intro2ml

October 5, 2023

1 Introduction to Machine Learning

Machine learning (ML) is at the core of this course. But before we dive into the ML algorithms, we want to be clear what exactly machine learning is. If you've heard the phrase "Machine Learning", you must have heard the more famous phrase, "Artificial Intelligence" or AI, and you've also probably heard the phase "Deep Learning". These three phrases are among the most recently popular buzzwords. Many people think they are the same, and often use the three phrases interchangeably. Well these things are related to each other, but they are not same.

1.1 Artificial Intelligence, Machine Learning and Deep Learning

A picture is worth a thousand words. The image below shows the relationship between AI, machine learning, and deep learning. Simply said, deep learning is a subset of machine learning, and machine learning is a subset of AI. Machine learning and deep learning are ways of achieving AI.

1.1.1 Artificial Intelligence

Roughly speaking, AI is a system that mimics the cognitive functions of humans, and carries out tasks based on algorithms in an "intelligent" manner. The term "Artificial intelligence" was first brought up in the 1950's. AI is a popular topic in movies. Every one loves C-3PO, and many are afraid of the Terminator. C-3PO and the Terminator are examples of so called "General AI", or "Strong AI", which refers to a system that is able to cope with any generalized task which is asked of it, much like a human. The truth is, we are not able to establish a general AI yet. All existing AI are so called Narrow AI, or Weak AI.

Narrow AI refers to AI which is able to handle particular tasks. A spam filtering tool, a face recognition system, a recommended playlist from YouTube, or even a self-driving car—all of which are sophisticated uses of technology that can be defined as "Narrow AI". To achieve AI(narrow AI for now and general AI in the future), we rely on machine learning.

1.1.2 Machine Learning

Machine Learning is a technique of parsing data, learning from that data and then applying what has been learned to make an informed decision. Machine learning focuses on designing algorithms that can learn from, and then make predictions on the data. The learning can be supervised or unsupervised.

- **Supervised Learning** is where the algorithm generates a function that maps inputs to desired outputs. One standard formulation of the supervised learning task is the Classification problem: The learner is required to learn, (to approximate the behavior of), a function which maps a vector into one of several classes by looking at several input-output examples of the function. One example of supervised learning is a spam filter: Based on past information about spams, it filters out a new incoming email into Inbox (normal) or Junk folder (Spam).
- **Unsupervised Learning** models a set of inputs without labels. Unsupervised learnings are designed to extract structure from data samples. One unsupervised learning example is Customer Segmentation, which divides customers into different groups based on certain characteristics like gender, age, and purchase pattern.

There are other types of learning like semi-supervised learning and reinforcement learning, but they are not in the scope of this course.

In this course, for supervised learning, we will introduce classification and regression; for unsupervised learning, we will introduce clustering.

Classification is used when the outcome is discrete. For example, whether a bank should approve a loan application given the applicant's financial situation is a classification problem, since the outcome is either yes or no.

Regression is used when the outcome is continuous, or a real number. For example, regression can be used to determine how much premium should be charged for an auto insurance policy given the vehicle and the driver information. It's a regression problem because the outcome is an amount, which is a real number.

Clustering is a task of grouping objects into groups, such that the objects in same group are more similar then those in other groups. Clustering can be used to unfold customer segment for marketing purposes.

We will also briefly introduce **Text Analysis** and **Time Series Analysis** in this course. Text analysis is the automated process that extracts and classifies information from text, such as emails, news, product reviews, etc. Time series analysis is a special kind of machine learning that deals with an ordered sequence of values of a variable at equally spaced time intervals. Time series analysis is widely used in business forecasting.

1.1.3 Deep Learning

Deep learning is a branch of machine learning completely based on artificial neural networks. Neural networks mimic the stucture of the human brain so deep learning also mimics how the human brain works. The main difference between deep learning and machine learning is that machine learning models become better progressively, but the models still need some guidance. In deep learning, we don't need to explicitly program everything. The concept of deep learning is not new. It has been around for a couple of years now. It's on hype nowadays because earlier we did not have that much processing power and as much data. However, in the last 20 years, as the processing powers and data increased exponentially, deep learning came into the picture.

Notable deep learning applications include:

- Self-driving car
- Image recognition
- Cancer detection
- Automatic machine translation
- Automatic text generation

- Speech recognition
- Alpha-go

In May, 2019, Google published a study on applying deep learning on lung cancer detection. In Google's case, it trained its deep learning algorithm on 45,000 chest CT scans, some of which featured cancer in various stages. The algorithm was then put to work using a single CT scan for diagnosis, with the accuracy of the algorithm compared to that of six board-certified radiologists. Google says it detected five percent more cancer cases and reduced false-positives by more than 11 percent.

While deep learning already has some impact on accounting, like text analysis on regulatory filings, or transcripts of conference calls, it still has relatively low adoptions in accounting world, partly due to the lack of interpretability of the deep learning algorithms. In this course, we will focus on other machine learning algorithms and their application in accounting field.

1.2 Machine Learning and Accounting

Among all industries, financial services are early leading adopters of machine learning, together with high tech and telecom. The majority of financial services are delivered by cross-functional teams that use machine learning.

According to PWC, future auditors will be able to audit 100% of companies' financial transactions with the help of machine learning. Machine learning algorithms will process and review the data, recognize anomalies and compile a list of outliers for auditors to check. Instead of spending most of their time checking data, auditors can apply their skills to investigating and deducing the reason behind a pattern or anomaly.

Machine learning also has big impact on management accounting. It can be applied in many parts of the organization to ensure proper governance and internal controls throughout the organization.

Accounting would be benefited through machine learning in many ways, including but not limited to:

- Auditing of expense submission
- Reducing the documentation burden
- Risk assessment
- Automated invoice categorization
- Business forecasting
- Decision making in management activities

In this course, we will introduce classification, regression, clustering, text analysis and time series analysis which have various applications in accounting industry.

- Classification
- Auditing
- Risk assessment
- Decision making
- Regression
- Price prediction
- Operating efficiency improvement
- Text Analysis
- Sentiment analysis

- Content recommendation
- Clustering
- Customer segmentation
- Time series analysis
- Business forecasting

1.3 Ancillary Information

The following links are to additional documentation that you might find helpful in learning this material. Reading these web-accessible documents is completely optional.

1. Why Accountants Must Embrace Machine Learning

2.

1.4 An overview of machine learning