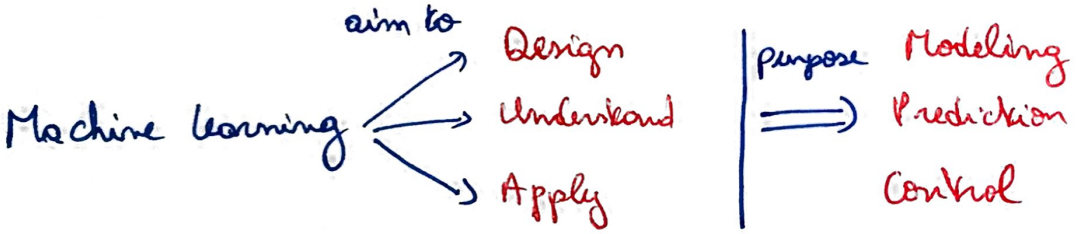
Unit 1 - lecture 1: Introduction to machine learning.

Objectives:

- Understand the goal of machine learning from a movie recommender example
- Understand elements of supervised learning, and the difference between the training set and the test set.
- Understand the difference of classification and Regression - two representative kinds of supervised learning.

What is Machine learning?

Machine learning as a discipline aims to Design, understand and apply computer programs that learn from experience (i.e data) for the purpose of modeling, prediction and control. We will start with Prediction as a core machine learning task.

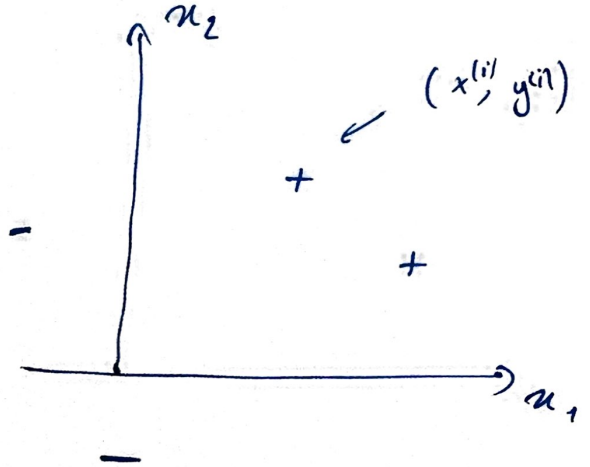


$$S_m = \{ (x^{(i)}, y^{(i)}), i = 1, \dots m \}$$

training set — feature vectors $\in \mathbb{R}^2$ — labels $\{-1, 1\}$

$$h : X \to \{-1, 1\}$$

classifier that maps the $X \to$ labels



The classifier divide the space in 2 parts → part 1 labels +1
→ part 2 label -1

$$\varepsilon_m(h) = \frac{1}{m} \sum_{i=1}^{m} [\![h(x^{(i)}) \neq y^{(i)}]\!]$$

$= 1$ if error
$= 0$ otherwise

fraction of errors = Training error.

Classification maps feature vectors to categories. The number of categories need not be two - they can be as many as needed.

Regression maps feature vectors to real numbers.

A Concrete Example of a supervised learning Task:

We have a movie recommending system that reads description of each movie and determines some important characteristics of the movie. In particular, it examines whether each of the criterion below is true for that movie:

1. Is it a comedy movie?
2. Is it an action movie?
3. Was the movie directed by Spielberg?
4. Do dinosaurs appear in the movie?
5. Is it a Disney film?

When the recommending system reads descriptions of "Jurassic Park", the answers for the five questions above will be "no, yes, yes, yes, no"

$$X_{\text{Jurassic Park}} = [0, 1, 1, 1, 0] \quad \text{(Dimention of } X \text{ is 5)}$$

feature vector

If the person like a movie $\Rightarrow$ label = +1 , dislike = −1

We try to create a recommender system from the training data to predict whether someone will like the movie.

Different types of learning:

Labelled training and test examples → Supervised learning

Using knowledge from one task to solve another task → transfer learning

learning to navigate a robot → Reinforcement learning

Deciding which examples are needed to learn → active learning

Data with no annotation → Unsupervised learning

Training and test examples with limited annotation → Semi-Supervised learning