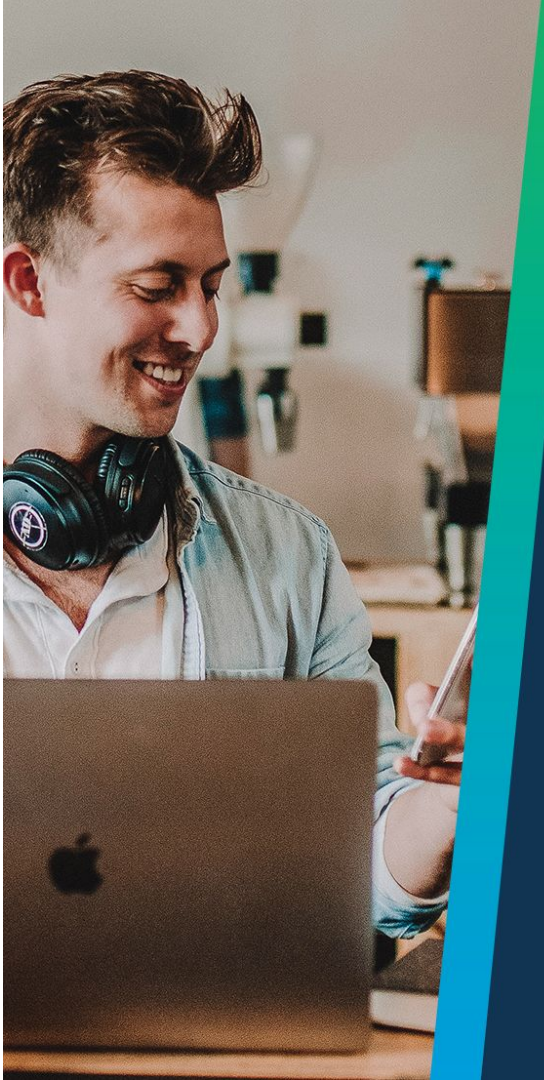


Data Science Bootcamp

Hyperiondev

Introduction to Machine Learning

Welcome

Your Lecturer for this session

Sanana Mwanawina

Lecture - Housekeeping

- The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment for all - please engage accordingly.
- No question is daft or silly - **ask them!**
- There are Q/A sessions midway and at the end of the session, should you wish to ask any follow-up questions.
- You can also submit questions here: hyperiondev.com/sbc4-ds-questions
- For all non-academic questions, please submit a query: hyperiondev.com/support
- Report a safeguarding incident: hyperiondev.com/safeguardreporting
- We would love your feedback on lectures: https://hyperionde.wufoo.com/forms/zsgv4m40ui4i0g/

Objectives

- Learn about supervised and unsupervised learning

Defining Machine Learning

- ★ The term “machine learning” is often used interchangeably with the term “artificial intelligence” (AI). But, while the two are very much related, they are not the same thing.
- ★ There is much debate about the difference between the two, but a simple way to look at it for our purposes is to see Machine Learning as a type of artificial intelligence.
- ★ Any program that completes a task in a way that can be considered human-like can be considered an example of artificial intelligence, but only programs that solve the task by learning without pre-programming are machine learning programs.

Input and Output

- ★ Whatever it is that we want a machine-learning algorithm to learn, we first need to express it in a numerical way.
- ★ The machine-readable version of a task consists of an input and an output. The input is whatever we want the algorithm to learn from, and the output is the outcome we want the algorithm to be able to produce e.g., image identification system
- ★ An example of input would be the budget and number of awards of a movie, and an example of output would be the box office sales of that movie

