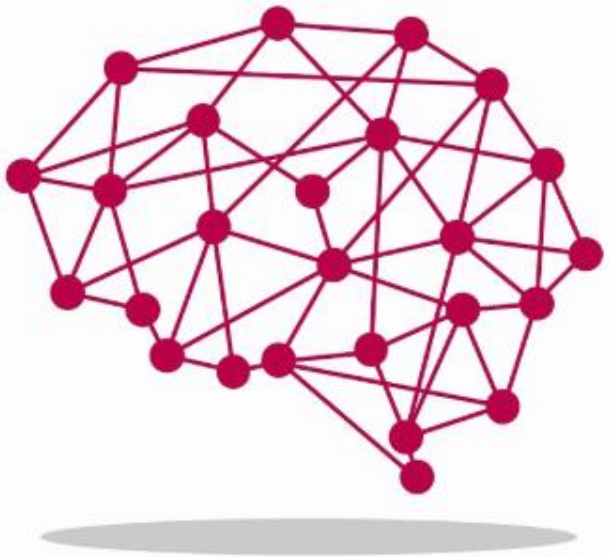
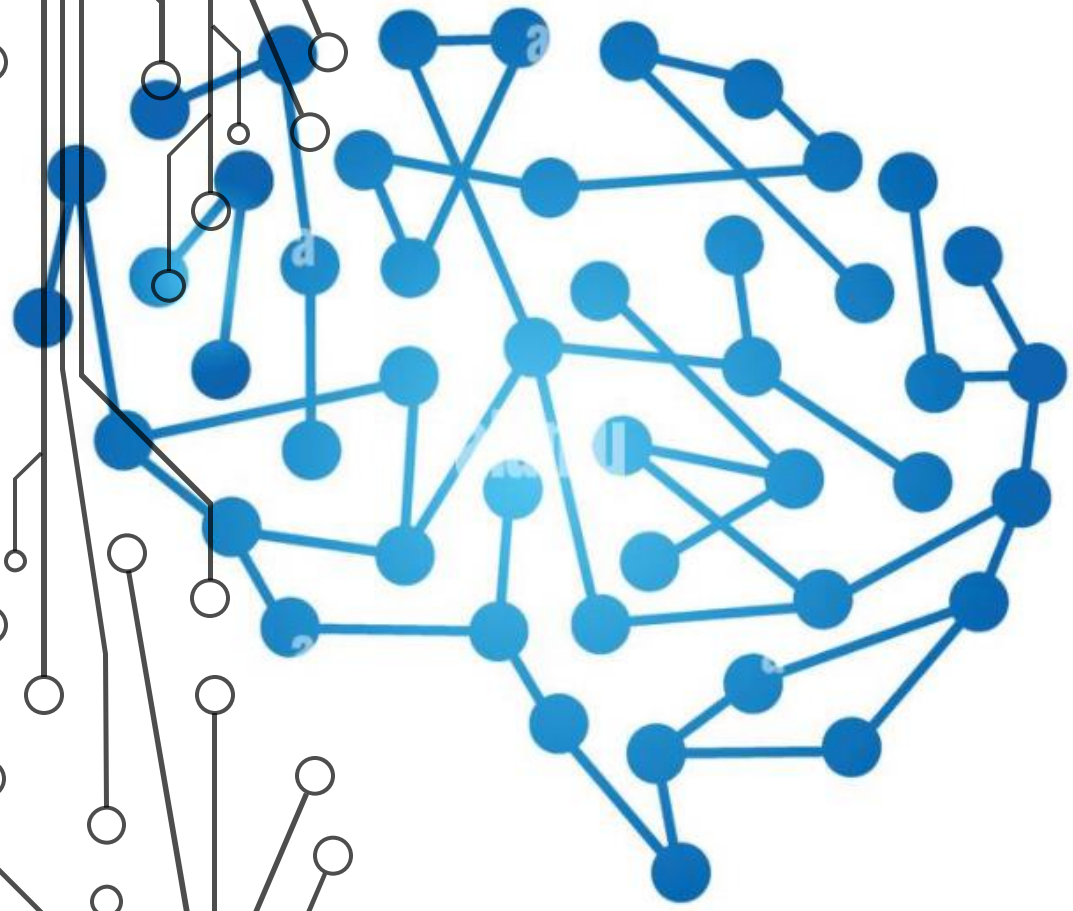
C379 EMERGING TECHNOLOGIES

