

Introduction to Python and Machine Learning

Unit 7

Working draft

Nepal Data Literacy Program, 2019

Organized by



Supported by



Modules in Unit 7

- **Module 1:** Introduction to Google Colab and Basic Python Syntax
- **Module 2:** Basic Probability in Python
- **Module 3:** Introduction to Seaborn for Visualization
- **Module 4:** Introduction to Pandas
- **Module 5:** Introduction to Machine Learning

Module 1

Introduction to Google Colab and
Basic Python Syntax

Google Colab

Colaboratory is a free web based application for data analysis that requires no setup and runs entirely in the cloud.

With Colaboratory you can write and execute code, save and share your analyses, and access powerful computing resources, all for free from your browser.

co

Drive FUSE example.ipynb ☆

File Edit View Insert Runtime Tools Help

CODE TEXT CELL CELL

CONNECTED EDITING

Table of contents Code snippets Files

UPLOAD REFRESH

drive

My Drive

Colab Notebooks

07_structured_data.ipynb

BigQuery recipes

Copy of nima colab.ipynb

Exported DataFrame sheet.gsheets

Pickle + Drive FUSE example.ipynb

Sample Excel file.gsheets

Sample Excel file.xlsx

TFGan tutorial in Colab.txt

created.txt

foo.txt

script.ipynb

to_upload (1).ipynb

to_upload (2).ipynb

to_upload (3).ipynb

to_upload.ipynb

variables.pickle

sample_data

[1]

```
1 # Load the Drive helper and mount
2 from google.colab import drive
3
4 # This will prompt for authorization.
5 drive.mount('/content/drive')
```

Mounted at /content/drive

```
1 # After executing the cell above, Drive
2 # files will be present in "/content/drive/My Drive".
3 !ls "/content/drive/My Drive"
```

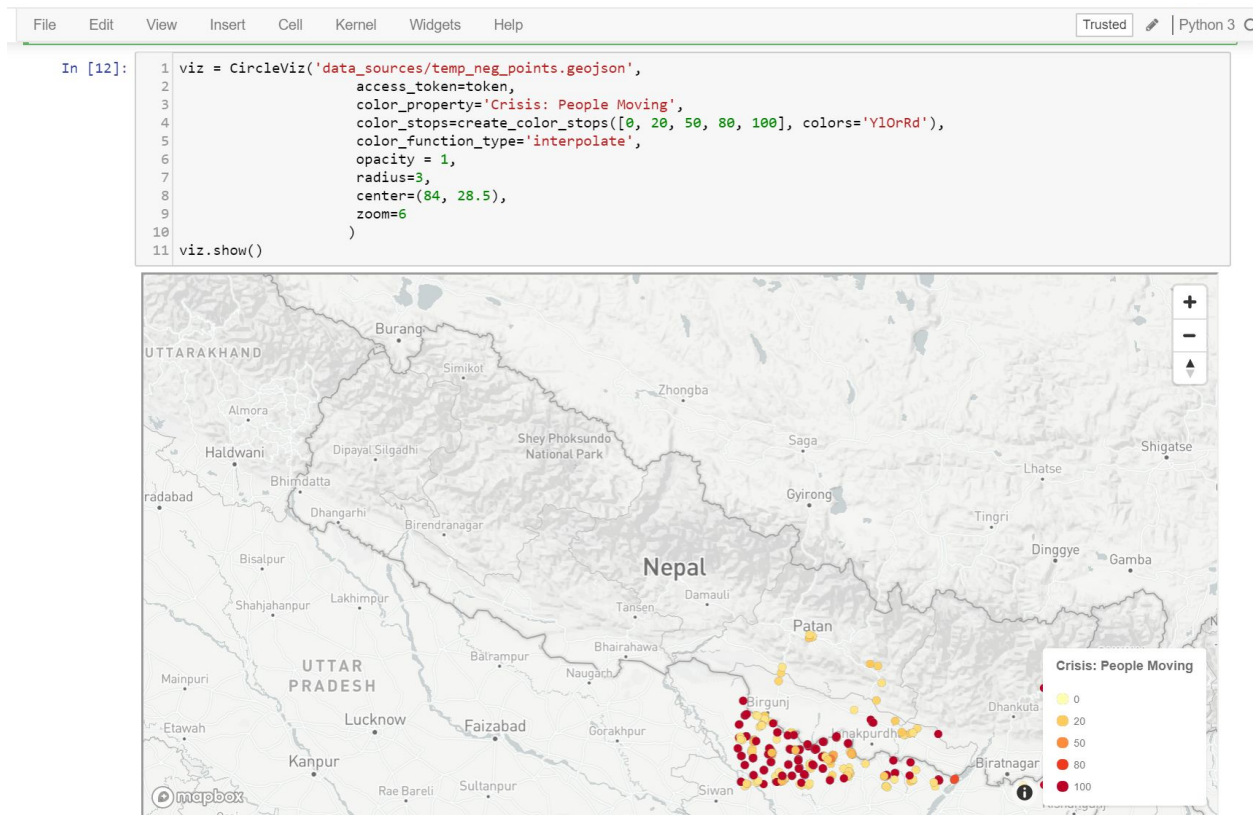
07_structured_data.ipynb	Sample Excel file.xlsx
BigQuery recipes	script.ipynb
Colab Notebooks	TFGan tutorial in Colab.txt
Copy of nima colab.ipynb	to_upload (1).ipynb
created.txt	to_upload (2).ipynb
Exported DataFrame sheet.gsheets	to_upload (3).ipynb
foo.txt	to_upload.ipynb
Pickle + Drive FUSE example.ipynb	variables.pickle
Sample Excel file.gsheets	

