Class

Introduction to Machine Learning

Topic

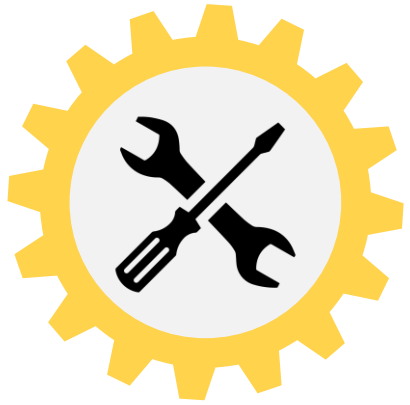
Overview

Agenda

Relationship between Artificial Intelligence and Machine Learning

Machine Learning and commonly used Machine Learning algorithms

Machine Learning course structure



Premises of using Machine Learning

Differences between Machine Learning and Statistical Modelling

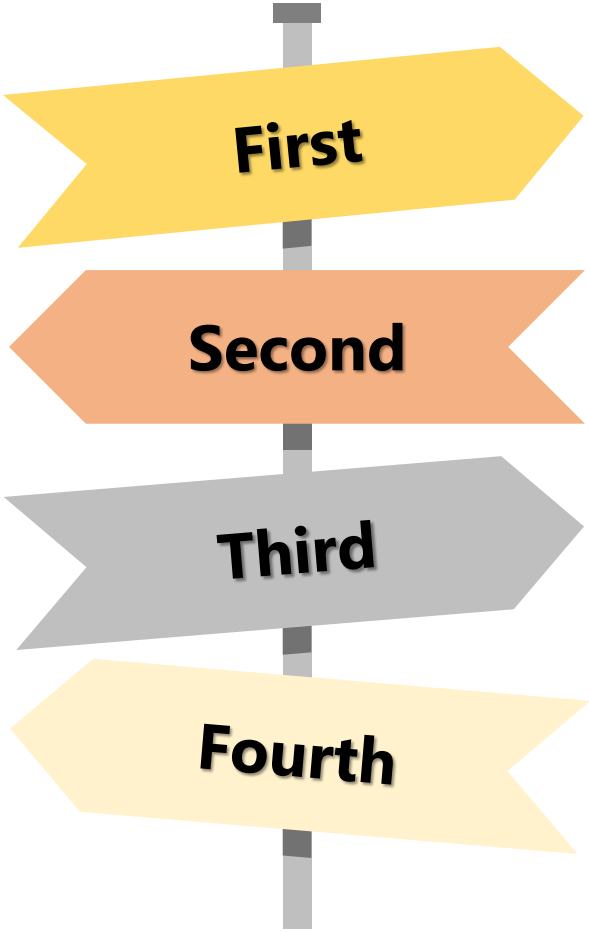
Genesis of Machine Learning

1950'S - Artificial intelligence

- Sub-discipline of CS
- Develop computer programmes that could think and evolve

Machine learning

- Sub-field of AI
- Programs could be created that learn from data



Computer science

Chess and Sudoku – computer programmed games

- In a traditional computer program explicit rules were coded
- Computers couldn't differentiate between picture and sounds
- And, creating a program to do so using an explicit set of rules was not practical

Illustrating Use of Machine Learning

As an email service provider how to filter out any spam messages to give a better user experience?

Words such as lottery, win, million dollars etc in emails are used to flag emails as spam

Write a program to check each email from some specific email ids that historically sends spam messages

Scan for some keywords in the email body to check whether it's a spam

Problems with the approach

Tedious to implement

Prone to errors

Not always easy to manually generate a list of email ids to be black listed or gather keywords to be searched

Illustrating Use of Machine Learning

As an email service provider how to filter out any spam messages to give a better user experience?

Rules are fed into a traditional program to decide whether an email is spam

Historical Data	
Email 1	S
Email 2	S
Email 3	NS
Email 4	S
Email N	NS

Human

Rules

Machine Learning

As an email service provider how to filter out any spam messages to give a better user experience?

Historical Data	
Email 1	S
Email 2	S
Email 3	NS
Email 4	S
Email N	NS

An alternate way to solve this problem

Human

Rules

Machine Learning

As an email service provider how to filter out any spam messages to give a better user experience?

An alternate way to solve this problem

Historical Data	
Email 1	S
Email 2	S
Email 3	NS
Email 4	S
Email N	NS

Algorithm

Rules

Machine Learning

"Machine learning is a field of computer science, that gives computers the ability to learn with data without being explicitly programmed"
- Samuel, Arthur (1959)

Application of Machine Learning

Automated cars – calculating the forthcoming traffic sign to control motion accordingly

Banks - making credit decisions

Companies – choosing the best marketing channels

Online experience - filtering out spam messages, articles recommended based on reading preferences etc.

