Machine Learning

Introduction

Souhaib Ben Taieb

February 3, 2020

University of Mons

Table of contents

Contact details

- Prof. Souhaib BEN TAIEB
- De Vinci Building, second floor, room 2.15
- Email: souhaib.bentaieb@umons.ac.be

Prerequisites

- Probability and statistics
- Multivariate calculus
- Linear Algebra
- Optimization (non-linear)

Communication

- **Website**
  - `https://github.com/bsouhaib/ML20`
  - Lecture notes
- **Moodle**
  - `https://moodle.umons.ac.be/course/view.php?id=2785`
  - Forum for asking questions, etc.
  - Assignment submissions
  - **No email please — use the forum**

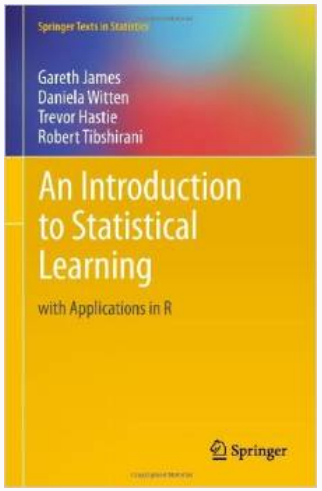
R programming language and Rstudio IDE

Key reference

James, Witten, Hastie and Tibshirani (2012) An Introduction to Statistical Learning. Springer.

www.statlearning.com

- Free pdf online
- Data sets in associated R package **ISLR**
- R code for examples



Assessment

- Final exam (*Open book*): **60%**
  - To pass, your exam mark must be $\geq \frac{50}{100}$
- One project (group of 2 students): **20%**
- Four assignments (group of 2 students): **20%** (5% each)

Task	**Due Date**	**Value**
Final exam	Official exam period	60%
Project	TBA	20%
Assignments 1–4	TBA	20%

Main topics

- Statistical learning
- Linear regression
- Classification with logistic regression and linear discriminant analysis
- Model assessment/selection and resampling methods
- Dimensionality reduction
- Advanced regression
- Advanced classification (tree-based methods)

Learning from data

- **Better understand** or **make predictions** about a certain phenomenon under study
- **Construct a model** of that phenomenon by finding relations between several variables
- If phenomenon is complex or depends on a large number of variables, an **analytical solution** might not be available
- However, we can **collect data** and learn a model that **approximates** the true underlying phenomenon

Learning from data

"Machine learning is a **scientific discipline** that explores the **construction and study of algorithms** that can **learn from data**."

- The essence of machine learning
  - A pattern exists
  - We cannot pin it down mathematically
  - We have data on it
- Learning examples
  - Spam Detection
  - Product Recommendation
  - Credit Card Fraud Detection
  - Medical Diagnosis

Related fields and other views of "learning form data"

"Statistics is the **science of learning from data**, and of **measuring**, **controlling**, and **communicating uncertainty**; [...]"

"Data mining, [...], is the **computational process** of **discovering patterns** in **large data sets** involving methods at the intersection of **artificial intelligence**, **machine learning**, **statistics**, and **database systems**."

"Data Science means the **scientific study** of the **creation**, **validation** and **transformation of data** to **create meaning**."

"Artificial Intelligence is the theory and development of **computer systems** able to perform tasks normally requiring **human intelligence**, such as **visual perception**, **speech recognition**, **decision-making**, and **translation between languages**."

Machine learning problems?

Which of the following problems are best suited for Machine Learning?

1. Classifying numbers into primes and non-primes.
2. Detecting potential fraud in credit card charges.
3. Determining the time it would take a falling object to hit the ground.
4. Determining the optimal cycle for traffic lights in a busy intersection.

- 2 and 4
- 1 and 2
- 1, 2, and 3
- 3
- 1 and 3

Different learning problems

- Supervised learning
  - (input, correct output)
- Unsupervised learning
  - (input)
- Semi-supervised learning
  - (input, correct output) for some observations, and only (input) for others.
- Reinforcement learning
  - (input, *some* output, grade for this output)
  - (state, action, reward)

Other types of learning: online learning, active learning, etc.

In practice, it is important to identify which learning problem is best suited for the application and the data available.

Different learning problems

For each of the following tasks, identify which type of learning is involved (supervised, unsupervised or reinforced) and the training data to be used. If a task can fit more that one type, explain how and describe the training data for each type.

- Recommending a book to a user in an online bookstore
- Playing tic-tac-toe
- Categorizing movies into different types
- Learning to play music