BU551W - Advances in Machine Learning in Finance
Lecture Notes

Rodrigo Miguel

January 29, 2023

Contents

Chapter 1

Lecture 1 - Introduction to Machine Learning in Finance

1.1 Big Data in Finance

1.1.1 What is Big Data?

Definition Volume (the amount of data), velocity (the speed of data processing) and variety (the number of types of data).

1.1.2 Big Data Applications

- Analyse customer satisfaction;
- Speed up manual processes with transactional data;
- Analyse financial performance;
- Customer recommendation.

1.1.3 Big Data in Finance

From a **practitioners'** perspective, big data refers to petabytes of structure and unstructured data that can be used to anticipate customer behaviour and create strategies for banks and other financial institutions.
From an **academic** perspective, big data in finance research is large sized, high dimension and with a complex structure:

- Large size: the data is large in an absolute and relative sense, e.g. transaction-level market microstructure data;
- High dimension: the data has many variables relative to its sample size;

- Complex structure: unstructured data includes text, pictures, videos, audio and voice.

1.1.4 Big Data Challenges in Finance

- **Regulatory framework:** strict regulatory requirements that need to be followed, e.g. Fundamental Review of the Trading Book, governing access to critical data and demands accelerated reporting.
- **Data privacy/security:** unauthorised data collection, hackers, etc;
- **Data quality:** ensuring information is accurate, usable and secure is a challenge;
- Cloud solutions for financial data.

1.2 The Role of Machine Learning

1.2.1 What is Machine Learning?

Machine Learning Is a field of study that allows computers the ability to learn and improve without being programmed continuously.
Traditional programming is done by a programmer that programs a computer and runs the software to get an output. To improve it, more programming needs to be done by the programmer. Machine learning reads inputs and outputs, the computer automatically processes these and outputs a program.

1.2.2 Applications of Machine Learning

- E-mail spam detection;
- Face detection and matching, e.g. Face ID;
- Self-driving cars;
- Language translation;
- Drug design;
- etc.

1.2.3 Machine Learning in the field of Finance

In 2020, a definition of machine learning was proposed:

- A diverse collection of high-dimensional models for statistical prediction;
- "Regularisation" methods for model selection and mitigation of overfit;

- Efficient algorithms for searching among a vast number of potential specifications.

1.2.4 How Machine Learning is Used in Finance

- Fraud detection and prevention: machines learn to identify and combat fraudulent financial transactions;
- Robo-advisors;
- Credit rating;
- Algorithmic trading;
- Return prediction.

1.2.5 Terminology

- **Algorithm:** set of rules that a machine follows to achieve a certain goal. Example: Cooking recipes are algorithms, the cooking steps are algorithms instructions, and ingredients are inputs. The cooked food is the output.
- A **Learner** or **Machine Learning Algorithm** is a program used to learn a machine learning model from data.
- A **Black Box Model** is a system that does not reveal its internal mechanisms, opposite of a **White Box**;
- **Interpretable Machine Learning** refers to the methods and models that make the behaviour and predictions of the machine learning systems understandable to humans;
- A **Dataset** is a table with the data from which the machine learns from;
- An **Instance/Observation** is a row in the dataset;
- The **Features** are the inputs used for prediction or classification (column in the dataset).
- A **Machine Learning Task** is the combination of a dataset with features and a target. E.g. classification, regression, survival analysis and/or clustering;
- The **Prediction** is what the machine learning model "guesses" what the target value should be, based on the features given.

Chapter 2

Lecture 2 - Financial Data Science with Python

Chapter 3

Lecture 3 - Web Scraping with Python

Chapter 4

Lecture 4 - Machine Learning Basics

Chapter 5

Lecture 5 - Supervised Learning

Chapter 6

Lecture 6 - Supervised Learning - Classifiers

Chapter 7

Lecture 7 - Unsupervised Learning and Cross-Validation

Chapter 8

Lecture 8 - Text as Data

Chapter 9

Lecture 9 - Financial Technology

Chapter 10

Lecture 10 - Fully Automated Trading System