Google Colab Screenshot

Jupyter Notebook

Jupyter notebook is open-sourced. Feel free to download [Anaconda](#) to get Jupyter Notebook setup on your computer. For this course, however, we will be using Google Colab.

Google Colab also provide free GPU support, which is very useful for Machine Learning Models (especially Deep Learning Models which is resource intensive).



Jupyter Notebook Screenshot

Why use Jupyter or Colab as researchers?

1. Literate programming
2. Reproducible research
3. Teaching and learning

Link to Google Colab: <https://colab.research.google.com/>

Google Colab Activity

- Once you have opened the link and signed in, start a new Python 3 notebook by clicking on **New Python 3 Notebook**.
- If you click on Untitled1.ipynb, you should be able to rename the file. Rename the notebook to **Intro to Colab**
- You can share the file similar to how you shared your google sheets and google docs. Try it out.
- Click on +Text. This is useful if you want to provide additional information
- Write an introduction text to describe your notebook. For example “The purpose of the notebook is to learn the basics of Google Colab.”
- After you are done, hit Shift+Enter, or click on the run button on the left side of the cell.
- By default, if you do not select +Text, the cells are for code. Go to another cell and type `2+3`

:

- Try a few arithmetic operations (use * for multiplication, / for division and % for modular arithmetic)
- See what happens if you try dividing something by 0
- Try adding comments to your code by using #
- Create a variable x and assign a value of 5 by typing x = 5
- Create a variable y and assign a value of 10 by typing y = 10
- Add x and y and press Shift+Enter

Python Basics

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes **code readability** with its notable use of significant whitespace.

Open the notebook

https://colab.research.google.com/drive/1ywtKMeaXitdIMbBF0n3vz_cvrBMOmWSp

and copy it to your google drive.

The notebook covers topics on

1. Arithmetic Operations and Math Operators
2. Working with variables in Python
3. Getting help and using tab complete
4. Basic data types
5. Manipulating String
6. Data types: Lists
7. Manipulating Lists
8. Logic and Control Flow
9. Loops and iterables
10. Logic Operators
11. Conditional Statements
12. Testing Conditionals inside a loop

Lab 1: Problem 1

Open a new Google Colab to print if a number is even or odd.

1. Break down the problem
2. How do you know if a number is even or odd? Can you use the modulo operation?
3. Use if-else based on modulo operation result.
4. Complete the problem

Lab 1: Problem 2

Open a new Google Colab notebook or continue in your previous notebook and write a simple program to calculate the sum of the first 100 numbers (1-100)

1. Break down the problem
2. Choose the loop you prefer (for/while)
3. First of all, print the first 100 numbers
4. Complete the problem

Lab (Module 1) Solutions

<https://colab.research.google.com/drive/1X43SiE3bvWErQnrDeC1rUpQ4RDae3Gmc>

Module 2

Basic Probability in Python

Open the notebook and copy it to your Google Drive

<https://colab.research.google.com/drive/1O4bEoAq2xPV5h1s7AG0cM3h8GoM0uxxg>

The notebook covers topics on:

1. Using Random integers for basic simulations
2. Monte Carlo simulation
3. Importing Datasets in Google Colab from Google Drive
4. Dataframes
5. Shape, Describe, Sum etc in Dataframe
6. Calculating Probabilities
7. Conditional Probabilities

Module 2 Solutions

<https://colab.research.google.com/drive/1LygekNV9va00iRF2wMNIVHEQYTMx2YyW>

Module 3

Introduction to Seaborn for Visualization

The notebook covers topics on

1. A very simple plot on Seaborn
2. Adding a title to the plot
3. Adding x and y-axis labels
4. Changing the figure size
5. Different types of plot
6. Plotting using the Pandas plot function
7. Horizontal and vertical bar charts
8. Grouped bar plots
9. Monte Carlo plots for a coin flip
10. Introduction to histograms and real-world Normal Distributions

