

GETTING STARTED WITH MACHINE LEARNING

LECTURE 1

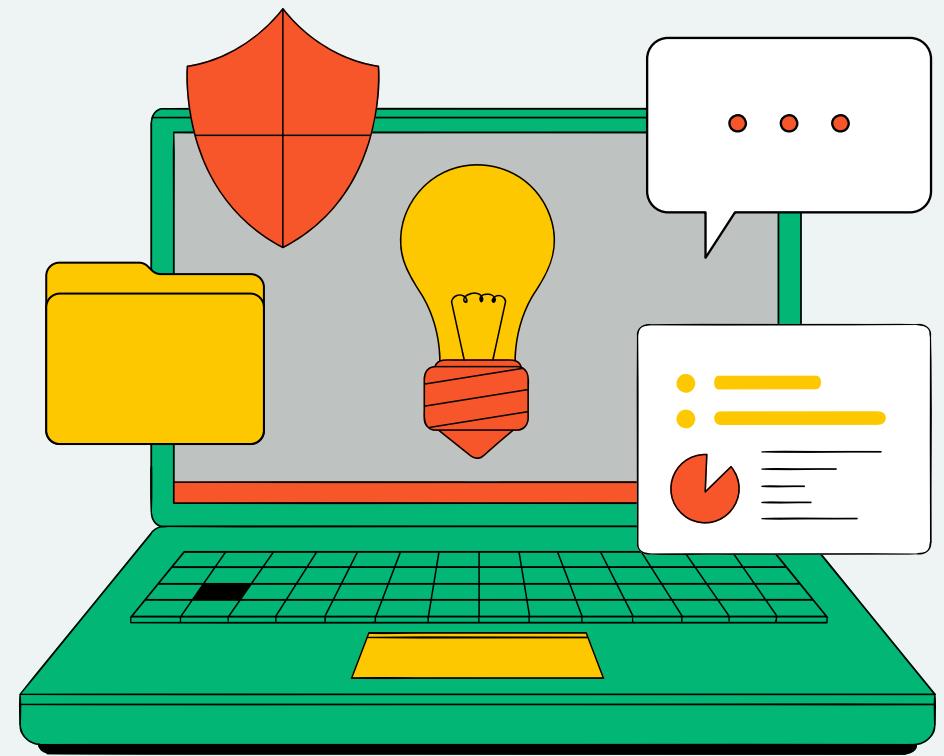
OUTLINE

- What is Machine Learning?
- Why is Machine Learning Important?
- Terminology of Machine Learning
- The Many Flavours of ML
 - Supervised Learning
 - Unsupervised Learning
 - Semi-supervised Learning
 - Reinforcement Learning
 - Active Learning
- Supervised Learning
 - Regression
 - Classification



