Learning Objectives

What is Machine Learning?

Machine Learning Categorization

Classification and Regression

DPhi

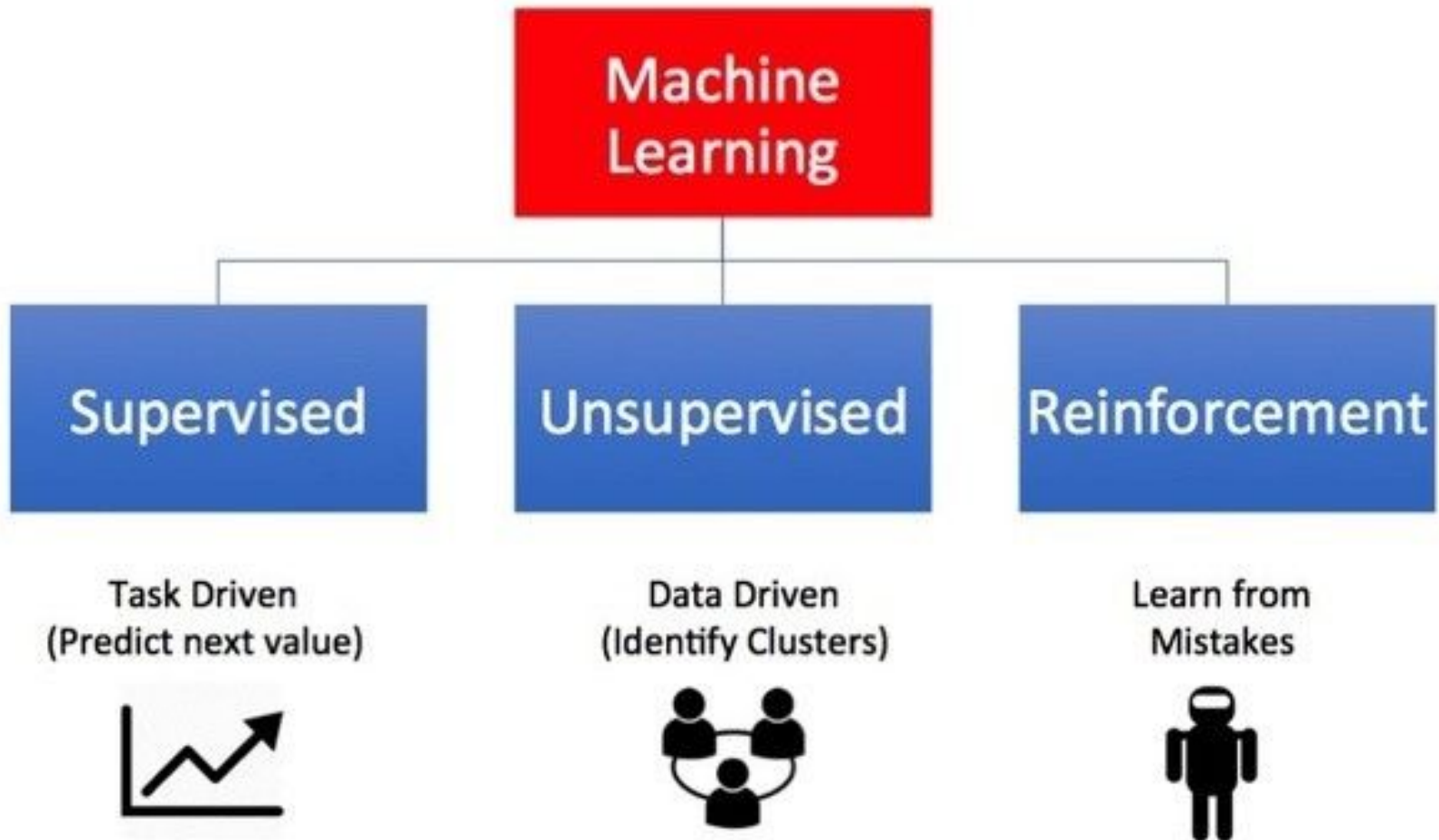# What is Machine Learning?

Machine Learning Categorization

Supervised Learning Algorithms

Let's talk about the datasets that have both input variables and target variables (labels for the data). Ranging from predicting the survival rate of a person in Titanic Dataset where Survival Rate is already given to predicting the House Price according to house characteristics where the house prices are provided.

The algorithms that work on such datasets are known as **Supervised Learning Algorithms.**

It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

DPhi

Unsupervised Learning Algorithms

Unsupervised learning is where you have unlabeled data (or no target variable) in the dataset.