Open the notebook and copy it to your Google Drive

<https://colab.research.google.com/drive/1LPP7Kvso-zgS2CLYgG70UvEI2A4WSTmk>

Module 4

Introduction to Pandas

Open the notebook and copy it to your Google Drive

<https://colab.research.google.com/drive/1A2nRZaMkhmhdAAoE3x7Xbr78DEKq4ErF>

The notebook covers topics on

1. Creating series by passing a list
2. Creating an empty DataFrame, and add new columns to it
3. Creating a DataFrame using dictionaries to specify column name and data
4. Viewing data using head, tail and sample functions
5. Min and max in a column
6. Selecting and manipulating data
7. Sorting values
8. Plotting outputs
9. Concatenating DataFrames
10. Performing EDA in python using pandas
 - a. Reading data from CSV
 - b. Understanding data
 - c. Handling missing data

Linear Regression

Open the notebook and copy it to your Google Drive.

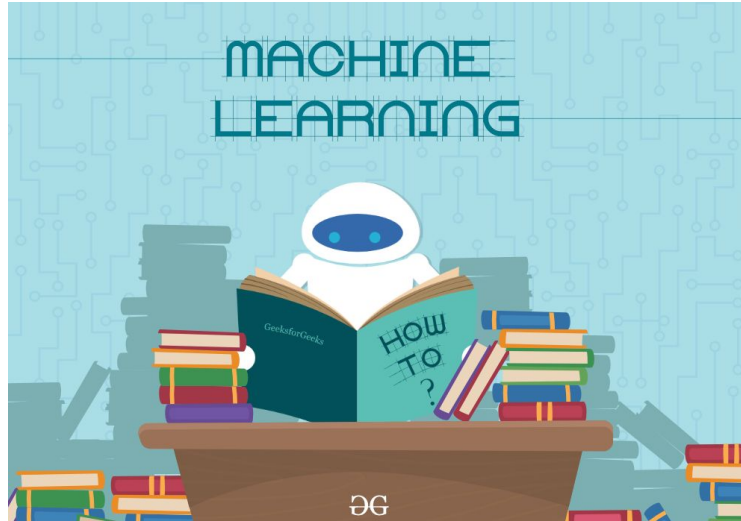
<https://colab.research.google.com/drive/1Mn9DzyXtlkn-PlqgyMdv6IECWdJFDyIB>

The notebook covers topics on

1. Correlation between two variables
2. Linear Regression
3. Predicting using Linear Regression
4. Model Score
5. Coefficient of Determination

Module 5

Introduction to Machine Learning



Artificial Intelligence (AI) and Machine Learning

Artificial intelligence is an area of computer science which has to do with creation of intelligent machines that work and act like humans.

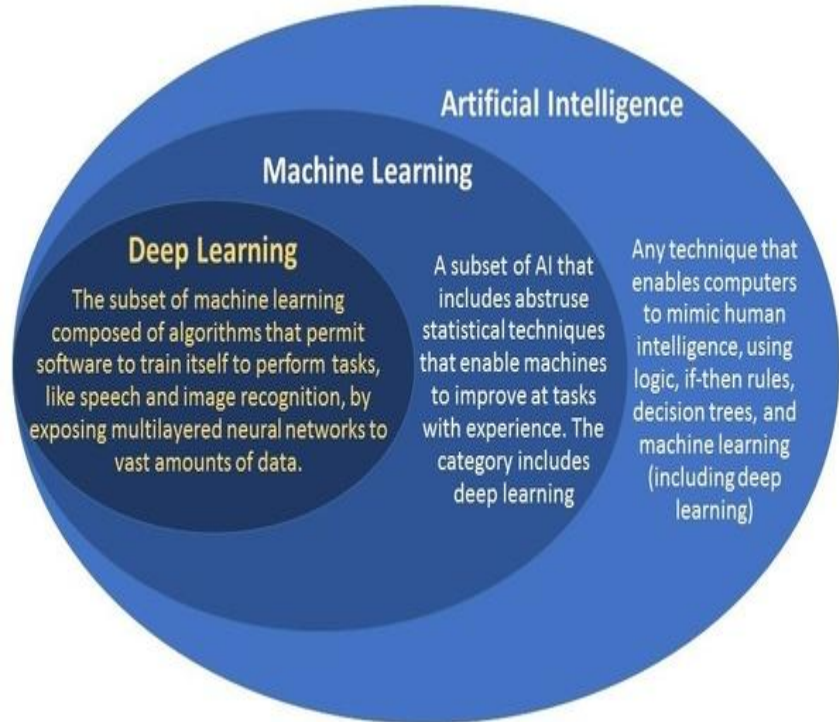


Image credit: [Vikas Kumar Sharma, Quora](#)

Artificial Intelligence (AI) and Machine Learning

Machine Learning is the practice of using statistics and algorithms to parse data, learn from it and make a prediction/determination about something without having to explicitly program

Goal of Machine Learning: How can we build a computer system that automatically improves with experience, and what are the fundamental laws that govern all these learning processes?