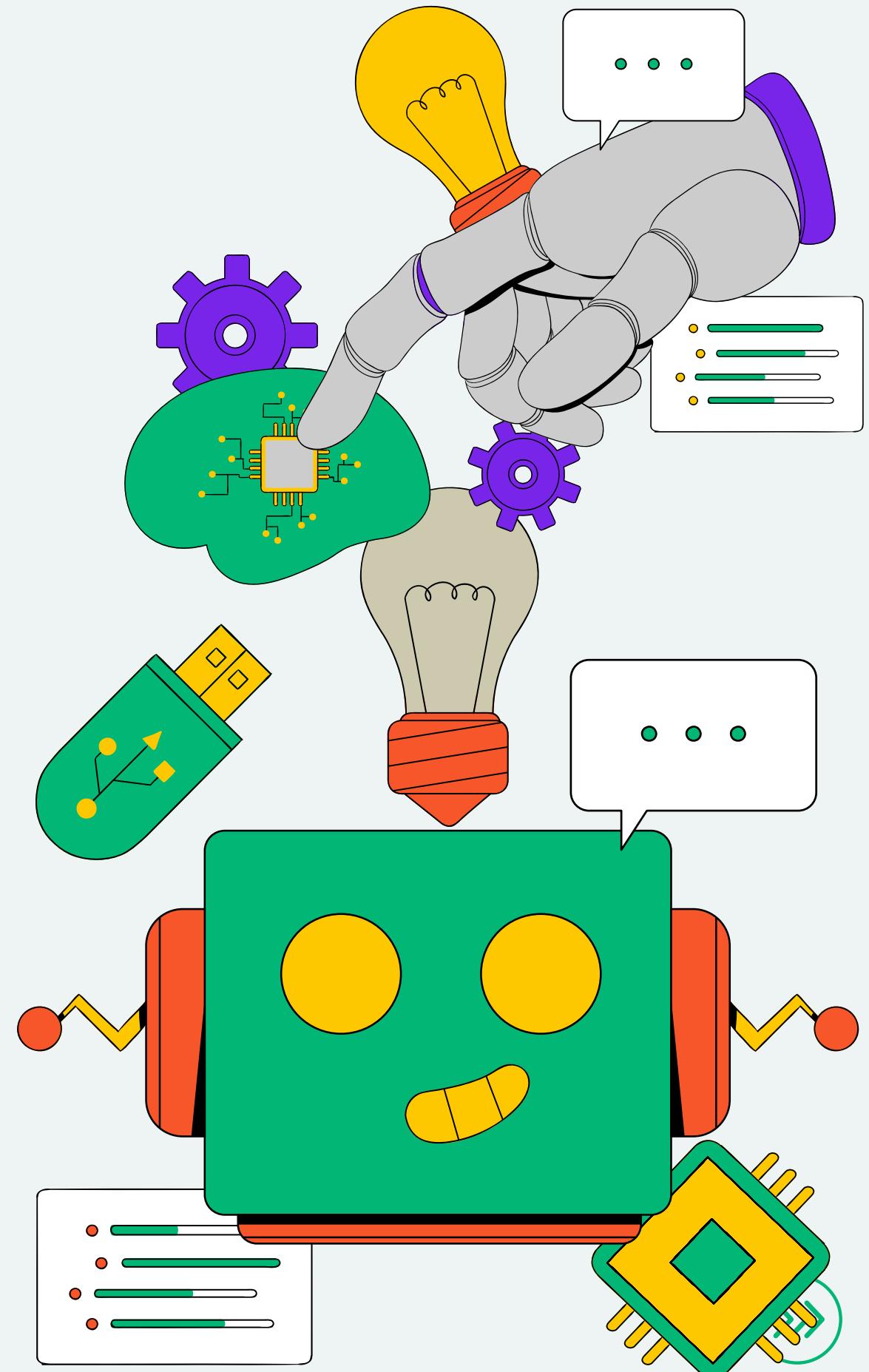
WHAT IS MACHINE LEARNING?

- A branch of artificial intelligence, concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data.
- In traditional programming paradigms, programmers are seen writing programs **P** such that for a given set of inputs **I**, the set of outputs **O** is produced when **P** acts on **I**.
- However, in machine learning, the computer is provided with the set of inputs **I**, and the corresponding outputs **O**(or not :)) to get a program **P** which can act on unseen data **I'** which is similar to **I**.
- Just like humans are expected to learn from experiences, computers can learn from data, which can represent some “past experiences” of an application domain.

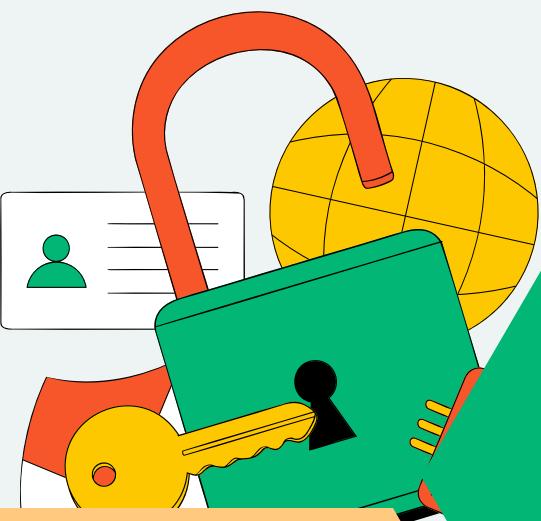


WHY IS MACHINE LEARNING IMPORTANT?