The goal of Unsupervised Learning Algorithms is to **find some structure in the dataset**.

These are called unsupervised learning because unlike supervised learning, there are no correct answers and there is no teacher. Algorithms are left to their own to discover and present the interesting structure in the data.

DPhi

Reinforcement Learning Algorithms

A robot takes a big step forward, then falls. The next time, it takes a smaller step and is able to hold its balance. The robot tries variations like this many times; eventually, it learns the right size of steps to take and walks steadily. It has succeeded.

What we see here is called **reinforcement learning**. The robot learns how to walk based on reward (staying on balance) and punishment (falling). This feedback is considered "reinforcement" for doing or not doing an action.

Reinforcement learning, in a simplistic definition, is learning best actions based on reward or punishment.

Types of Supervised Learning Algorithms
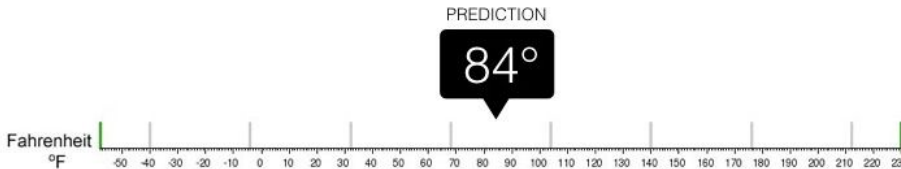
Supervised Learning can be further divided into 2 types:
1. Classification
2. Regression
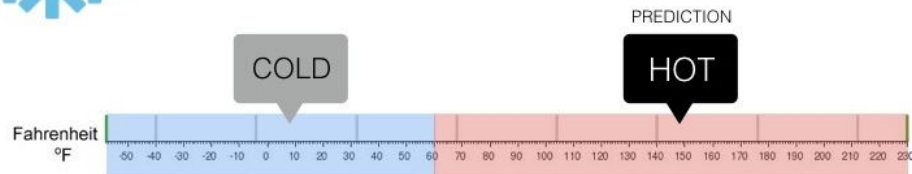
DPhi

Classification vs Regression

Regression

What is the temperature going to be tomorrow?

PREDICTION

84°

Fahrenheit
°F   -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230

Classification

Will it be Cold or Hot tomorrow?

PREDICTION

COLD   HOT

Fahrenheit
°F   -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230

In order to decide whether to use a regression or classification model, the first questions you should ask yourself is:

Does your target variable have a continuous value or is it discrete (binary or multi-class)?

DPhi

Regression

If your answer is continuous values, you're dealing with **Regression**.

This means that if you're trying to predict quantities like height, income, price, or scores, you should be using a model that will output a continuous number.

So if your objective is to determine tomorrow's temperature, you should use a regression model.

DPhi

Classification

Let's come to the second case where if you can clearly see that the target variable is divided into classes. You'll be using **Classification**.

- When the number of classes is 2, it is known as Binary Classification. Eg. Will it be hot or cold tomorrow is a binary classification problem with 2 classes: Hot and Cold.

- When it is more than 2, it is known as Multi Class Classification. Eg. Classifying movies in Good, Average or Bad according to reviews.

DPhi

What is Classification?

Let's learn with some examples:

- In **Classification we** classify the outcome
- **Examples:**
  - Predict whether a transaction is fraud or not fraud
  - Predict whether to give loan or not
  - Predict whether to give college admission or not
  - Predict the grade (Grade A, B, C, D)
  - Note: Classification can be more than two