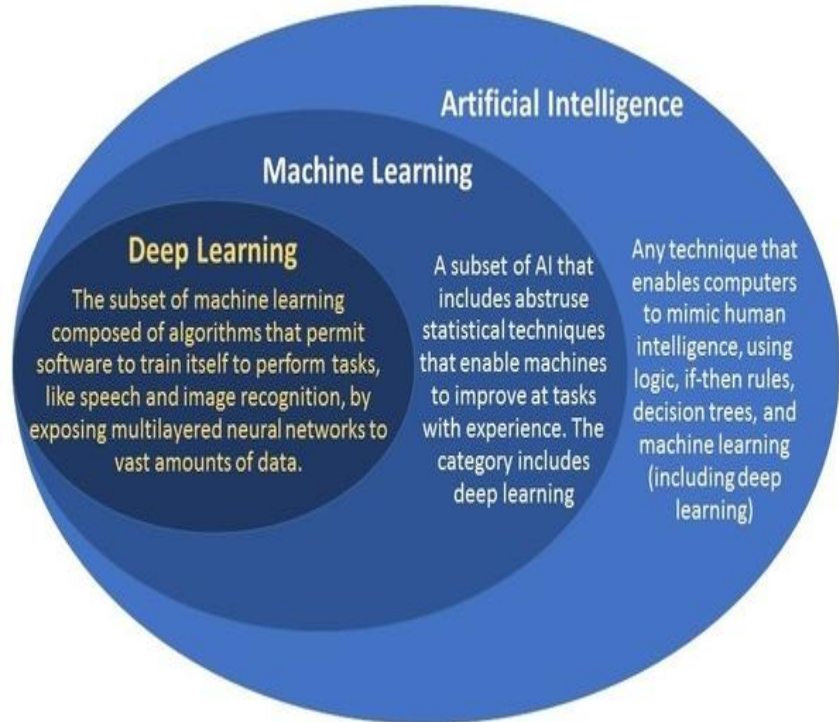


Image credit: [Vikas Kumar Sharma, Quora](#)

Artificial Intelligence (AI)

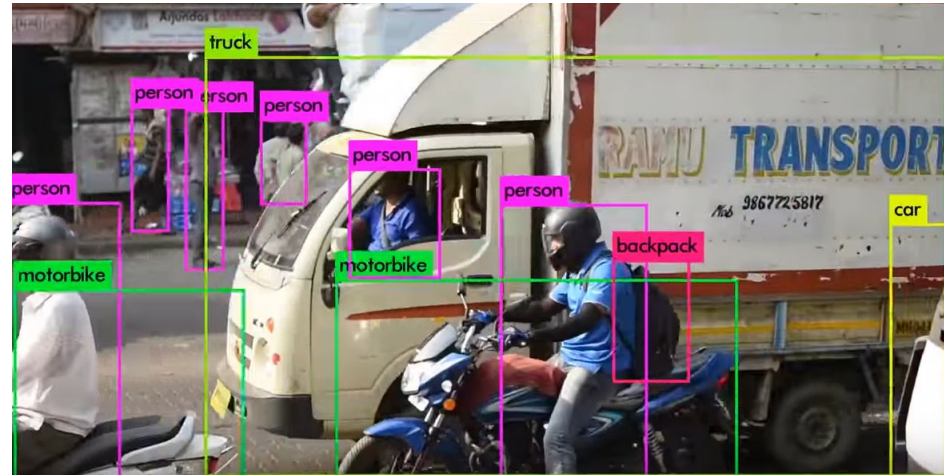
From "Mainstream Science on Intelligence" (1994),

A very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings — "catching on," "making sense" of things, or "figuring out" what to do.

Applications

YOLO Real Time Object Detection: Real time object detection video using state-of-the-art YOLOv3 architecture.

