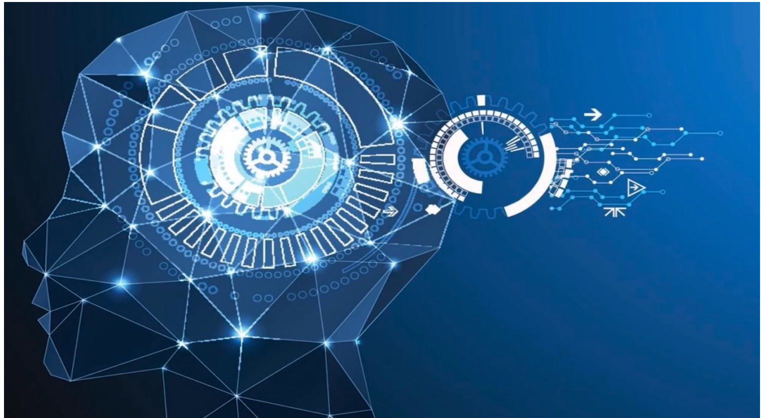
# Statistical and machine learning methods for engineering mechanics

Wengang Mao (Professor at Marine Technology)
Department of Mechanics and Maritime Sciences,
Chalmers University of Technology,
Goteborg, Sweden

# Outline of this lecture

- Course objectives
- Teaching format, examination and course schedule
- **Definition of various terminologies (AI, ML, Data Science, Deep learning, …)**
- **Overview of Machine learning techniques**
- Course contents on ML methods
- Other matters



# Course objectives

- Teach theories of statistical and machine learning methods
- Deal with two types of data, i.e., independent observations from multi-variables with uncertainties/noises, and strongly dependent observations of time series signals.
- Some practical projects to show how to use the knowledge in this course.
- Since this is a PhD course, it will rather run in a little fast pace. We hope that you should actively seek for literature to understand some concept by yourself. Or I can give your reference for further reading.

# Teaching format and examination

TEACHING, LANGUAGE AND EXAMINATION

LECTURE HOURS AND ROOMS

DISCUSSIONS

## Teaching Format

- Mainly composed of lectures (80%): theory + examples
- Project assignments (20% as homework)
- Small exercise when request

## Program/language for ML algorithms

- Python: sci-kit + several ML libraries
- R: statistical learning packages
- Others are also encouraged: Tensorflow, Pytorch, Keras

## Examination and lectures

- Lectures mainly in the campus (2-4 hours with flexible times and locations)
- Examination is the presentation for the project assignment and seminars

# Course schedule and examination

## Project assignment (homework)

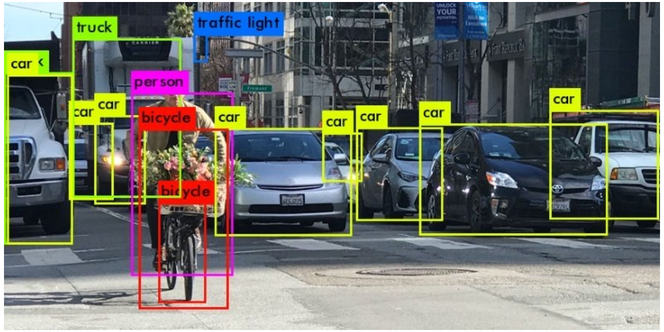
- 3 projects assignments will be handed out
- You choose 2 of them to finish.
- Replace one with your own project.

| Type of class | Date | Time | Room | Teachers | Contents |
|---|---|---|---|---|---|
| **Week 12, 2021: study week 1** | | | | | |
| **Lecture 1** | Wed 24/3 | 14.15-16.00 | Zoom | **WM** | Course contents: Introduction to the course, learning objectives, and criteria for examination |
| **Week 13, 2021: study week 2** | | | | | |
| **Lecture 2** | Mon 29/3 | 10.15-12.00 | Zoom | **WM** | Regression and statistical interpretation |
| **Lecture 3** | Wed 31/3 | 14.15-16.00 | Zoom | **WM** | Polynomial and Spline regression |
| **Week 14, 2021: study week 3** | | | | | |
| **Lecture 4** | Wed 7/4 | 14.15-16.00 | Zoom | **WM** | Model parameter estimation - gradient |
| **Week 15, 2021: study week 4** | | | | | |
| **Lecture 5** | Mon 12/4 | 10.15-12.00 | Zoom | **WM** | Generalized Linear Models and Additive models |
| **Lecture 6** | Wed 14/4 | 14.15-16.00 | Zoom | **WM** | Logistical regression and classification |
| **Week 16, 2021: study week 5** | | | | | |
| **Lecture 7** | Mon 19/4 | 10.15-12.00 | Zoom | **WM** | ML algorithms 1- neural network |
| **Lecture 8** | Wed 21/4 | 14.15-16.00 | Zoom | **WM** | ML algorithms 2 - support vector machines |
| **Week 17, 2021: study week 6** | | | | | |
| **Lecture 9** | Mon 26/4 | 10.15-12.00 | Zoom | **WM** | ML algorithms 3 - Decision trees and boost methods |
| **Lecture 10** | Wed 28/4 | 14.15-16.00 | Zoom | **WM** | ML algorithms 4 – XGBoost for regression and examples |
| **Week 18, 2021: study week 7** | | | | | |
| **Lecture 11** | Mon 3/5 | 10.15-12.00 | Zoom | **Anders** | Time Series 1 - Transformation and Gaussian process |
| **Lecture 12** | Wed 5/5 | 14.15-16.00 | Zoom | **WM** | Time Series 2 – Basic properties of random process |
| **Week 19, 2021: study week 8** | | | | | |
| **Lecture 13** | Mon 10/5 | 10.15-12.00 | Zoom | **WM** | Time Series 3 - Time series analysis and model exploration |
| **Lecture 14** | Wed 12/5 | 14.15-16.00 | Zoom | **WM** | Time Series 4 – Autoregressive integrated Moving Average model (1) |
| **Week 20, 2021: study week 8** | | | | | |
| **Lecture 15** | Mon 17/5 | 10.15-12.00 | Zoom | **WM** | Time Series 5 - Autoregressive integrated Moving Average model (2) |
| **Lecture 16** | Wed 19/5 | 10.15-12.00 | Zoom | **WM** | Time Series 6 – A few examples of using ARIMA model applications |