- A lot of times we come across problems in life where it is hard to define algorithms for a procedure
- For instance think of the simplest example possible, handwritten digit recognition. Can you think of an algorithm which would be suitable for this task ?
- Remember the sheer amount of variety the data can have ! (different handwriting, different font ... and so on)
- In such a case it becomes difficult to handweave an algorithm that can be capable of working successfully all the time
- This calls for the need of mechanism that would be able to learn by itself, what to look for when its trying to classify a digit



TERMINOLOGY OF ML



Data

This refers to the information that is fed into the computer of which it is expected to learn patterns from. Data is composed of one or more datapoints, which may be labelled or unlabelled.

Labels

The datapoints in the dataset may be associated with one or more labels which can be considered the final output.

Features

This refers to the various available inputs that we may (or may not) use in order to predict the target value

Model

It refers to the mechanism that would take the features and then use them to predict the labels

Loss Function

A function that governs how off the model's predictions are with respect to the ground truth

Training

This refers to the process of computation of the model parameters in order to give the best predictions. Such computations work in order to minimize the loss function



Testing and Inference

This refers to using the trained model to obtain predictions on unseen data and evaluate its performance



THE MANY FLAVOURS OF MACHINE LEARNING

01

SUPERVISED LEARNING

In this machine learning paradigm, the datapoints of the training dataset contain both features and the labels.

Examples of subtypes include regression, classification etc.

02

UNSUPERVISED LEARNING

In this machine learning paradigm, the datapoints of the training dataset contain only features., and no labels.

Thus, the model relies on grouping similar datapoints together, relying on the intrinsic pattern of input data. Examples of subtypes include clustering, anomaly detection etc.

03

SEMI-SUPERVISED LEARNING

Semi-supervised learning is a type of machine learning that lies between supervised and unsupervised learning. It uses a combination of a small amount of labeled data and a large amount of unlabelled data. This is done to increase model accuracy as well as to economize since getting labelled data can be difficult.



04

ACTIVE LEARNING

This is a special case of supervised machine learning. In this a human annotator is used to increase the accuracy of our model.

The model training process begins with labelled data. This trained model is then applied on unlabelled datapoints. The model selects a subset of unlabelled datapoints on which it has low confidence, and asks the user to label these.

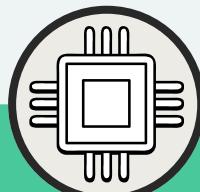
05

REINFORCEMENT LEARNING

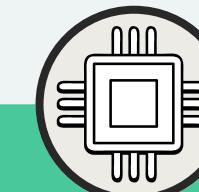
Reinforcement learning generally involves an agent that works by taking random actions from a set of actions, and is rewarded or punished for doing the same, if the result is desired or undesired respectively. For example, training an agent to play chess intelligently, so that its chances to win are maximised.



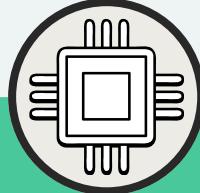
SOME EXERCISES!



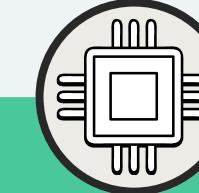
The mails in your gmail inbox are labelled primary, updates etc(provided you're organised :))
Which ML paradigm is this?



Some of you may dropout(college is hard :)) and decide to become realtors. You will have to constantly estimate the price at which your house should be sold. Use ML!



AWS Deepracer.
Google it.
What is the paradigm being employed here?
Your answer can determine your future with Tesla.