Input and Output

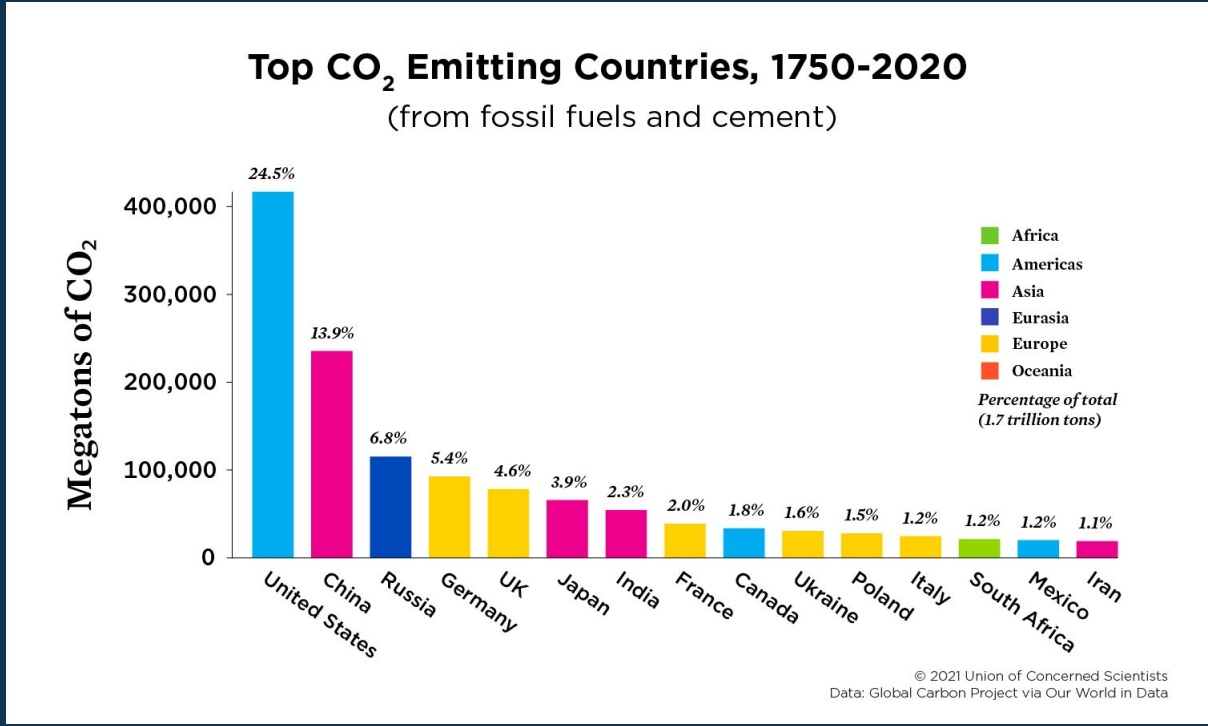
- ★ Since machine learning is a young field that overlaps several other disciplines, including statistics, the input and output may be referred to by several other names.

- ★ For input, these include features (named after the fact that inputs typically ‘describe’ something), independent variables, and explanatory variables (named due to the fact that the output is usually assumed to depend on or be explained by the input).

- ★ For output, alternate terms are labels, predictions, dependent variables, and response variables.