LESSON 15: INTRODUCTION TO MACHINE LEARNING



L15 LEARNING OBJECTIVES

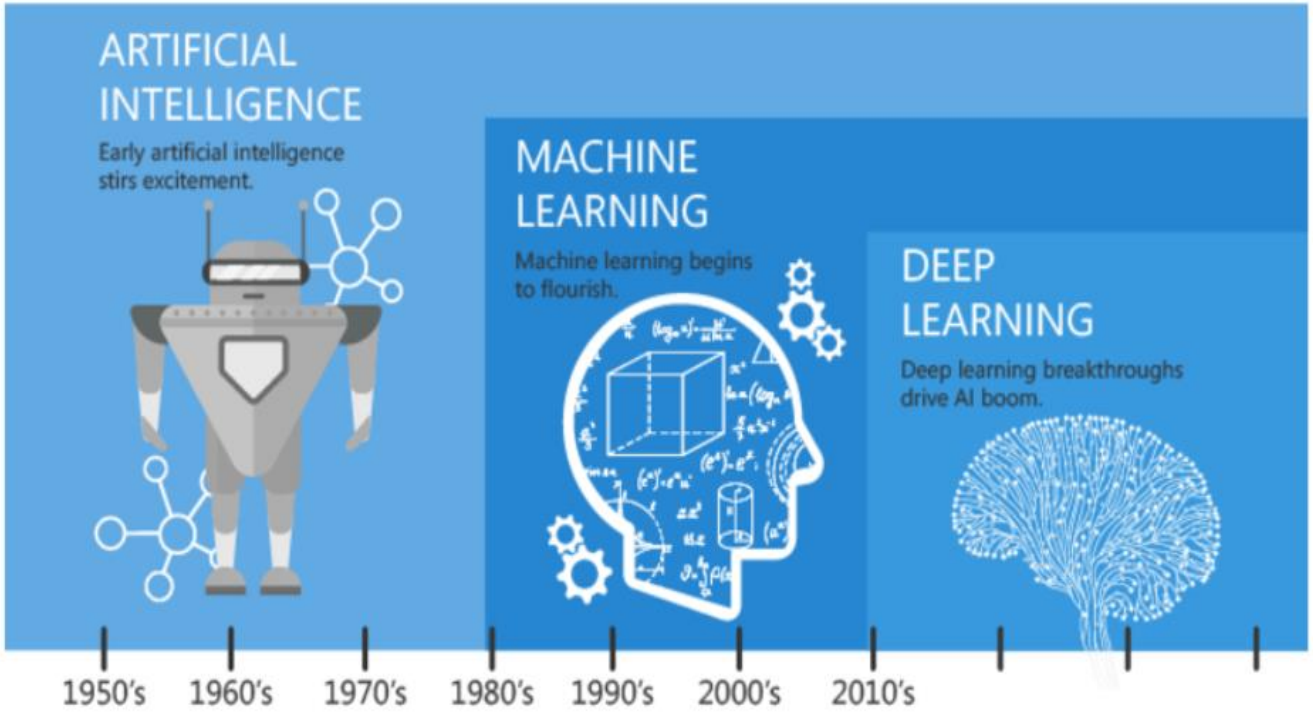
- Understand the fundamental concepts of machine learning, including supervised, unsupervised, and reinforcement learning.
- Explain the difference between classification, regression, and clustering problems.
- Understand the workflow of Machine Learning
- Identify the challenges and limitations of AI