# Definition of various terminologies (AI, ML, Data Science, Deep learning, …)

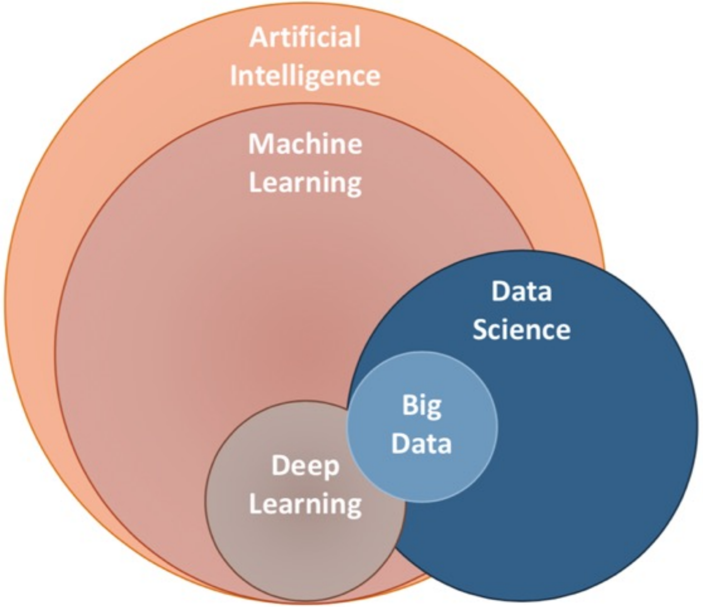
# Artificial Intelligence (AI)

- AI is a field of development of intelligent machines that work and react like humans
  - See like a human: image processing;
  - Listening like a human: speech recognition
  - Feel like human: Sensing, data collection
  - Act like a human: autonomous driving, custom support
  - Adapt like a human: decision make, obstacle avoidance





# AI, ML, Deeplearning, Data Science …



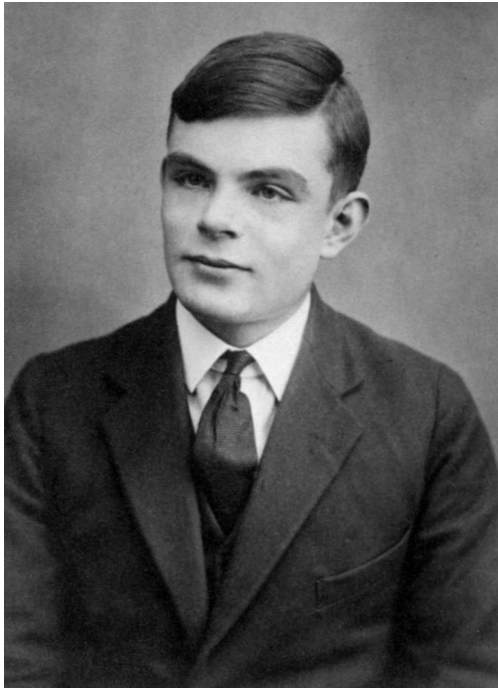

- **AI** is a program that can sense, reason (model), act and adapt for certain applications. The process is normally done by a machine or a non-living artificial.
- **ML** refers to algorithms/methods that can establish models to describe performance, scenarios, systems from data
- **Deep Learning**, a subset or advancement of ML (neural network models), is used when ML cannot fully deliver desired outcomes, e.g., dataset/features too large.
- **Data Science** is about data, a multidisciplinary field focused on drawing INSIGHTS that can help us make better decisions. It is the basis for the AI, ML, Deep Learning, etc.

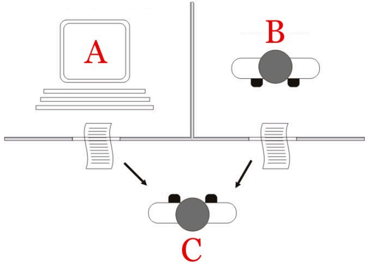
# AI (history)

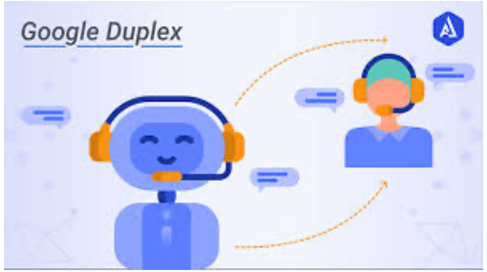
## The Turing Test (1950s)

- If a machine exhibits intelligent behavior equivalent to, or indistinguishable from, that of a human.
  - For example, a three-person game "imitation game“: Can the evaluator “C” distinguish the human and the machine.
- Google AI Duplex: who is talking to you?