<https://pjreddie.com/darknet/yolo/>

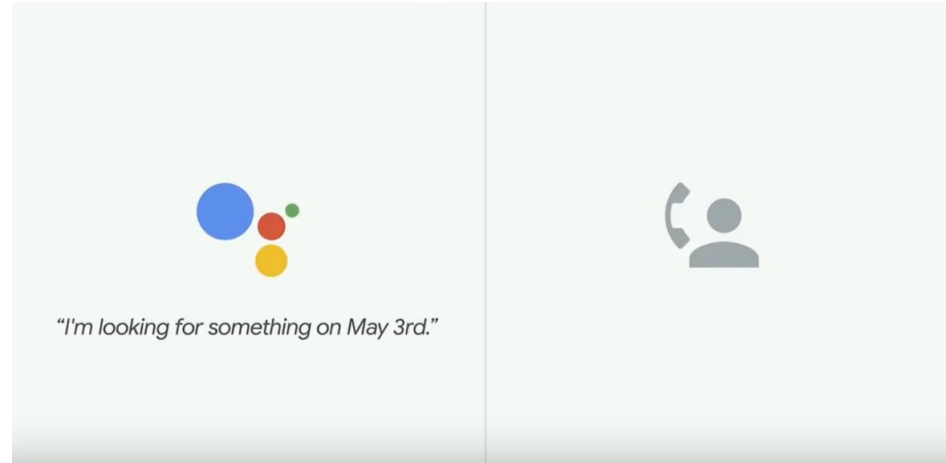


<https://www.youtube.com/watch?v=MPU2HistivI>

Applications (continued)

Google Duplex: A.I. Assistant Calls Local Businesses To Make

Appointments: AI chat bot that can carry out “real world” tasks, such as making a reservation or appointment, over the phone.

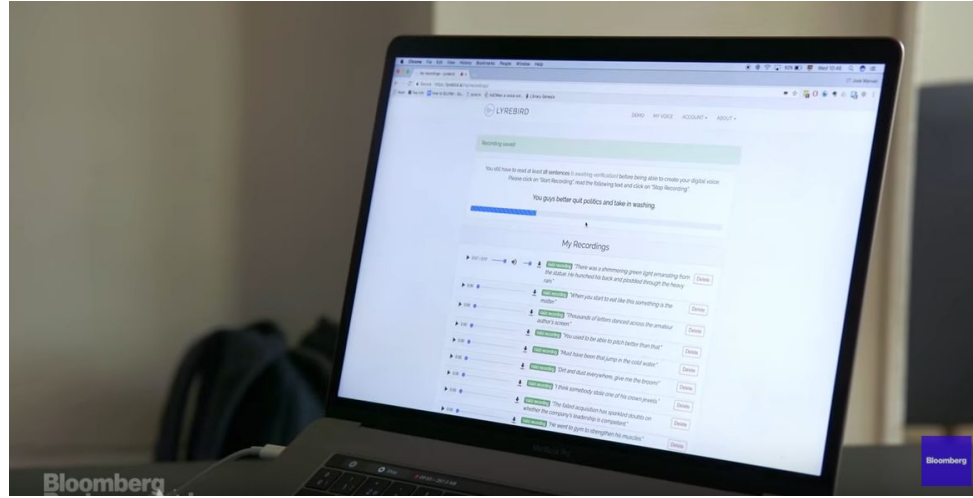


<https://www.youtube.com/watch?v=D5VN56jQMWM>

Applications (continued)

Voice Cloning: Lyrebird is a startup which used AI to clone human voices with good precision. This video features journalist Ashlee Vance who travels to Montreal, Canada to meet the founders and try the application

<https://www.youtube.com/watch?v=VnFC-s2nOtI>



Applications (continued)

Holographic Objects using Mixed Reality and Language Translation using Azure: Use of Azure AI, mixed reality and real time translation to render holographic speaker delivering speech in other languages.

<https://medialist.info/en/2019/07/22/hololens-2-and-azure-ai-real-time-translation-into-mixed-reality/>



<https://www.youtube.com/watch?v=auJJrHgG9Mc>

Applications (continued)

NVIDIA AI Car

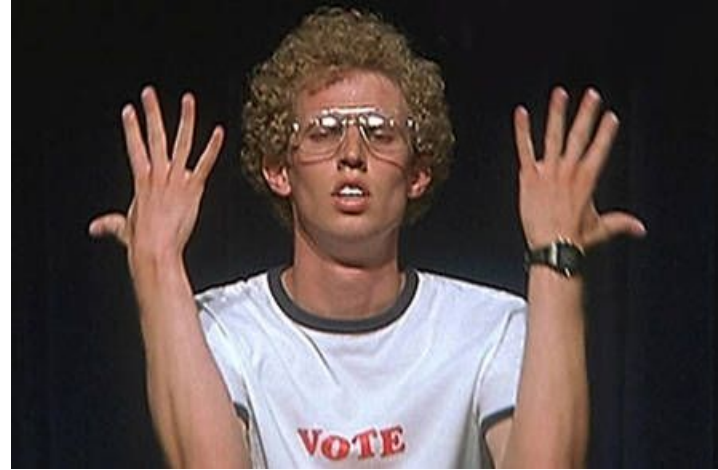
Demonstration: *BB8* learns to drive the car after watching humans drive in all kinds of roads and road conditions.