MACHINE LEARNING

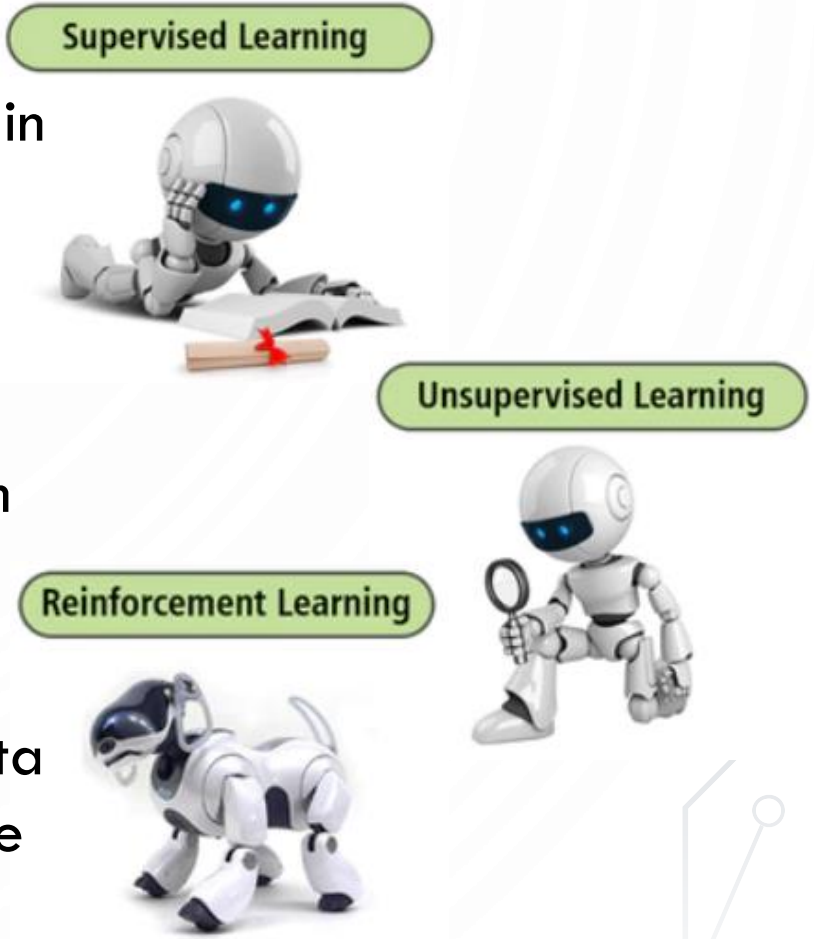
MACHINE LEARNING

- AI describes the broad approach of using machines to imitate intelligent human behavior to solve problems.
- Machine Learning (ML) is a technology used to achieve AI.
  - If one will to develop an algorithm to detect fraud in financial data, this would be typical AI.
  - If this algorithm still learns itself and also recognizes new facts, it would be called ML.



TYPES OF MACHINE LEARNING

- Machine Learning (ML) is a component of AI
- It is the ability for an algorithm to learn from **prior data** in order to produce a **desired behaviour**.
- It teaches machines to **make decisions** in situations they have never seen.
- The mainstream approach to ML is showing the algorithm a **data set** of situations and telling it what the **right decision** is, through training a model.
- Once the model has been trained, you can feed new data through the algorithm. Hopefully, the machine could make *intelligent* decisions from these new situations.