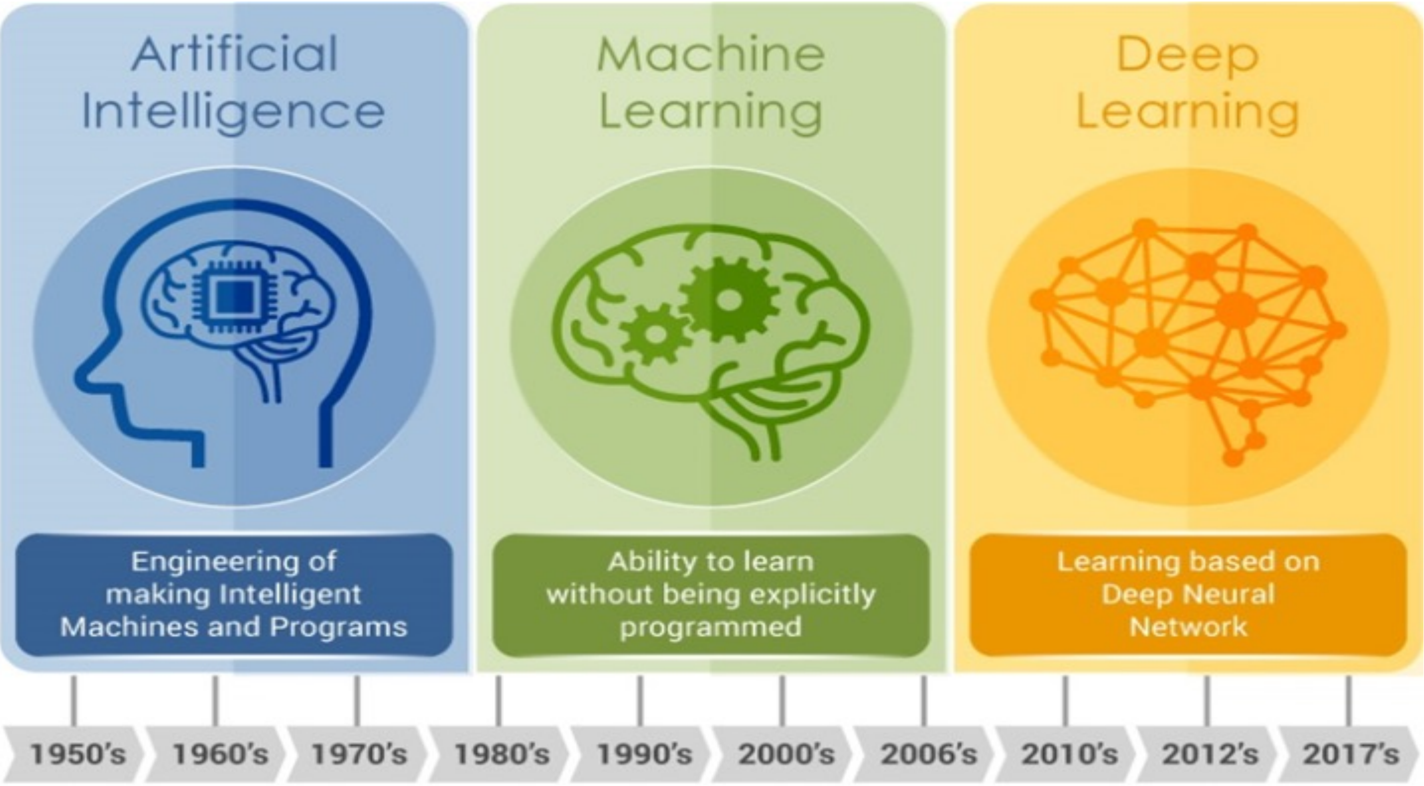
Alan Turing





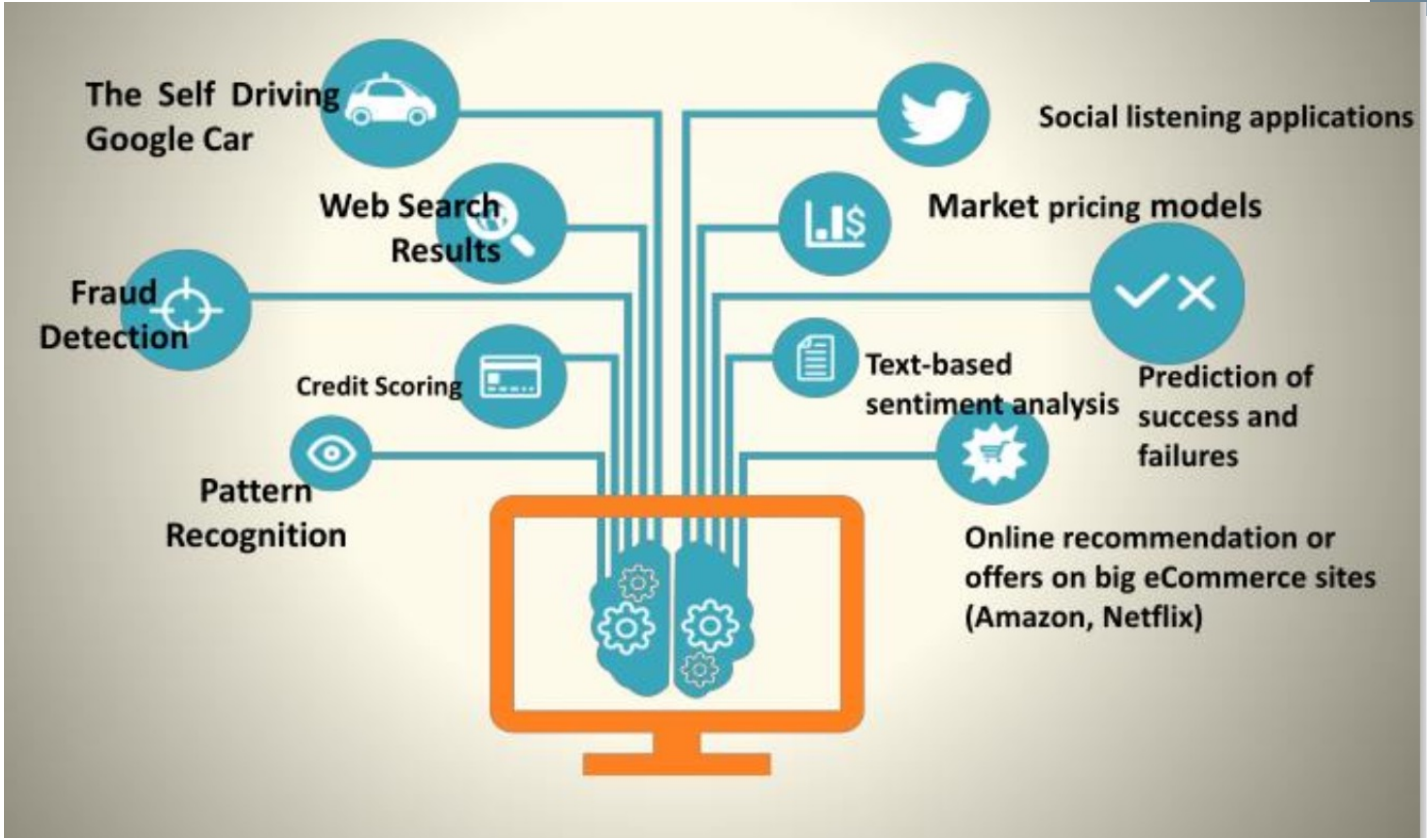



# AI, ML, Deep Learning (history)

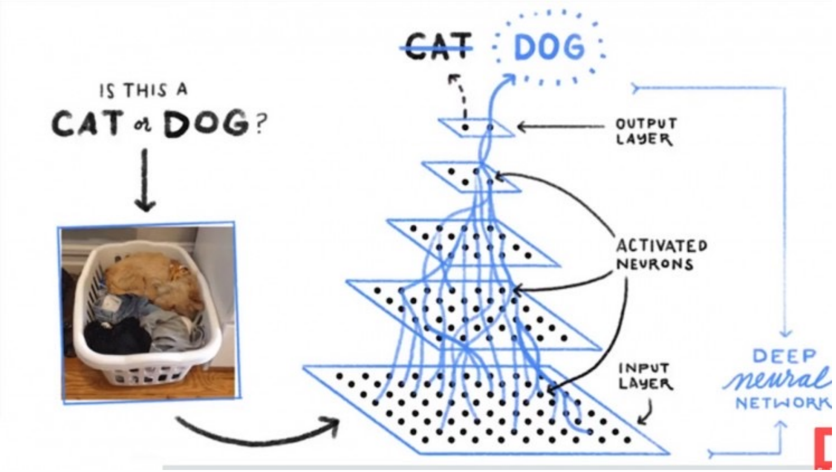


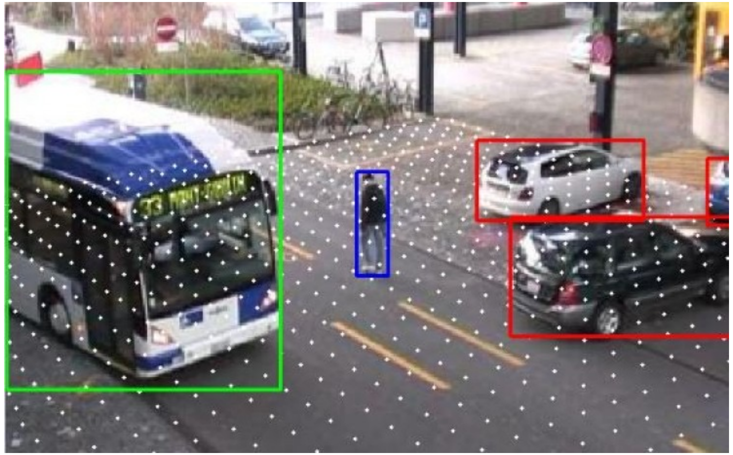
# Overview of Machine learning techniques

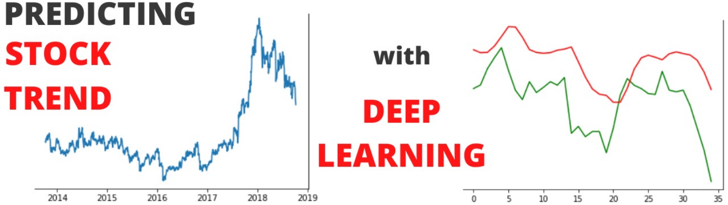
# Examples of Machine learning (1)

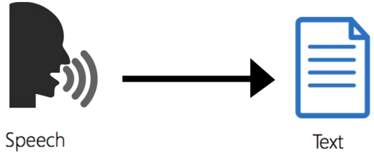


