

# Introduction to Machine Learning

TA Session - 1

8th Aug 2022



## Some general info

- **Official Github Repository:**

[https://github.com/vijay-jaisankar/ML\\_TA\\_IITB\\_2022](https://github.com/vijay-jaisankar/ML_TA_IITB_2022)

- **Where you can reach me:**

Slack (preferred) or Mail

- **Find Good Resources?**

Share it on the slack channel!

Make a PR here - <https://github.com/zense/helpful-resources>



## Some questions you might have:

- Do I need a fancy computer? What interfaces are we going to use?  
No!
- Do I need to be good at programming? (DSA trauma... I understand)  
Don't panic! Basics required.
- Do I need to be a pro at maths?  
Nope.

**Just don't panic.**  
**You'll do it** 😊

—



# What we're gonna do today

- A small general discussion
- Overview on basic components
- Supervised v/s Unsupervised Learning
- A brief explanation of loss function and where it comes into play
- Pen and Paper problem-solving (closed-form and gradient descent)
- Introduction to some tools and libraries used

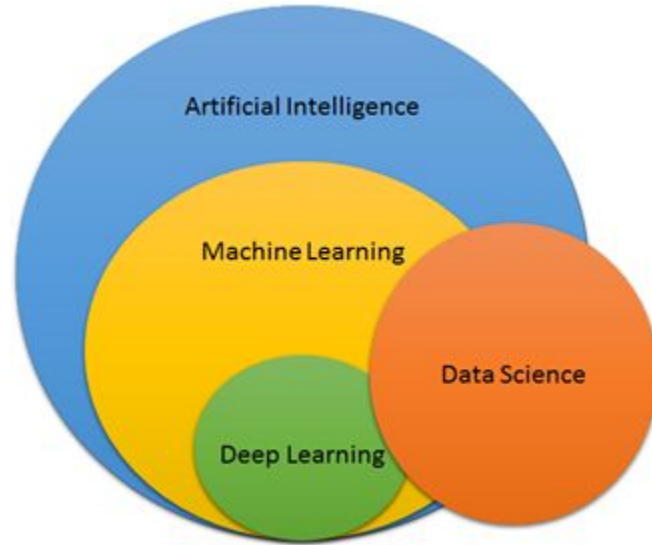
# What is ML?

---

Another word you might've heard  
being thrown around...

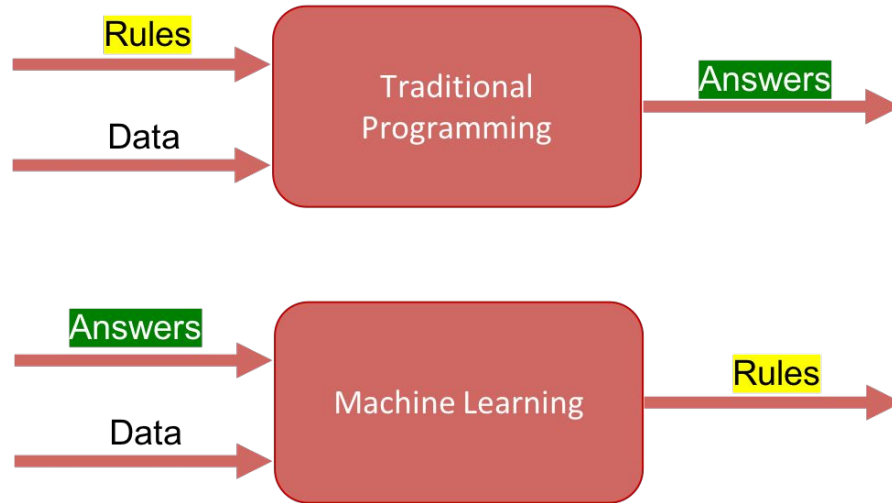
**AI?**

---



Source: <https://medium.com/@dilip.rajani/comparing-ds-ml-dl-and-ai-65627109e67a>





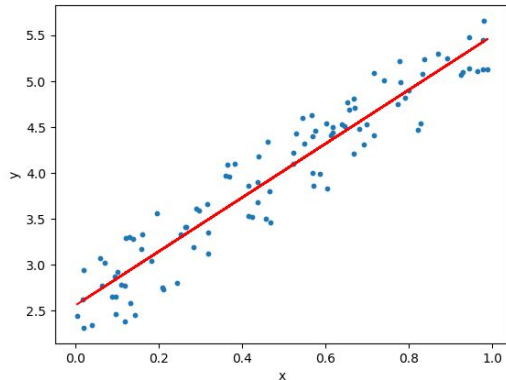
<https://www.congrelate.com/15-programming-without-machine-learning-images/>

# Is it all a scam?

---

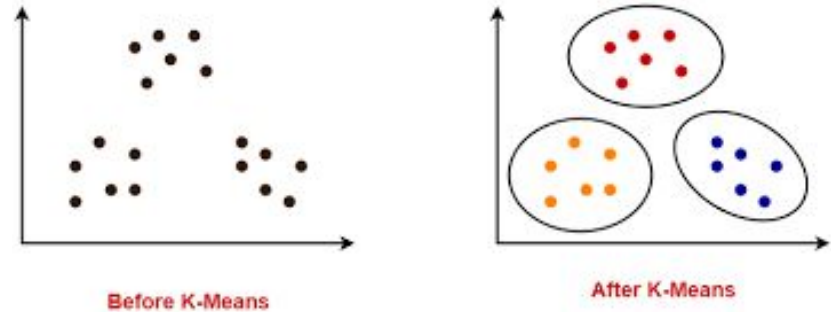
# Supervised v/s Unsupervised Learning

- Where your data has labels, it is supervised learning i.e. someone is telling you what the true label is for the training data.
- Simple linear regression is an example.



<https://towardsdatascience.com/linear-regression-using-python-b136c91bf0a2>

- Unsupervised learning, there are no labels at all. No 'teacher'. No ground rules. No true labels.
- Clustering is an example.



<https://www.gatevidyalay.com/k-means-clustering-algorithm-example/>



# Jargon to keep in mind

- Tasks
- Datasets
- Features
- Models



# Formal definitions

- **Tasks** - Type of prediction being made, based on the question that is being asked and the available data.
- **Datasets** - The raw data available.
- **Features** - The factors or characteristics taken into consideration.
- **Models** - Almost like a function. Algorithms are used to tune this model.

Let's discuss an  
example

—

**What is loss function?**  
**Why is it important?**

—

# Problems.

---





# Libraries and Tools

- Jupyter
- Kaggle - Hosts Datasets, Notebooks, and Contests
- Colab - Interface Google Drive with Jupyter
- Pandas - Dataset operations
- Numpy - Mathematical functions
- Matplotlib - Plots and Charts



## Good Resources

- Personally, B1(Burkov, Andriy - The hundred-page machine learning book-Andriy Burkov) which sir might've already suggested. Do give it a read!
- On youtube: StatQuest
- Good articles for beginners on Medium and Towards data science (might have to open in incognito to escape the paywall)
- When you start coding, use the documentations of the libraries. They're all mostly enough