<https://www.youtube.com/watch?v=-96BEoXJMs0>

Applications (continued)

Everybody dance now: use generative adversarial networks (GANs) that can copy moves by professional performers, and map them to anyone's face and body.

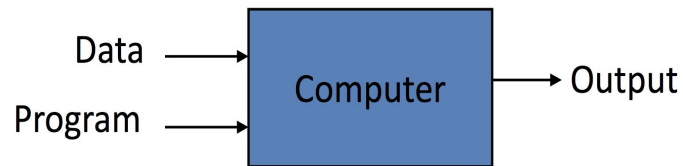


Source: https://www.theregister.co.uk/2018/08/24/ai_dancing/

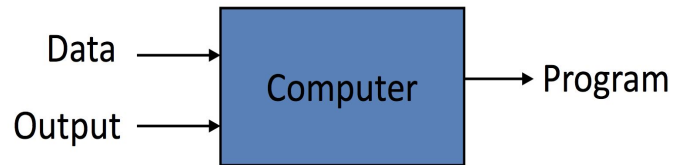
“Learning” in Machine Learning

a computer program learns from experience E with respect to some tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

Traditional Programming



Machine Learning



Credit: Tom Mitchell (1998)

“Learning” in Machine Learning

Example1: A handwritten digit recognition learning problem

Task T : recognizing and classifying digits within images

Performance measure P : percent of digits correctly classified

Training experience E : a database of handwritten digits with given classifications

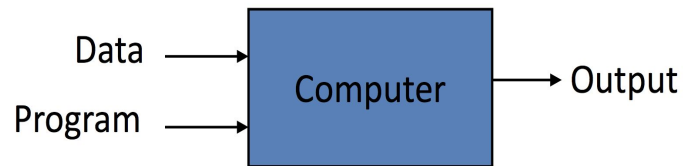
Example2: A robot driving learning problem

Task T : driving on public roads with vision sensors

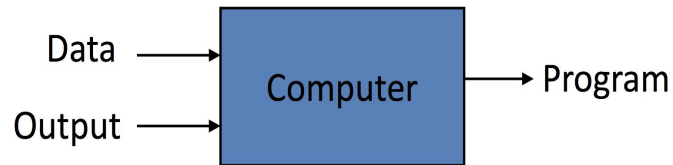
Performance measure P : average distance traveled before an error as judged by human overseer

Training experience E : a sequence of images and steering commands recorded while observing a human driver

Traditional Programming



Machine Learning

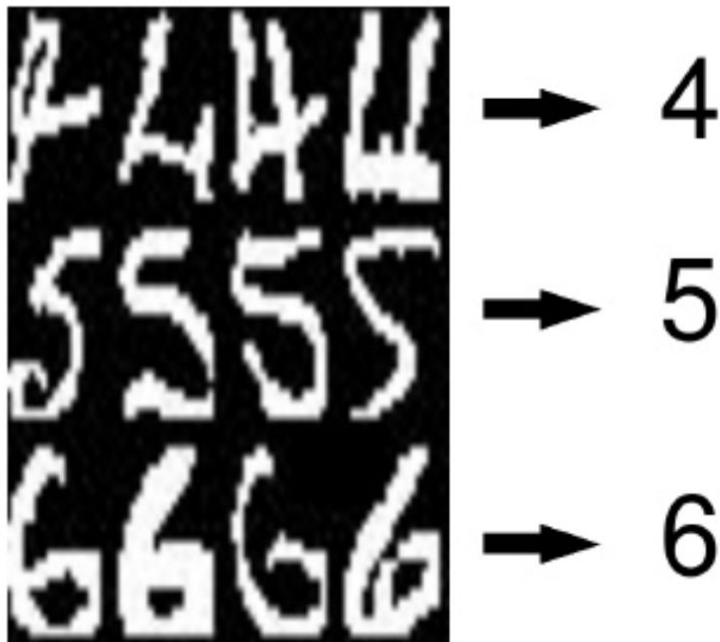


Credit: Tom Mitchell (1998)

Supervised Learning

Given a data set, try to infer a relationship between input and output variables. Given X , make a prediction of a y value.