# Examples of Machine learning (2)

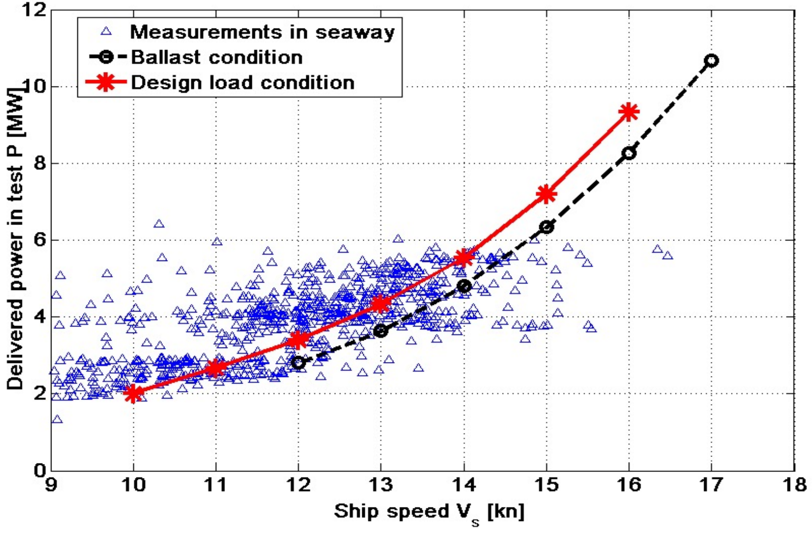


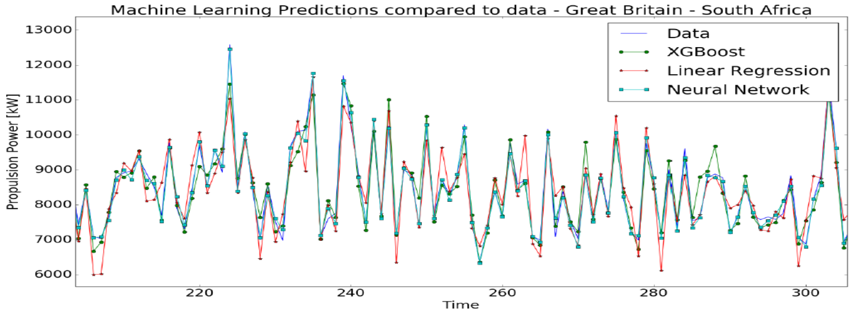



# Examples of Machine learning (3)



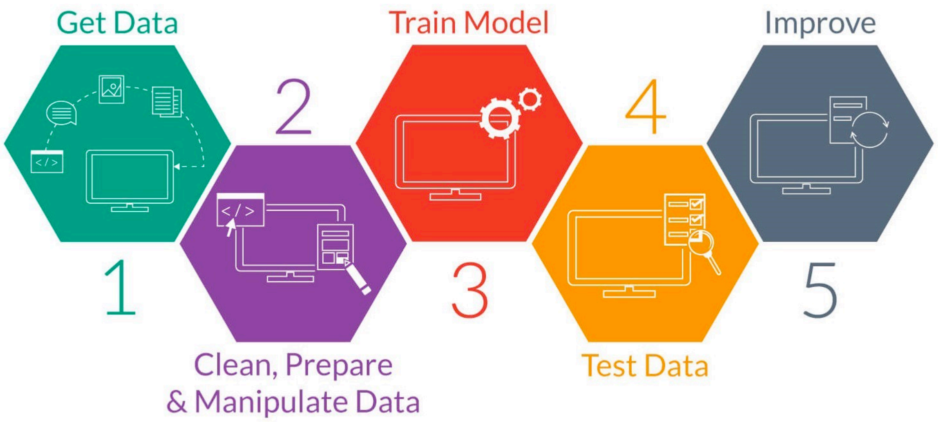




| | Time | MAE Propulsion | R2 |
|---|---|---|---|
| Linear Regression | 0.01 s | 460 kW | 0.75 |
| XGBoost | 18.67 s | 147 kW | 0.90 |
| Neural Network | 3.15 s | 276 kW | 0.97 |

