

# 01 Introduction

Pradeep Ravikumar(Lecturer), HMW-Alexander(Noter)

January 19, 2017

---

[Back to Index](#)

---

## Contents

<b>1</b>	<b>Info</b>	<b>1</b>
1.1	Final Project . . . . .	1
<b>2</b>	<b>What is Machine Learning?</b>	<b>2</b>
2.1	Different Goals with Different Fields . . . . .	2
<b>3</b>	<b>Three axes of ML</b>	<b>2</b>
3.1	Data . . . . .	2
3.2	Algorithms . . . . .	3
3.2.1	Model-based ML . . . . .	3
3.2.2	Model-free ML . . . . .	3
3.3	Knowledge/Tasks . . . . .	4
<b>4</b>	<b>Machine Learning Subfields</b>	<b>4</b>

## Resources

- [Lecture](#)
  - [Textbook](#)
- 

## 1 Info

### 1.1 Final Project

Kaggle style data analysis project, with leaderboard.

What is Kaggle<sup>1</sup>?

- In 2010, Kaggle was founded as a platform for predictive modelling and analytics competitions on which companies and researchers post their data, and statisticians and data miners from all over the world compete to produce the best models.

---

<sup>1</sup><https://en.wikipedia.org/wiki/Kaggle>

- In April 2015, Kaggle released the first version of their **Scripte**<sup>2</sup> product onto their platform. Scripts allows users to write, run, and publicly share their code on Kaggle. On 8 July 2016, Kaggle renamed its **Scripts** product to **Kernels**.
- In January 2016, Kaggle released their **Datasets**<sup>3</sup> product, making a selection of public datasets available on Kaggle.

## 2 What is Machine Learning?

**Algorithms** that improve their **knowledge** towards some task with **data**.



### 2.1 Different Goals with Different Fields

- Machine Learning: the underlying mechanisms and algorithms that allow improving our knowledge with more data.
- Statistics: the understanding of the data at hand.
- Artificial Intelligence: build an intelligent agent.
- Data Mining: to extract patterns from large-scale data
- Data Science: the science encompassing collection, analysis, and interpretation of data.

## 3 Three axes of ML

- Data
- Tasks
- Algorithms

### 3.1 Data

- Fully observed
- Partially observed
  - systematically not observed.
  - missing some of the time.
- Actively collect/sense data.

---

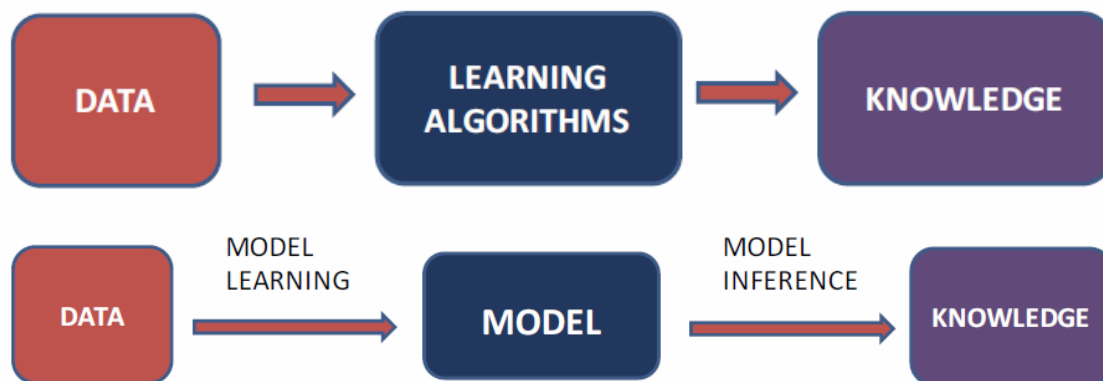
<sup>2</sup><https://www.kaggle.com/kernels>

<sup>3</sup><https://www.kaggle.com/datasets>

## 3.2 Algorithms

- Model-based methods
  - Probabilistic model of the data
  - Parametric models
  - Nonparametric models
- Model-free methods

### 3.2.1 Model-based ML



#### Learning and Inference

- Learning: from data to model
  - A model thus is a summary of the data, and also a story of how the data was generated.
  - Could thus be used to describe how future data can be generated.
- Inference: from model to knowledge
  - given the model, what can we answer to the questions.

#### Parametric and Nonparametric

- Parametric models
  - Fixed-size models that do not grow with the data.
  - More data  $\Rightarrow$  fit the model better.
  - Model: data = point on line + noise
- Nonparametric models
  - Models that grow with the data.
  - More data  $\Rightarrow$  more complex model.
  - Model: data = point on smooth curve + noise

### 3.2.2 Model-free ML

No modeling assumption.

### 3.3 Knowledge/Tasks

- Prediction: estimate output given input

given  $X \in$  feature space  $\mathcal{X}$ , predict  $Y \in$  label space  $\mathcal{Y}$

- Classification: discrete labels
- Regression: continuous labels

- Description (also called unsupervised learning)

Given  $X \in$  feature space  $\mathcal{X}$ , learn  $f(X)$

- E.g.
  - Density estimation
  - Clustering
  - Dimensionality reduction

## 4 Machine Learning Subfields

- Supervised learning
  - Data consists of both inputs and outputs
  - Tasks consists of prediction
- Semi-supervised learning
  - Data consist of inputs and only some of them with outputs
  - Tasks consists of prediction
- Reinforcement learning
  - Data consists of rewards that come through taking **actions** that has a feedback
  - Task: maximize reward.
- Multi-agent systems
  - multiple agents with same setup as in reinforcement learning.
  - the data also consists of other agents' actions.