Machine Learning : Supervised and Unsupervised Learning

Deepak Narayan
University of New Hampshire,
Durham, New Hampshire, USA.

Introduction

- Ability of machines to learn without being explicitly programmed.(Arthur Samuel, 1950s)
- If we have a Task T, and if Performance P increases with Experience E, then the program/machine is said to learn from experience E. (Tom Mitchell, 1990s)

- Don’t let the term ‘Machine Learning’, put you off.
- Under the hood, it is very simple math.

Definitions and Terminologies

Iris Data set (R.A.Fisher, statistician and biologist):

- Target (species of Iris plant) : output variable
- PL, PW, SL, SW : input variables
- Output variable or response variable : our primary focus of interest
- It is also called as dependant variable, since it is dependant on input variables.
- Input variables : Features, Predictors, Inputs

2 kinds of problems are seen :

- Supervised Learning :
- Will have a output/response variable.
- Output variable helps to ‘supervise’ our analysis, will look more at this later.

- Unsupervised Learning :
- Lack of a response variable. Eg : clustering techniques
- Eg : search results from a search engine in good old days.

Definitions and Terminologies

Models :

-They are basically frameworks.
-They can be as simple as a line or can be as complicated as a deep learning neural network containing thousands of layers.
- A model is the algorithm used + parameters calculated by feeding data to it + hyperparameters.

- 2 main types of modelling :
- Explanatory models (this is not our focus today)
- Predictive models (this is our focus)

- A really good podcast on predictive and explanatory models with Galit Shmueli.
•https://www.youtube.com/watch?v=IzZrGBIotro (watch from 17:30).

Definitions and Terminologies

Explanatory models (this is not our focus today):
- used for '*explain*'ing our data.
- for inference.
- Eg : Covid cases
- Used largely in the Statistical community.

- Predictive models (this is our focus) :
- used for '*predict*'ing outcome for new inputs.
- for prediction.
- Data Driven
- Eg : Weather prediction, cancer detection
- Used largely in the Data Science and Machine Learning community.

- A simple comparison example :
- Is it true that exercising regularly (say 30 minutes per day) leads to lower blood pressure? – Explanatory model
- If I exercise 30 minutes per day to what extent is my blood pressure likely to drop? – Predictive model