**Based on mechanics data, ML can help us build models to predict performance and estimate responses**

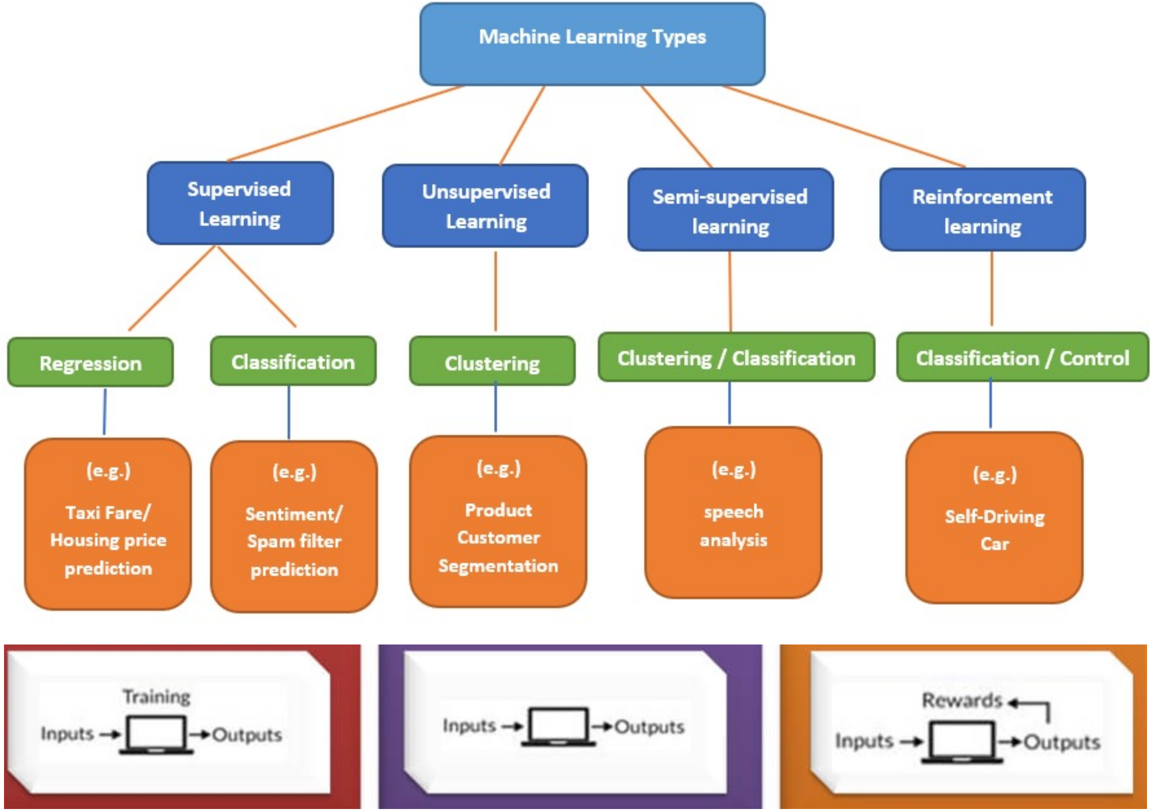
# Machine learning definition

- Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed (Arthur Samuel 1959).
- Well-posed Learning Problem: A computer program is said to *learn* from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E (Tom Mitchell 1998).




- Data types and randomness
  - Data/observations are often random variables
  - Categorical, logical or **numerical** data
  - Correlated or independent
  - Retrain your model (with train/test splits) you want to generate a new prediction.
  - The uncertainty of the forecast is just as important as, or even more so, than the forecast.