Examples: Handwriting recognition, Cancer diagnosis, Spam and Fraud detection, poverty prediction, house price prediction



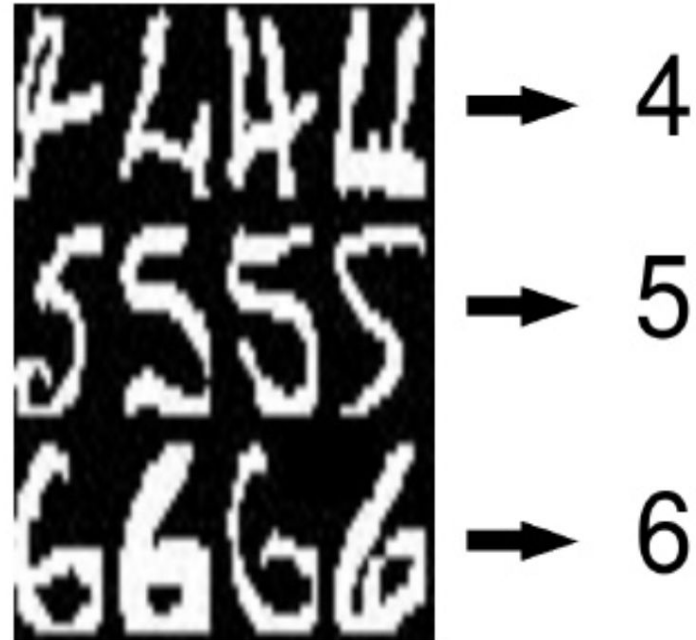
Supervised Learning

Regression: map input variables to some **continuous** function (output)

Eg. Predicting housing price from different features of houses like area, number of rooms, latitude, longitude is regression as it is predicting the continuous variable (housing price)

Classification: map input variables to some **discrete** function (output)

Eg. - spam filtering, digit recognition, object detection, poverty prediction

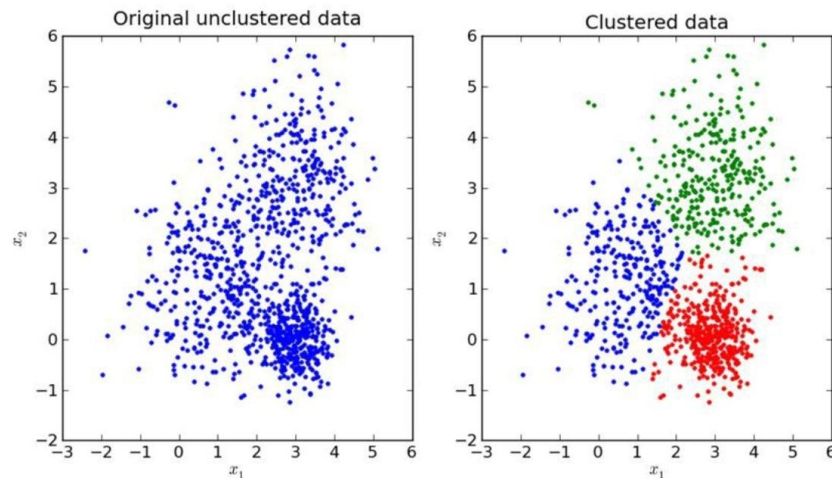


Unsupervised Learning

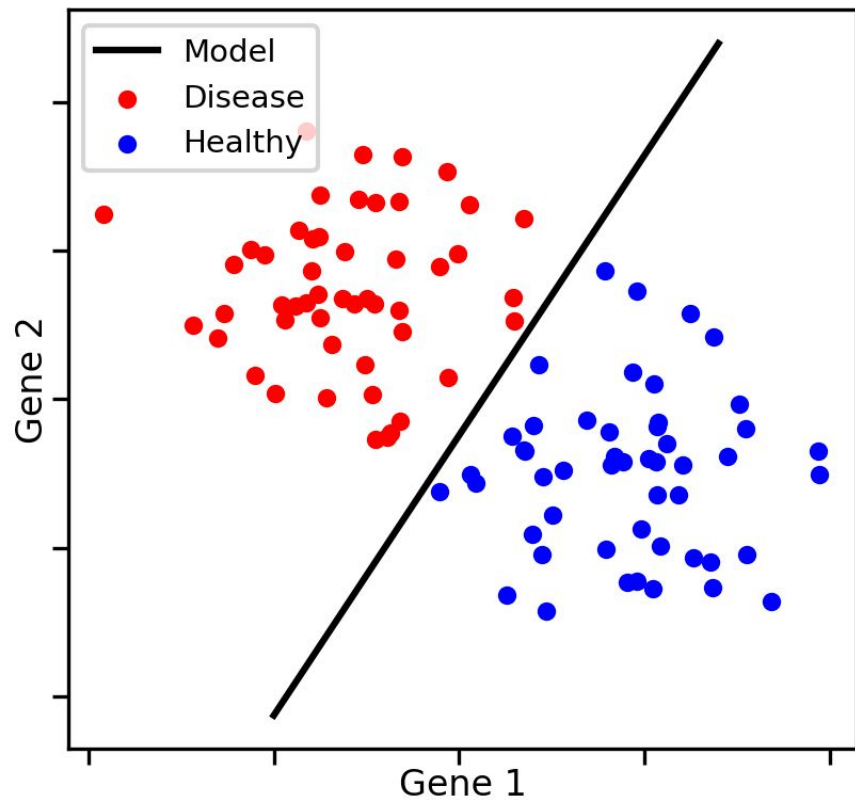
No output variables or labels in the data set. Derive structure/trends based on the relationships among the variables in the data

Clustering: Grouping data into different clusters.