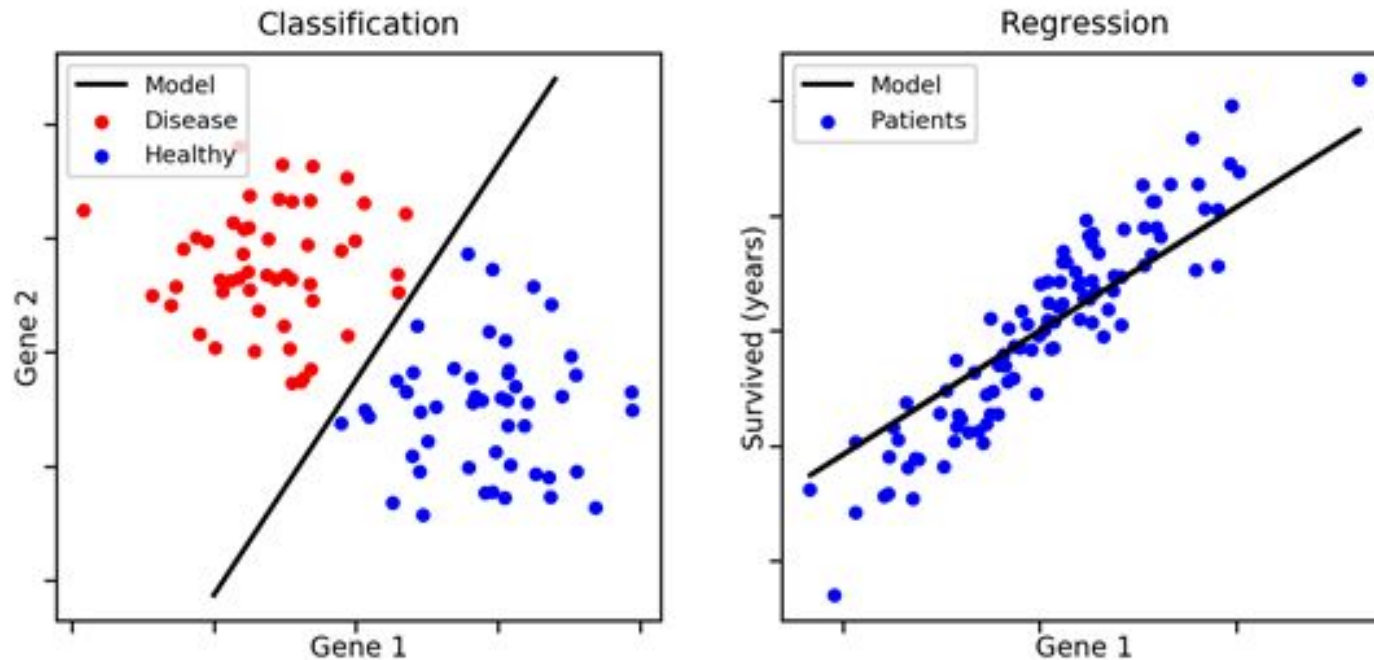|	
Feature	Tumor Age and Tumor Size
Label	Tumor (Benign or Malignant)
Goal/ Aim	We want to predict whether a tumor is benign or malignant from the given age and tumor size

Age

○ Benign
X Malignant

Tumor Size

Phi
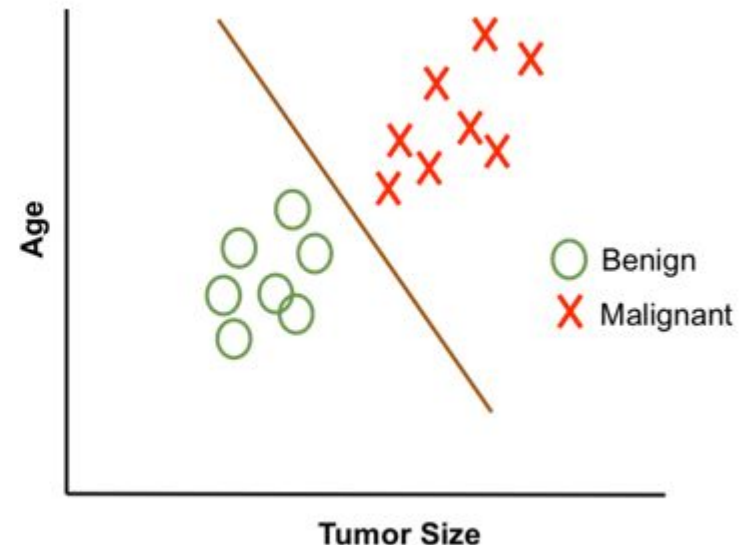
What is Multi-Classification?

It is as simple as dividing waste into 4 categories - plastic, glass, metal, paper

Classification vs Regression

In the above image, you can see that the classification line is dividing the data into 2 parts or 2 classes - red and blue. On the other hand, the regression line is going along the direction of data and not segregating it.

It's important to understand the characteristics of your target variable before you begin running models and forming predictions.

DPhi

Slide Download Link

You can download these slides from the below link:

[https://docs.google.com/presentation/d/1Ope6aITBTCnGtF-9-FxZvm OvfoovaszqAR4neGoIxjM/edit?usp=sharing](https://docs.google.com/presentation/d/1Ope6aITBTCnGtF-9-FxZvmOvfoovaszqAR4neGoIxjM/edit?usp=sharing)

DPhi

That's it for this unit. Thank you!

Feel free to post any queries on Discuss.

DPhi