HOW MACHINE LEARNING WORKS?

https://www.google.com/search?q=Ai+adventure+7+steps&rlz=1C1GCEU_enSG925SG925&oq=Ai+adventure+7+steps&aqs=chrome..69i57.12777j0j4&sourceid=chrome&ie=UTF-8#kpvalbx=_yBiSYarpGZvb9QPot6HYCg33

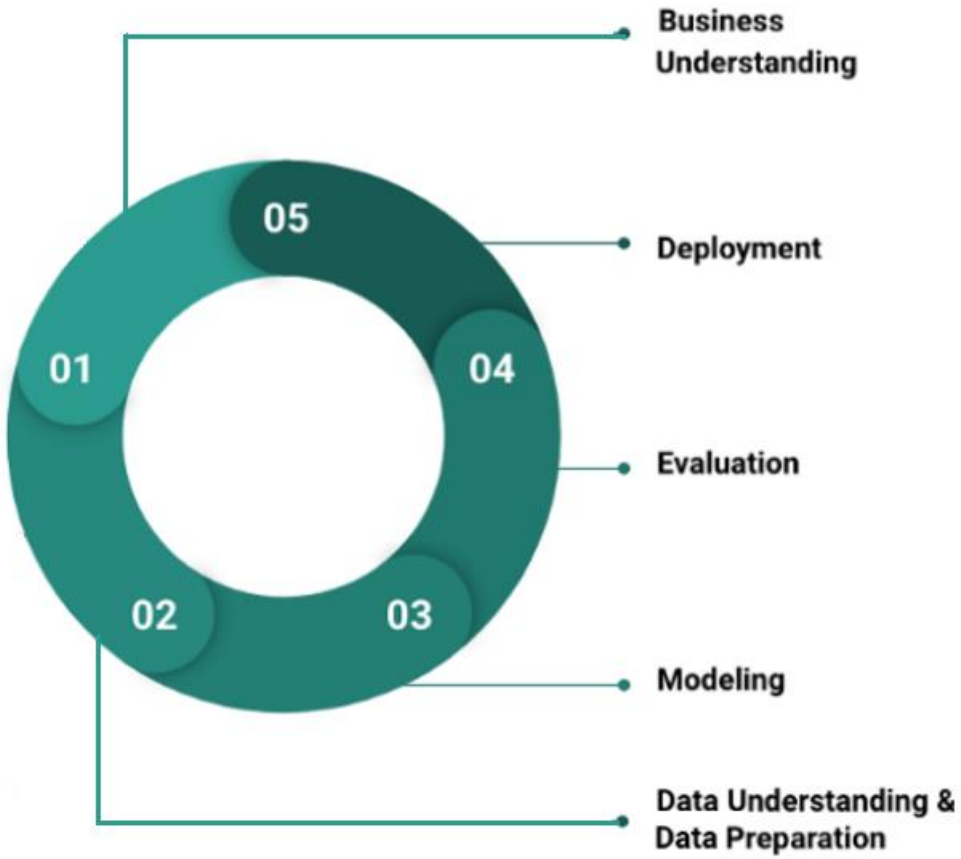
TEAM DISCUSSION

- Discuss within your team what type of machine learning is best/suitable for your assigned C379 project?
  - Supervised Learning
  - Unsupervised Learning
  - Reinforcement Learning
- Why do you think so?

WORKFLOW OF MACHINE LEARNING

The predictive analytics process is as follows:

1. Problem understanding and definition
2. Data collection, preparation and understanding
3. Model building
4. Model evaluation
5. Model deployment



TOOLS

- Programming Languages
  - Python, R, C#, SQL, etc.
- Services
  - Google, AWS and Azure. Often used services are e.g. image and video recognition, translation, NLP and many more.
- Software
  - Anaconda, Jupyter Notebook, R Studio, Tableau, etc.

CHALLENGES & LIMITATIONS TO AI

1. **Data**

   Since data is sourced from diverse sources, it is unstructured and requires technological processes and significant human effort to cleanse it.

2. **Increases Human Effort**

   Not only cleansing data, but also labeling it with the appropriate parameters so that machines can learn correctly, involve significant human effort.

3. **Lacks Emotions**

   Unlike humans, machines lack emotional intelligence and cannot comprehend the feelings behind the spoken words/sentences.