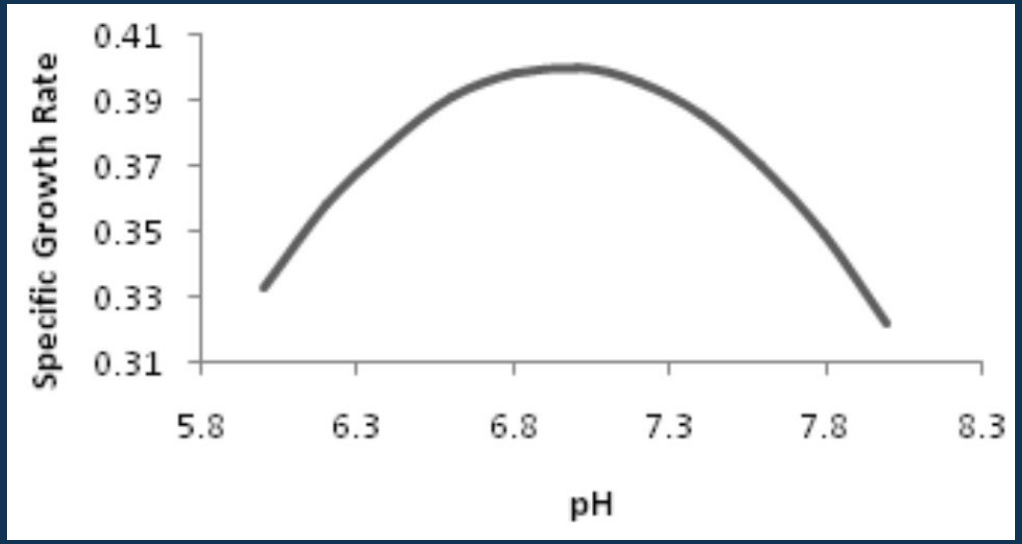
Input and Output

- ★ Dependent vs independent variables



Input and Output

- ★ Response vs explanatory variables



Supervised and Unsupervised Learning

- ★ Machine learning algorithms can be categorised according to whether they learn from seeing examples of the desired output or not.
- ★ Supervised methods learn by seeing such examples, which may be time-consuming and difficult to create.
- ★ Unsupervised approaches learn without needing any human involvement, but they often do not achieve the same results as supervised algorithms.

Supervised Learning

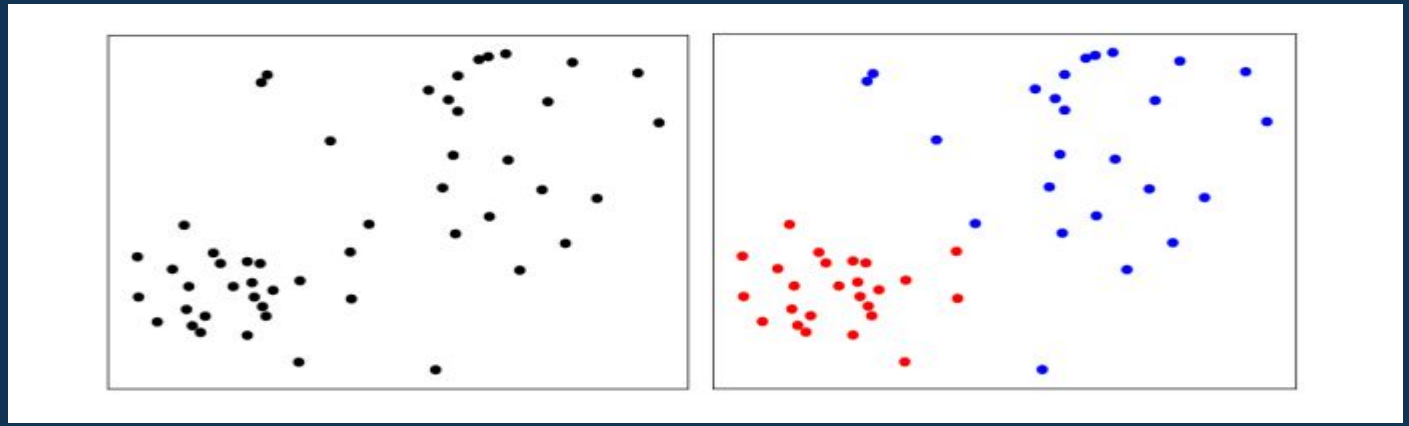
- ★ In supervised learning problems, a program predicts an output for an input by learning from pairs of inputs and outputs (labels); that is, the program learns from examples that have had the right answers assigned to them beforehand.
- ★ These assignments are often called annotations. Because they are considered the correct answers, they are also called gold labels, gold data, or the gold standard. The collection of examples that comprise the supervised experience is called a training set.
- ★ A collection of examples that are used to assess the performance of a program is called a test set. Like a student learning in a language course that teaches only through exposure, supervised learning problems see a collection of correct answers to a variety of questions, and must learn to provide correct answers to new, but similar questions.