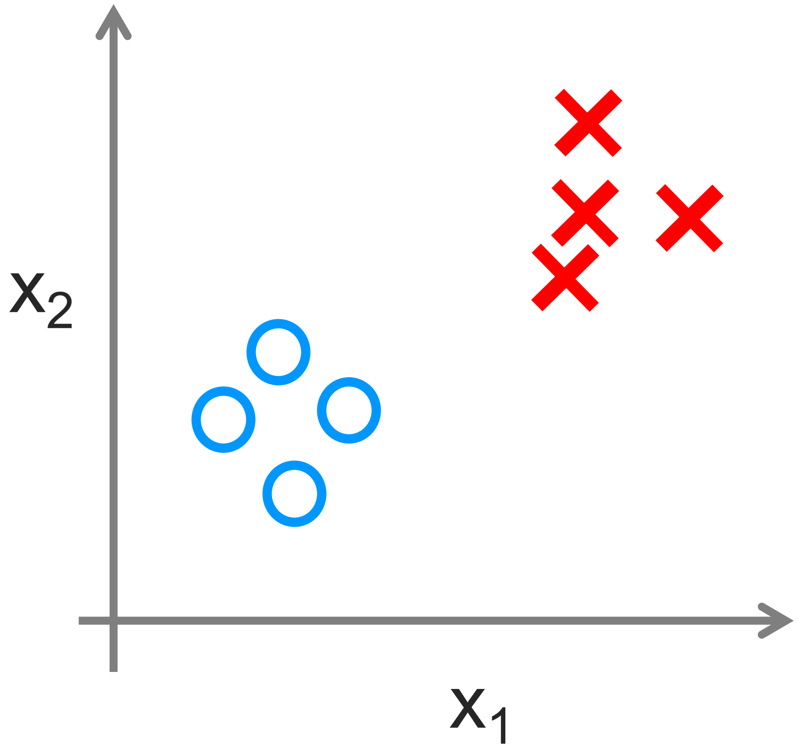
# Types of machine learning

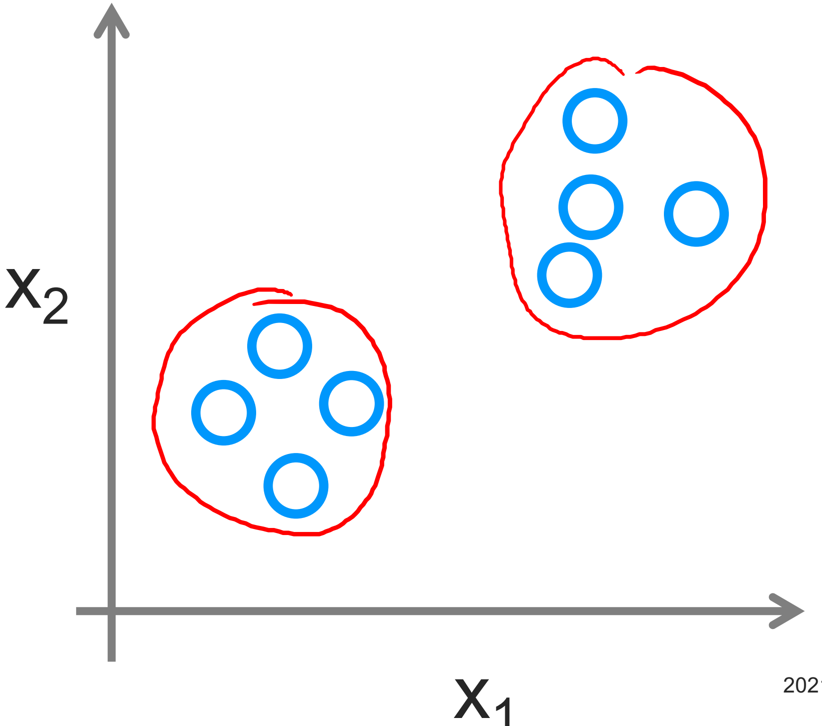



**The ML types depends on:**

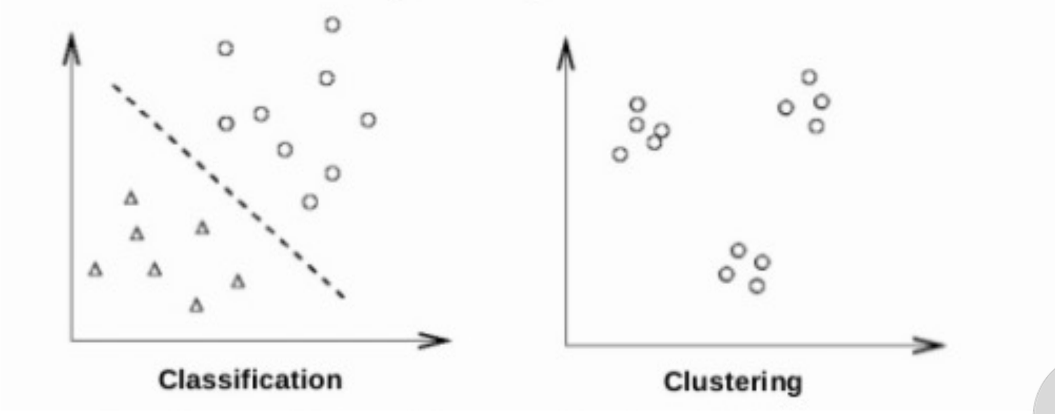
- Learning task of the ML methods
- Discrete the task into mathematical notations
- Information → data
- Input data types (logistic data, categorical data,
- Required output/prediction