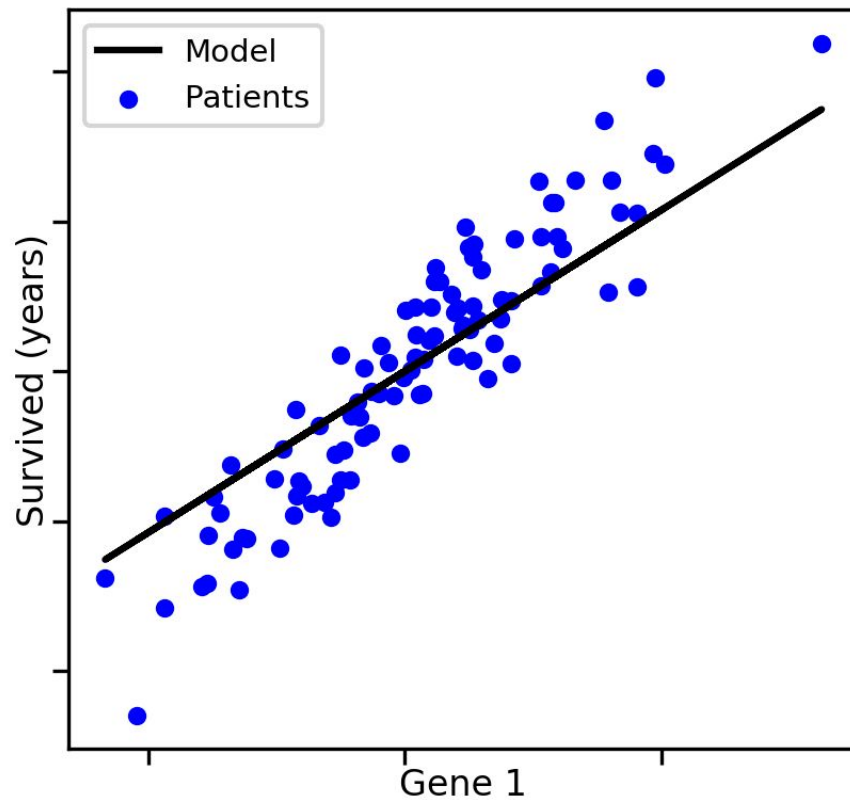
Examples of unsupervised learning are customers segmentation to do targeted marketing, movie recommendation engine etc.



Classification

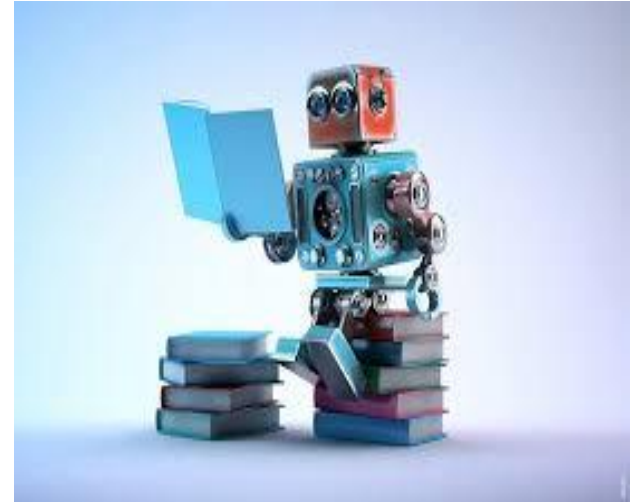


Regression



Many algorithms

- Linear Regression
- Logistic Regression
- Decision Trees
- Random Forest
- K-means clustering
- Neural Networks
- Principal Component Analysis
- Gradient Boosting algorithm
- Naive Bayes Algorithm



Machine Learning Workflow

1. Getting the data
2. Data Preprocessing
3. Models Research
4. Training and Testing
the model
5. Evaluation

Logistic Regression

Open the notebook and copy it to your Google Drive

https://colab.research.google.com/drive/1LeqK14vDQlpNFVrmk_8x-CY0Rr9atT2w

The notebook covers topics on

1. Cleaning the data
2. Handling Categorical data for Machine learning algorithms
3. Logistic Regression with Scikit Learn
4. Splitting data for training and testing
5. Model Accuracy
6. True Positives, False Positives, True Negatives and False Negatives
7. Type I and Type II error