Machine Learning Tasks

Regression: Predicting the value of continuous variable based on how a variable is related to some other variables

Example –

a. An e-retailer like Amazon or Flipkart can predict how much money a customer will spend in next 1 month based on his/her purchase history and user interaction on their mobile apps

Classification: Predicting the class of a given data point, given certain attributes

Example –

a. Based on the credit history, a bank can predict if a given person will pay his full due
b. An expert system controlling a self-driving car while determining the speed limit by following traffic signs

System will look at the picture of traffic signs to classify it correctly and take action

Unsupervised Learning: Finding similar data points in a dataset

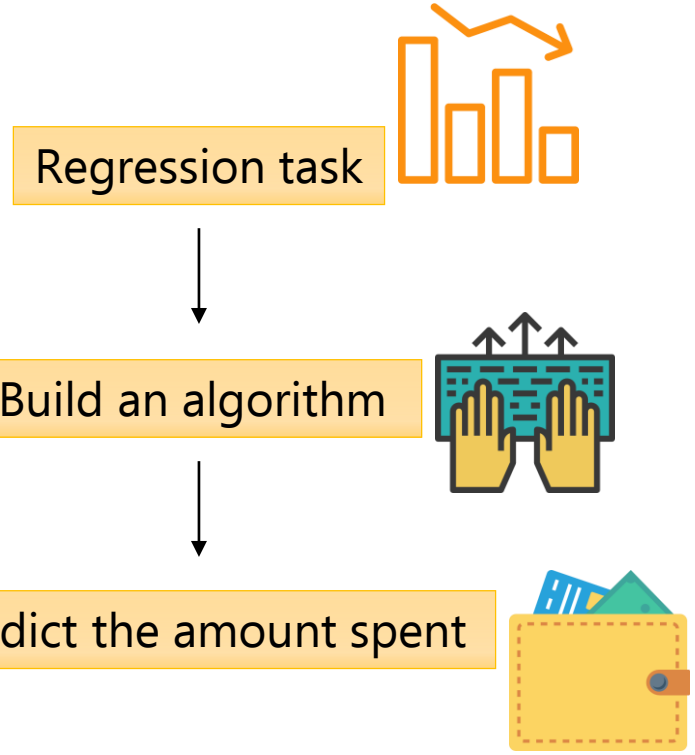
Example –

a. Grouping segments of customers who are similar on certain sets of attributes like demography, buying behaviour etc.
b. Building recommendation engines, which are also based on the notion of similarity

Regression Task

Historical data

Gender	Income	Age	Amount Spent
Male	40000	30	1000
Female	35000	26	500
Female	50000	32	2500
Male	50000	40	5000
Female	65000	35	5000



Algorithm will learn to predict Amount Spent of a new customer with the available information about their gender, age and income

Classification Task

Historical data

Gender	Income	Age	Good
Male	40000	30	Yes
Female	35000	26	No
Female	50000	32	Yes
Male	50000	40	No
Female	65000	35	No

Classification

Bank

Train an algorithm

Classify its current customers into good or bad customers

Help the bank in future to predict if a customer will be a good or not given his demographic data such as gender, income and age

Classification Task

Image labels and corresponding signs

Image	Sign
Image 1	Stop
Image 2	U Turn
…	…
…	…
Image N	Parking

Classification

Expert system recognizing images of different traffic signs

Image labels

Train an algorithm

Recognize the sign corresponding to the image label

Supervised Machine Learning

Classification Task

Gender	Income	Age	Good
Male	40000	30	Yes
Female	35000	26	No
Female	50000	32	Yes
Male	50000	40	No
Female	65000	35	No

Image	Sign
Image 1	Stop
Image 2	U Turn
...	...
...	...
Image N	Parking

Regression Task

Gender	Income	Age	Amount Spent
Male	40000	30	1000
Female	35000	26	500
Female	50000	32	2500
Male	50000	40	5000
Female	65000	35	5000

Contained a column that was needed to be predicted

Supervised Machine Learning
Whenever an algorithm is trained in a manner where the variable required by the algorithm to predict is present in the training data

Unsupervised Machine Learning

Recommendation engine

Customer	Item 1	Item 2	Item 3	Item 4
C1	Yes	Yes	No	No
C2	No	No	Yes	Yes
C3	Yes	?	No	No
C4	No	No	?	Yes

User behaviour with respect to product offerings

C3 and **C4** - not bought some items and their preferences are not known

C1 and **C3** – similar in terms of product preferences

Recommend **Item 2** to **C3**

C3 will like **Item 2**

C1 - liked **Item 2**

Unsupervised Machine Learning

Recommendation engine

Customer	Item 1	Item 2	Item 3	Item 4
C1	Yes	Yes	No	No
C2	No	No	Yes	Yes
C3	Yes	?	No	No
C4	No	No	?	Yes

C3 and **C4** - not bought some items and their preferences are not known

C2 and **C4** – similar in terms of product preferences

Recommend **Item 3** to **C4**

C4 will like **Item 3**

C2 - liked **Item 3**

User behaviour with respect to our product offerings

Unsupervised Machine Learning
When no target variable is present in the dataset of the algorithm

Recap

- Genesis of Machine Learning
- Illustrating Use of Machine Learning
- Machine Learning
- Machine Learning tasks
- Regression Task
- Classification Task
- Supervised Machine Learning
- Unsupervised Machine Learning