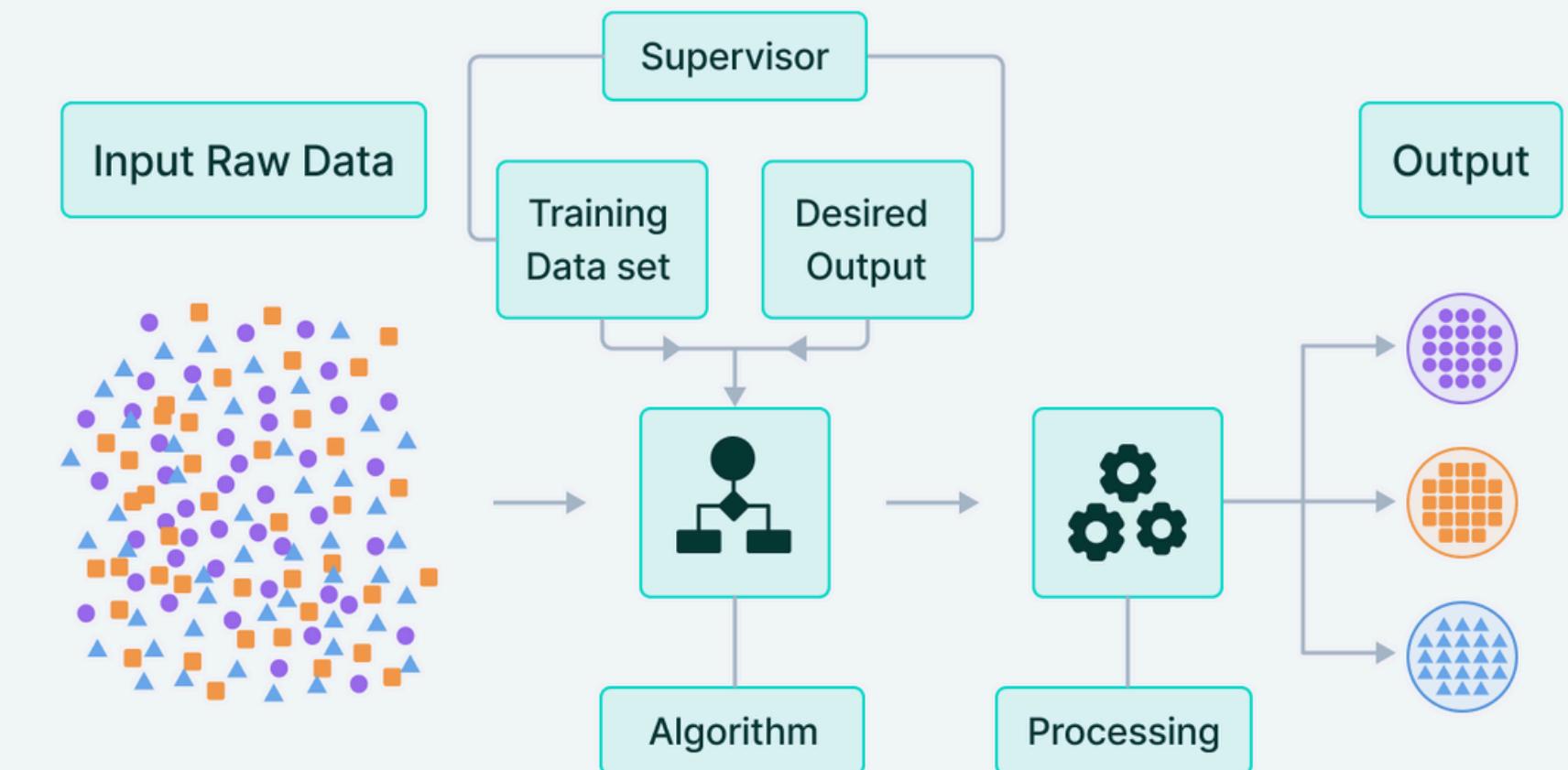
More often than not, things turn out to be faulty. Won't it be nice if we could detect these faults before someone dies?
Can an ML paradigm help us with this?