4. **Requires Supervision**

   Machines can only do what they are programmed or 'trained' to do. They can crunch data in real-time, but they cannot make judgments that require empathy. Similarly machines lack original thinking and creativity.

5. **High Costs**

   AI is a costly technology and requires significant investments to set up the infrastructure. The software needs regular updates and maintenance to adapt to the changing business dynamics.

TEAM DISCUSSION

- Discuss within your team which challenges/limitations will likely to impact your C379 project (e.g. Washing Cycle Prediction)?
- Why do you think so?

TOOLKITS REQUIRED FOR PREDICTIVE MODELING

- **Jupyter notebooks:** interactive coding and visualization of output

- **NumPy, SciPy, Pandas:** numerical computation

- **Matplotlib:** data visualization

- **Scikit-learn:** machine learning

INTRODUCTION TO NUMPY

- NumPy stands for Numerical Python
- It is a Python library used for working with arrays (lists)
- It also has functions for working in domain of linear algebra, Fourier transform, and matrices.

Example

Create a 2-D array containing two arrays with the values 1,2,3 and 4,5,6:

```
import numpy as np

arr = np.array([[1, 2, 3], [4, 5, 6]])

print(arr)
```

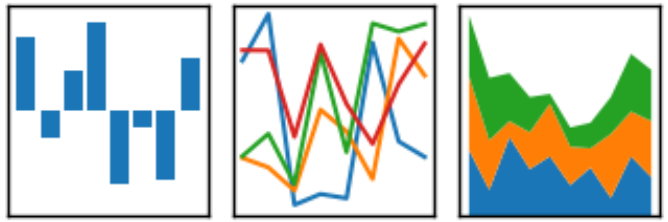
```
[[1 2 3]
 [4 5 6]]
```

INTRODUCTION TO PANDAS

- Library for computation with tabular data
- Mixed types of data allowed in a single table
- Columns and rows of data can be named
- Advanced data aggregation and statistical functions

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Source: http://pandas.pydata.org/

INTRODUCTION TO PANDAS

Basic data structures

Type	Pandas Name
Vector (1 Dimension)	Series
Array (2 Dimensions)	DataFrame

INTRODUCTION TO PANDAS

Pandas DataFrame Creation and Methods

Labeled columns and an index can be added

Code

```
# Add column names to dataframe
activity_df = pd.DataFrame(joined_data,
  index=pd.date_range('20150329', periods=6),
                columns=['Walking','Cycling'])

print(activity_df)
```