# Supervised VS unsupervised ML

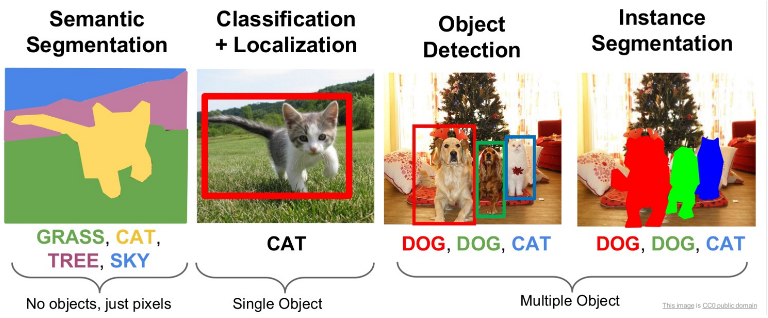


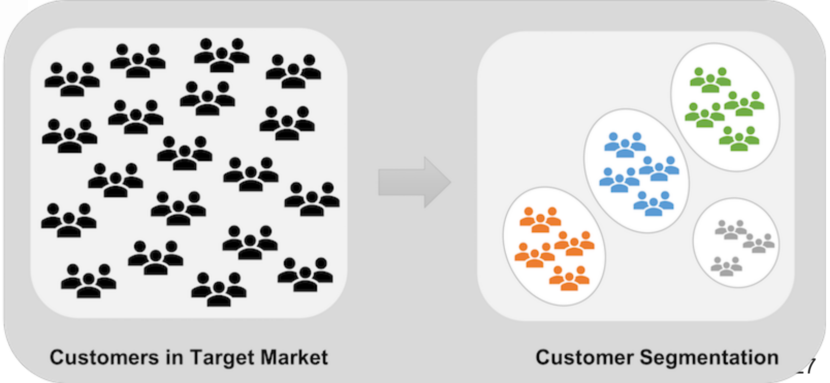



# Classification VS clustering

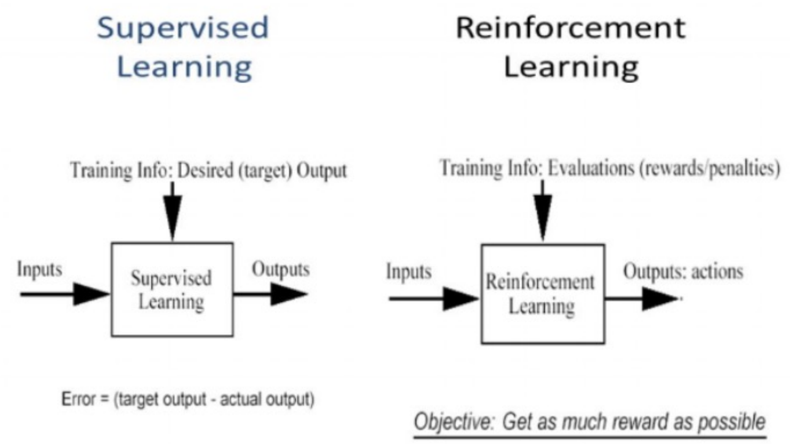


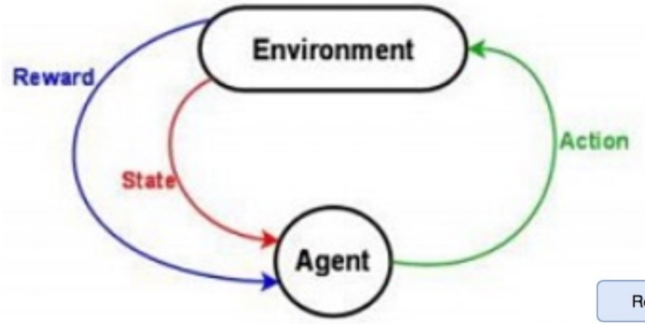

| Supervised Classification | Unsupervised Clustering |
|---|---|
| • known number of classes<br>• based on a training set<br>• used to classify future observations | • unknown number of classes<br>• no prior knowledge<br>• used to understand (explore) data |

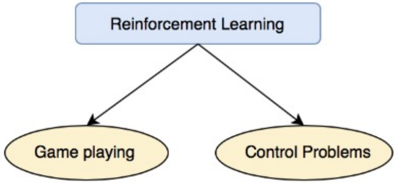


# Re-inforcement learning

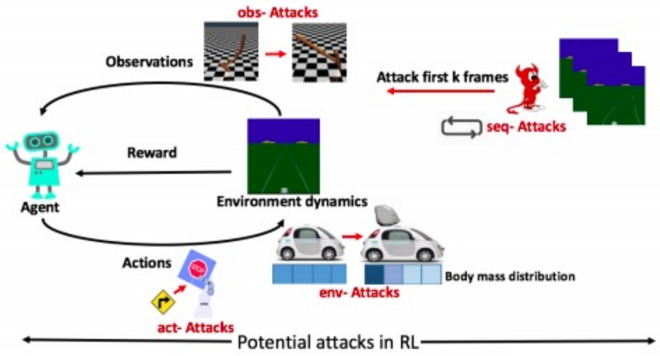



**The key features of reinforcement learning:**

- RL finds a compromise through trial and error between exploring new environment and the application of existing knowledge.
- RL explicitly considers the whole problem of a goal-directed agent interacting with an uncertain environment.
- All RL agents have explicit goals, can sense aspects of their environments, and can choose actions to influence their environments.
- The agent should operate in significant uncertain environment.



# ML – supervised machine learning

**In particular, we will focus on**

- Supervised machine learning with clear learning outcome/objectives as numbers
- Formulate the learning problem by mathematical models includes:
  - ✓ Input parameters and data
  - ✓ Possible formulas, relationship either as explicit or black box
  - ✓ Clear output variables and values
- Data be of the numerical values (forces, coefficients in a mechanics system)
- Data (values) can be collected independently or in time series format

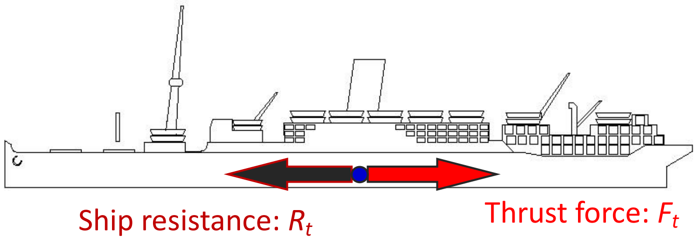
# Two supervised ML problems

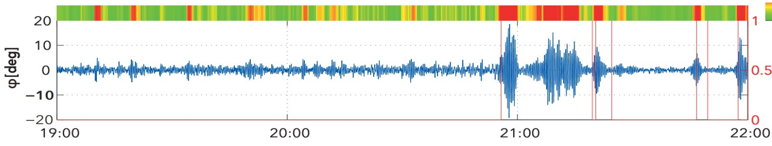
Establish a **model** to **predict** of an engineering mechanics model

- <u>Supervised Learning:</u> “right answers” given (or measured)
- <u>ML Regression:</u> to establish a model
- <u>Prediction:</u> use the model and new input to get new output

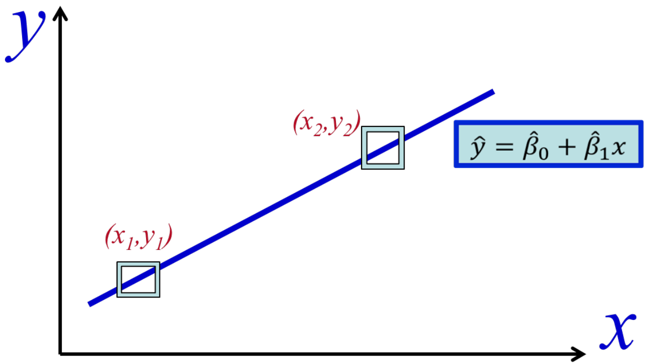
**Two examples of supervised ML methods:**

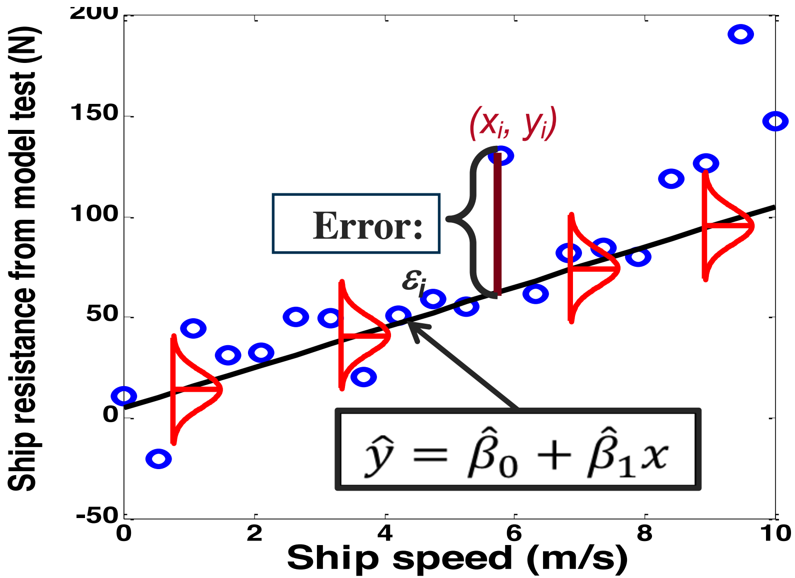
- Predict a ship speed-power model
- Predict a ship’s motion (parametric rolling)



