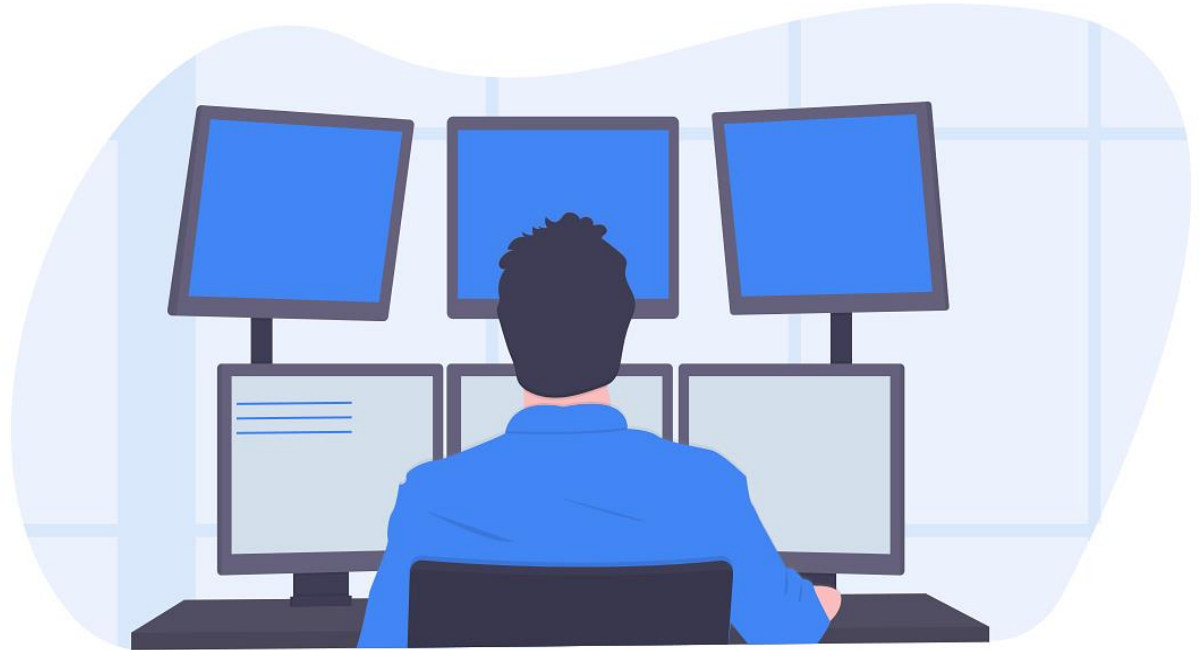


# Introduction to Machine Learning

Eng Teong Cheah



# Agenda

What is machine learning

Machine learning algorithms

Machine learning languages

# What is machine learning



# What is machine learning

Machine learning is a branch of science that deals with programming the systems in such a way that they automatically learn and improve with experience.

Here, learning means recognizing and understanding the input data and making wise decisions based on the supplied data.

# What is machine learning

It is very difficult to cater to all the decisions based on all possible inputs. To tackle this problem, algorithms are developed.

These algorithms build knowledge from specific data and past experience with the principles of statistics, probability theory, logic, combinatorial optimization, search, reinforcement learning, and control theory.

# Machine learning concepts and methodologies

When you start with machine learning, there are several concepts to consider:

The key steps in using and building machine learning models are:

# Machine learning concepts and methodologies

## 1. Obtain raw data

Data can be in a structured form such as a SQL database, or unstructured, such as email message text.

## 2. Preprocess the data

Typically, the source data will need to be prepared in some way, before it can be used – such as to extract just the features you're interested in, or to remove unnecessary noise from the data.

# Machine learning concepts and methodologies

## 3. Prepare the data.

Before you can use the data within your machine learning environment, you might to add labels to the data, change numerical formats, and so on.

