipulation is required

Thinking of a career in Machine Learning?

- Applications of ML exists in every field
- Maths and Programming are essential
- Often a postgraduate degree is required
- Job Market = Excellent!!!



References and Resources used

TensorFlow – A friendly introduction to linear algebra for ML (https://www.youtube.com/watch?v=LlKAna21fLE)

Questions??

Linear Algebra for Machine Learning – An Introduction

Bahavathy Kathirgamanathan

Contacts:
Email: Bahavathy.Kathirgamanathan@ucdconnect.ie
LinkedIn: www.linkedin.com/in/bahavathy-kathirgamanathan