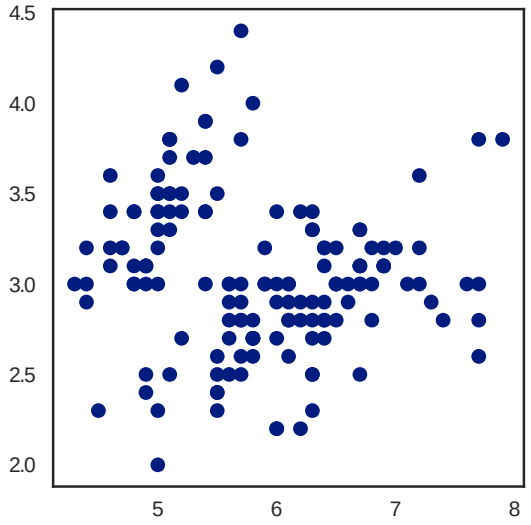
Output

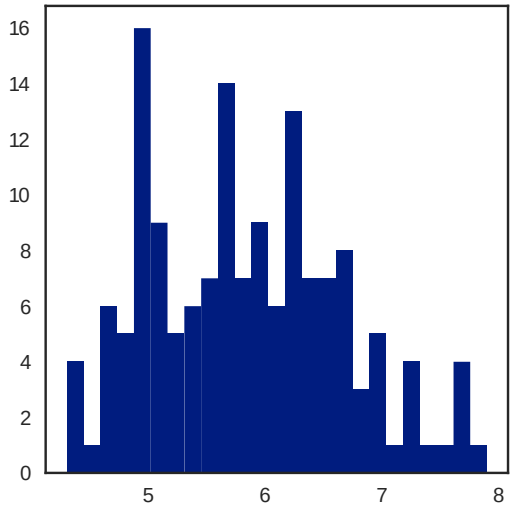
```
>>>
```

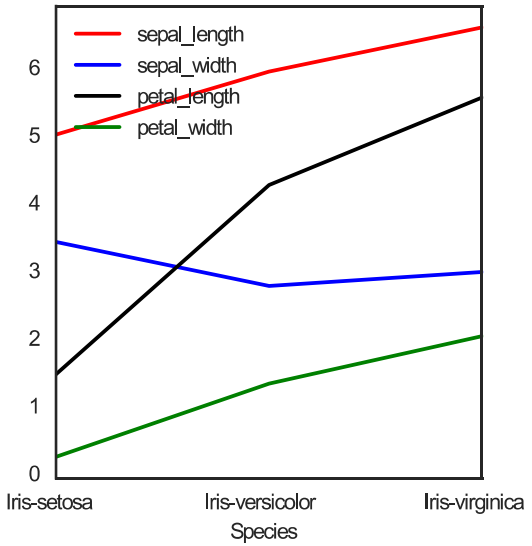
	Walking	Cycling
2015-03-29	3620	10.7
2015-03-30	7891	0.0
2015-03-31	9761	NaN
2015-04-01	3907	2.4
2015-04-02	4338	15.3
2015-04-03	5373	10.9

DATA VISUALISATION

- Matplotlib is the main library to create plots and graphs.
- It has a lot of features and tweaks







DATA VISUALISATION

Basic Scatter Plots with Matplotlib

Multiple layers of data can also be added

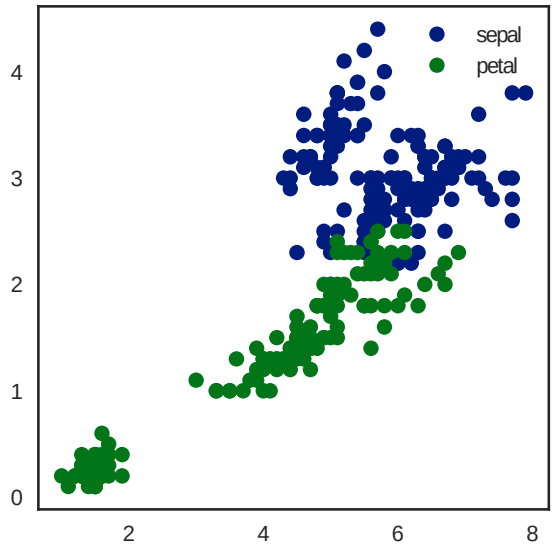
Code

```
plt.plot(data.sepal_length,
         data.sepal_width,
         ls ='', marker='o',
         label='sepal')

plt.plot(data.petal_length,
         data.petal_width,
          ls ='', marker='o',
         label='petal')

plt.legend()
```

Output



GOOGLE COLAB

- Google Colab, short for Colaboratory, is a free cloud-based platform provided by Google that allows users to write and execute Python code collaboratively in a Jupyter Notebook environment.
- Google Collaboratory notebook, is designed to facilitate machine learning (ML) and data science tasks by providing a virtual environment, Google colab python with access to free GPU resources.

COLAB TUTORIAL

https://www.youtube.com/watch?v=g0xu9DA4gDw

Duration: 12:33

GETTING YOU FAMILIAR WITH A DATA SCIENCE STACK

LAB 15-1

- COMPLETE THE LAB MATERIALS FOR NUMPY & MATPLOTLIB