## 4. Apply one or more machine learning algorithms to the data.

Typically, you will need to apply several algorithms, and repeatedly test your models, to identify the best algorithms to use.



# Machine learning concepts and methodologies

## 5. Determine the best model to use.

From your candidate models, you will need to apply real-world or business logic to help determine the best model to deploy and use.

## 6. Deploy the model.

You deploy a machine learning model so that it can be used in an application, and by end users; at this stage, you have program code as in any application.

# Machine learning algorithms



# Machine learning algorithms

Machine learning algorithms can be categorized based on the type of data and the typical questions that they are used to try to answer:

## Classification algorithms

used for yes/no questions, or to identify the most likely outcome from a multiclass list where there are more than two possibilities.

Classification algorithms will create a model that processes a dataset, and identify which of two or more predefined categories that data represents.

# Machine learning algorithms

## Regression algorithms

typically used to make predictions of outcomes, based on historical patterns. Regression algorithms will create a model that uses historical data, and then identifies relationships between values in that data to make accurate predictions for the future.

## Clustering algorithms

Used to examine large input datasets, and identify clusters or groupings within that data. You use clustering a model that can take a new piece of data and assign or allocate that data to the appropriate cluster for further action or processing.

# Machine learning algorithms

What machine learning algorithm should I use?

**It depends.**

It depends on the size, quality, and nature of data.

It depends on what you want to do with the answer.

It depends on how the math of the algorithm was translated into instructions for the computer you are using.

It depends on how much time you have.

# Classification algorithms

Two-classification algorithms:

- Average Perception
- Bayes Point Machine
- Boosted Decision Tree
- Decision Forest
- Decision Jungle
- Locally Deep Support Vector Machine
- Logistic Regression
- Neural Network
- Support Vector Machine

Multiclass classification algorithms:

- Decision Forest
- Decision Jungle
- Logistic Regression
- Neural Network
- One-vs-All Multiclass

# Regression algorithms

- Bayesian Linear Regression
- Boosted Decision Tree Regression
- Decision Forest Regression
- Fast Forest Quantile Regression
- Linear Regression
- Neural Network Regression
- Ordinal Regression
- Poisson Regression

# Clustering

- Often used during the initial stages of mode development
- Detects patterns and anomalies

Examples:

- K-Means Clustering.



# Supervised Learning

Supervised learning deals with learning a function from available training data.

A supervised learning algorithm analyzes the training data and produces an inferred function, which can use for mapping new examples.

# Unsupervised Learning

Unsupervised learning makes sense of unlabeled data without having any predefined dataset for its training.

Unsupervised learning is an extremely powerful tool for analyzing available data and look for patterns and trends.

It is commonly used for clustering similar input into logical groups.

# Anomaly detection

- Rare events
- Imbalanced data

Anomaly detection methods:

- Support Vector Machine (SVM)
- PCA-Based Anomaly Detection

# Machine learning languages



# R in machine learning

R is a workhorse for statistical analysis and by extension machine learning.

It is the platform to use to understand and explore your data using statistical methods and graphs.

It has an enormous number of machine learning algorithms, and advanced implementations too written by the developers of the algorithm.

# Python in machine learning

Python is a popular scientific language and a rising star for machine learning.

I'd be surprised if it can take the data analysis mantle from R, but matrix handling in NumPy may challenge MAATLAB and communication tools like IPython are very attractive and a step into the future of reproducibility.

# Python in machine learning

Python is a popular scientific language and a rising star for machine learning.

I'd be surprised if it can take the data analysis mantle from R, but matrix handling in NumPy may challenge MAATLAB and communication tools like IPython are very attractive and a step into the future of reproducibility.

# Demo

Quantile Regression: Car price prediction





# Resources

TutorialPoints

Microsoft Docs

[Lecture Collection | Convolutional Neural Networks for Visual Recognition \(Spring 2017\)](#)

[Python Numpy Tutorial](#)

[Quantile Regression: Car price prediction](#)

Thank you