Supervised Learning

- ★ Another example: an image identification system
- ★ Training data: lots of images of cheetahs
- ★ Test data: images of cheetahs it has never seen before

Unsupervised Learning

In unsupervised learning, a program does not learn from labelled data. Instead, it attempts to discover patterns in the data on its own. For example, suppose you have two classes scattered in a 2-dimensional space (as in the first of the images below) and you want to separate the two data sets (as in the second image). Unsupervised learning finds underlying patterns in the data, allowing the classes to be separated.



Unsupervised Learning

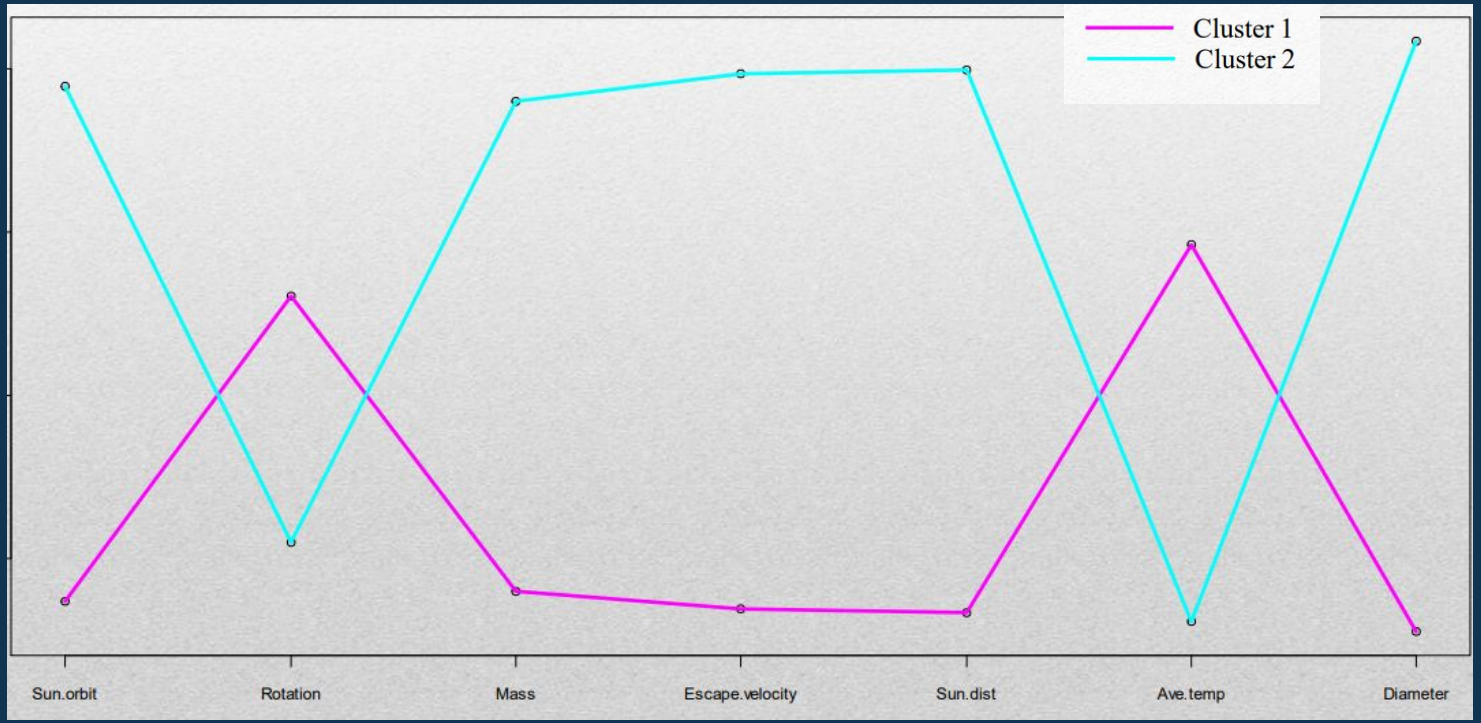
We would like to perform a cluster analysis on this dataset. That is, group them according to how similar they are.