Class

Introduction to Machine Learning


Topic

Common Algorithms and Course Overview

Commonly Used Algorithms

Classification

- Logistic Regression
- Decision Tree
- Random Forest
- Gradient Boosted Trees
- SVM
- Naïve Bayes
- Neural Networks
- Convolutional Neural Networks

Deep Learning: sub-field of Machine Learning

Regression

- Linear Regression
- Regression Tree
- Random Forest
- Gradient Boosted Trees
- SVM
- Neural Networks

Unsupervised

- K-Means
- Agglomerative Clustering
- SVD
- NMF

Recommendation Engines

Commonly Used Algorithms

Classification

- Logistic Regression
- Decision Tree
- Random Forest
- Gradient Boosted Trees
- SVM
- Naïve Bayes
- Neural Networks
- Convolutional Neural Networks

Structured Data

Unstructured Data

Images, text and video feeds

Commonly Used Algorithms

Classification

- Logistic Regression
- Decision Tree
- Random Forest
- Gradient Boosted Trees
- SVM
- Naïve Bayes
- Neural Networks
- Convolutional Neural Networks

Regression

- Linear Regression
- Regression Tree
- Random Forest
- Gradient Boosted Trees
- SVM
- Neural Networks

Unsupervised

Structured Data

Unstructured Data

Course Overview

Python – data analysis

- Working with structured data
- Creating basic visualizations
- Feature engineering
- Data exploration

Pandas package in python - handle structured data

Matplotlib and **Seaborn** – visualize data

Machine Learning Algorithms

- Linear models – Linear and Logistic
- Tree based models – Random Forest and Gradient Boosted Machines
- Neural Networks and Convolutional Neural Network
- K-means and Agglomerative Clustering
- Recommendation Engines – SVD and NMF

Scikitlearn library - perform most of the ML tasks

Keras with TensorFlow backend – build Neural Networks and Convolutional Neural networks

OpenCV – handle and process image data

When Should Machine Learning Be Used

When can Machine Learning Algorithms be used?

ML is not always used

Limited number of scenarios for using MLA

Through simple data analysis, it is possible to come up with insights or rules that can be fed into a traditional program

When Should Machine Learning Be Used

When can Machine Learning Algorithms be used?

While implementing a cyber security solution based on Machine Learning, petabytes of log files need to be handled

In absence of proper hardware and software stack, the task of training machine learning algorithms becomes difficult

Framework to use ML

1. Traditional programming isn't capable of solving the task

2. Availability of relevant historical data to train the algorithms

3. Ease of handling and processing the available data

When Should Machine Learning Be Used

When can Machine Learning Algorithms be used?

Traditional programming - deterministic solution

Machine Learning - experiment with different approaches to arrive at a reasonable solution

Framework to use ML

1. Traditional programming isn't capable of solving the task
2. Availability of relevant historical data to train the algorithms
3. Ease of handling and processing the available data
4. Enough time to iterate through different Machine Learning Algorithms

Unlike backend frame work, where there are set design patterns, in ML things are in more flux and usually **implementing an ML solution may mean iterating with several approaches over a substantial period of time**

Machine Learning Versus Traditional Modelling

Differences between modern Machine Learning and traditional statistical modelling

This section is meant for -

People who have built statistical model before and want to understand its difference with Machine Learning

This section is not meant for -

People new to model building

Return to these videos later in future after acquiring some perspective on modelling in general

Machine Learning Versus Traditional Modelling

Data science is fairly dated

Companies have been using models to keep the businesses better informed

Early adopters

- Banking 

- Insurance 
- Retail companies 

Using data analysis for management to make informed decisions

Traditional Modelling

- **Model building** - statistical methods 
- **Data** – structured in nature 

Machine Learning Versus Traditional Modelling

Reasons for popularity of statistical modelling with early adopters:

Structured Data – primarily analysed structured data stored in warehouse as SQL tables

Data with contextual business meaning - each column in SQL tables was mostly related to some business KPI

Provide model interpretation – required by companies while explaining the predictions of their models

Banks are bound by regulatory frameworks to explain their credit decisions

Machine Learning Versus Traditional Modelling

	Traditional statistical modelling	Modern machine learning
Origin	Statistical reasoning	Computer Science and Information Theory
Consumer	Banks, insurance company etc.	Big technology companies
Model	Emphasis on interpretation more than accuracy	Emphasis on accuracy more than interpretation
Data	Structured	Unstructured - text, images, audio files and video feeds
Aim	Gathering information for the decision makers in an organization	Create data-based products - digital assistants, self-driving cars, chatbots, spam filters etc

Machine Learning Versus Traditional Modelling

Traditional statistical modelling

Modern machine learning

Similarity

Differences

Help in creating models that learn from the data

End goals and data used for training

Despite the popularity of Machine Learning, traditional modelling is still prevalent

Many businesses need to only analyse data in a traditional manner

The company will have the final say on the method of data analysis

Recap

- Commonly Used Algorithms
- Course Overview
- When Should Machine Learning Be Used
- Machine Learning Versus Traditional Modelling