# Exmample 1: ML for power prediction

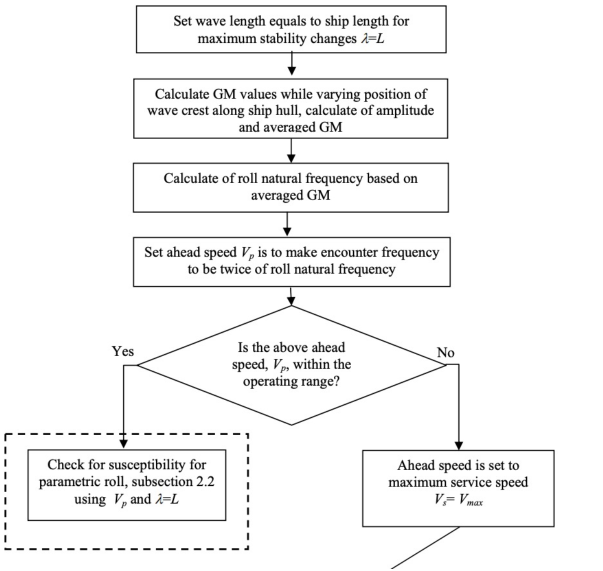



- X: ship speed
- Y: ship resistance/power measured
- Both variable may contain errors (random variables)

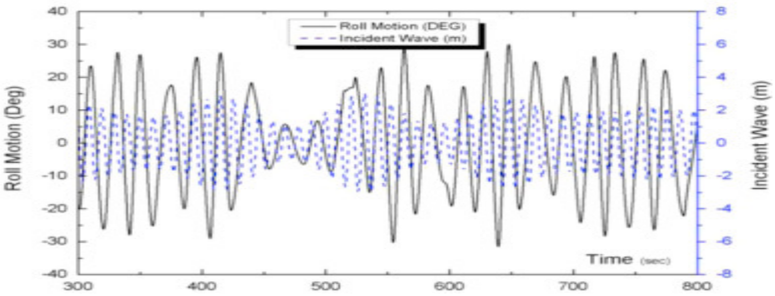
# Example 2: ML for parametric rolling forecast

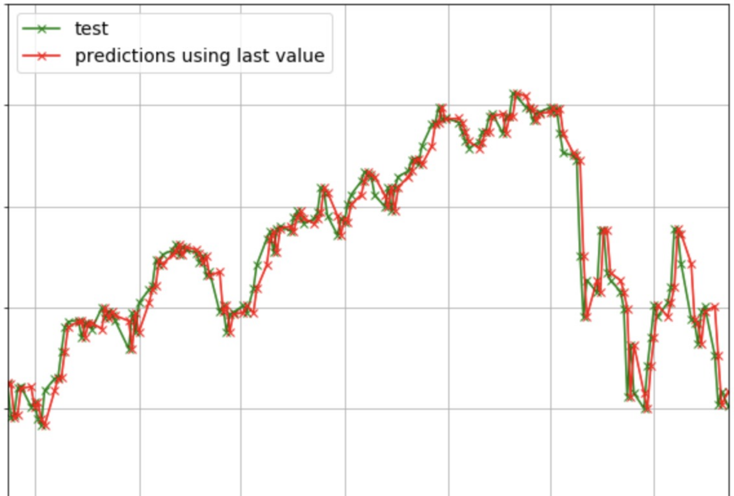
**Diagram Showing Selection of Wave Length and Ahead Speed**




**Method 1: the Mathieu Equation**

$$\ddot{\phi}(t) + \frac{B}{I_x}\,\dot{\phi}(t) + \frac{\Delta}{I_x}\,GM_i(t)\,\phi(t) = 0$$




**Method 2: modelling of random process**

The two figures are just for illustration purpose!

# Course contents on ML methods

# Course contents with focus on ML (1)

- **Part I:  Basic terminologies of Machine Learning**
  - AI, Machine Learning, Data Science, statistical learning
  - Supervised machine learning, Unsupervised, classification, data mining, Clustering
  - Useful toolbox (Python sci-kit learning, R, Tensorflow, Skaggle, etc.)
- **Part II: Common supervised learning (independent data)**
  - Linear regression, Polynominal regression
  - Spline regression, Logistical regression
  - Support vector machine
  - Decision trees, XGBoost,  Neural network

# Course contents with focus on ML (2)

- **Part III: Practical issues to use the common ML methods**
  - Feature selection
  - Model selection
  - Learning diagnostics
  - Cross validation, learning rate,  training/test/CV dataset
- **Part IV: Some more advanced statistical learning methods**
  - Generalized additive models (GAM)
  - Generalized linear models (GLM)
  - Mixed effect models (MEM)

# Course contents with focus on ML

- **Part V: ML for time series signals (correlated data signal)**
  - Basic statistics and correlation
  - Transformation (to Gaussian): Lognormal, exponential, Hermite polynomials
  - Moving average, AR, ARIMA
- **Part VI: Spatial-temporal modelling of Random field (optional)**
  - Prepare for data
  - Transformation to Gaussian
  - Model the correlation structure
  - Conditional prediction and simulation

# Course literature

- Hastie T., Ribshirani R. and Friedman J. (2003). The elements of statistical learning, Data mining, inference and prediction. Springer.
- Shalizi, C.R. (2019). Advanced data analysis from an Elementary point of view. Pre-print.
- Shumway, R.H. and Stoffer, D.S. (2016). Time series analysis and its applications with R examples, Fourth edition. Springer.
- Wei, W.W.S. (2006). Time series analysis Univariate and multivariate models, Second edition. Pearson Addison Wesley.

# Other ML algorithms/methods

1. Linear regression
2. Kernel ridge regression
3. Support Vector Machines
4. Stochastic Gradient Descent
5. Nearest Neighbors
6. Decision Trees (XGboost)
7. Ensemble methods
8. Multiclass and multilabel algorithms
9. Neural network models (supervised)

# Statistical learning algorithms

- Generalized linear model
- Spline models
- Generalized additive model
- Mixed effect model
- Gaussian processes/fields
- Autoregressive model
- Moving average model

# Other matters

AVANCEZ

1829

CHALMERS

UNIVERSITY OF TECHNOLOGY