SUPERVISED LEARNING

- The most basic form of machine learning, it involves the use of labelled dataset to train models to give label predictions on unseen data.
- For example, dog cat classification with images labelled as dog or cat
- The aim of supervised learning is to construct a map from inputs to outputs which can give accurate predictions on unseen data
- Based on the type of target to be predicted, supervised learning is divided into two type

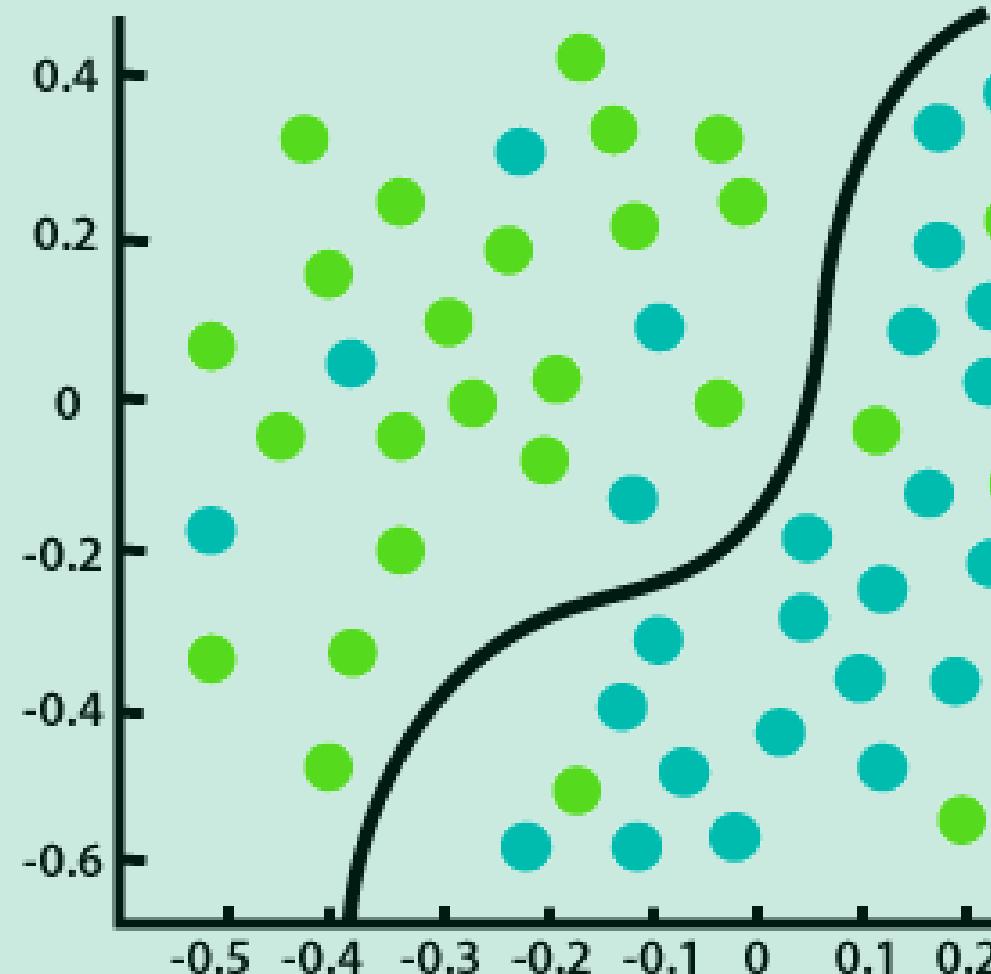
Supervised Learning



SUPERVISED LEARNING

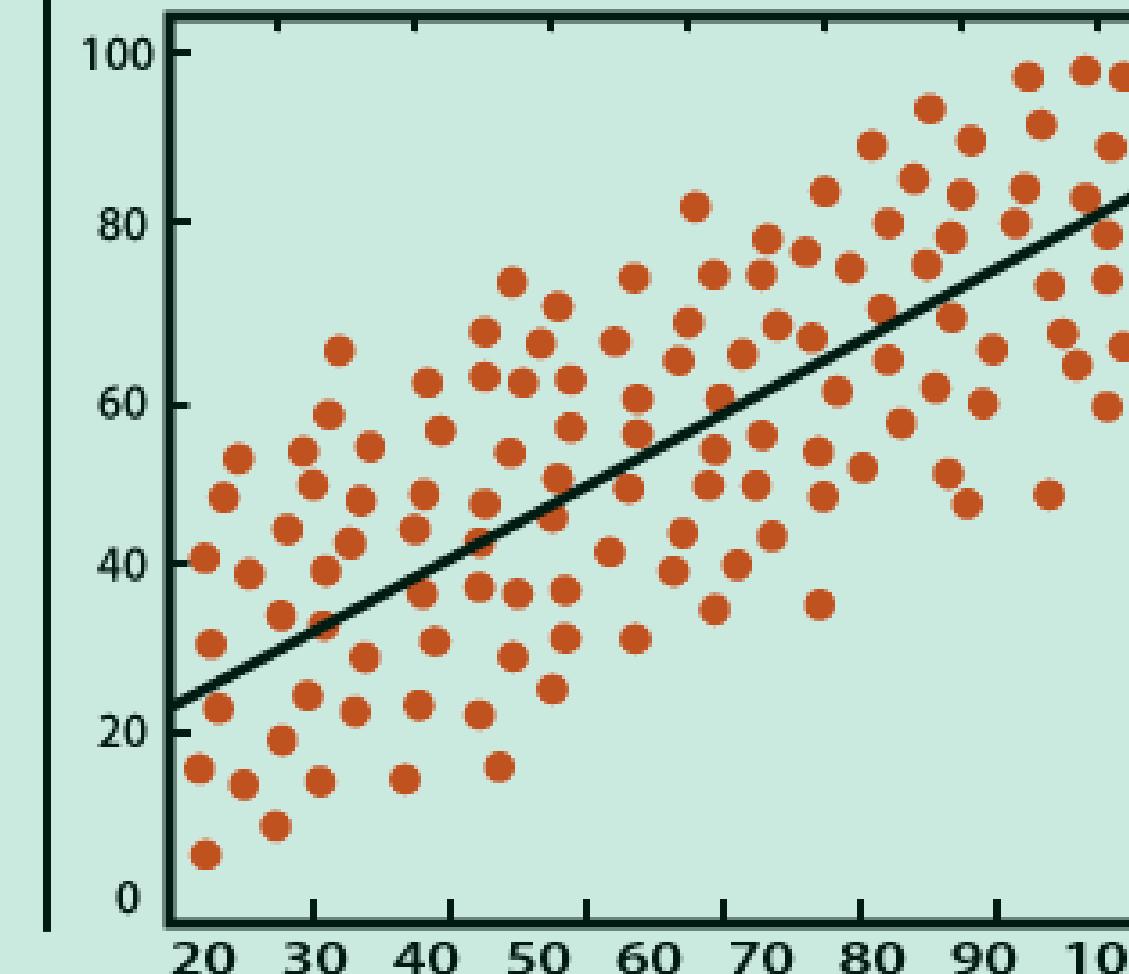
Classification

- This involves predicting discrete values, that are classes of the inputs
- For example, spam email verification, where our motive is to identify if an email is spam or not



Regression

- This involves prediction of continuous numerical values using the inputs
- For example, weather forecasting



REGRESSION

- **Goal**: given a set of inputs and corresponding outputs, create a mapping f such that the loss function between the predicted values using the mapping and the ground truth values is minimized
- Mathematically, consider the Mean Squared Error (MSE) loss function

$$L(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Where n is the number of samples, Y_i is the actual output for the i th sample, and \hat{Y}_i is the predicted output for the i th sample.

- We try to find the f such that this loss function is minimized , that is

$$\min_f L(Y, f(X))$$

- How do we do that ? Let's look at a sophisticated algorithm called **Gradient Descent**

GRADIENT DESCENT

1. **Initialize** the function f with some random parameters
2. **Forward Propagation**, for all inputs x evaluate $f(x)$ and find the Loss function, $L(x, f(x))$
3. Find the **Gradient** of the loss function, that is the set of partial derivatives of the loss function with respect to the function parameters.
4. For all parameters theta, update them as follows

$$\theta := \theta - \alpha \nabla J(\theta)$$

5. Repeat the procedure from step 2, until the loss function is minimized

WHY IT WORKS