	1	2	3	4	5	6	7	8
1	Planet	Orbit	Rotation	Mass	Velocity	Distance	Surface	Diameter
2	Venus	224	5836	0.81	37498	1.07E+08	449	12107
3	Earth	365	24	1	65578	1.5E+08	7.2	12755
4	Mars	687	25	0.1	18024	2.29E+08	-70	6794
5	Jupiter	4332	10	317	214203	7.77E+08	-153	142983
6	Saturn	10759	10	95	127782	1.43E+09	-184	120536

Unsupervised Learning

Planet	Sun orbit	Rotation	Mass	Escape velocity	Sun dist	Ave temp	Diameter
Venus	224	5836	0.81	37498	170000000	449	12107
Earth	365	24	1	67578	150000000	7.2	12755
Mars	687	25	0.1	18024	229000000	-70	6794
Cluster 1 mean	**425.3**	**1961.7**	**0.637**	**41033.3**	**183000000**	**128.7**	**10552**
Jupiter	4332	10	317	214203	777000000	-153	142983
Saturn	10759	10	95	127782	1429000000	-184	120536
Cluster 2 mean	**7545.5**	**10**	**206**	**170992.5**	**1103000000**	**-168.5**	**131759.5**

Unsupervised Learning



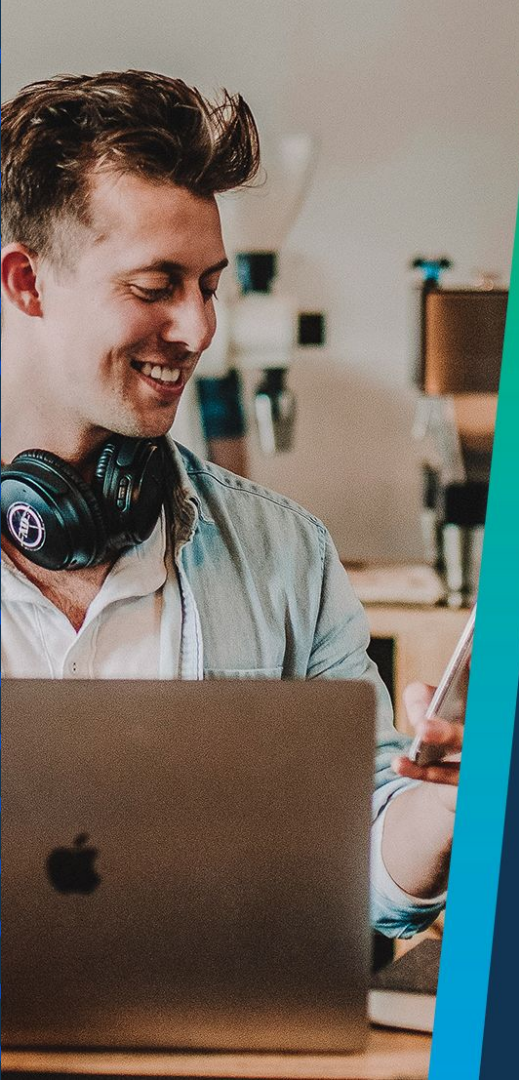
Supervised and Unsupervised Learning

- ★ To highlight the difference between supervised and unsupervised learning, consider the following example.
- ★ Assume that you have collected data describing the heights and weights of people.
- ★ An unsupervised clustering algorithm might produce groups that correspond to men and women, or children and adults.
- ★ An example of a supervised learning problem is if we label some of the data with the person's sex, and then try to induce a rule to predict whether a person is male or female based on their height and weight.

Hyperiondev

Q & A Section

Please use this time to ask any questions relating to the topic explained, should you have any



Hyperiondev

Thank you for joining us