Introduction to Machine Learning

Machine Learning for Process Engineers Workshop

Stellenbosch University

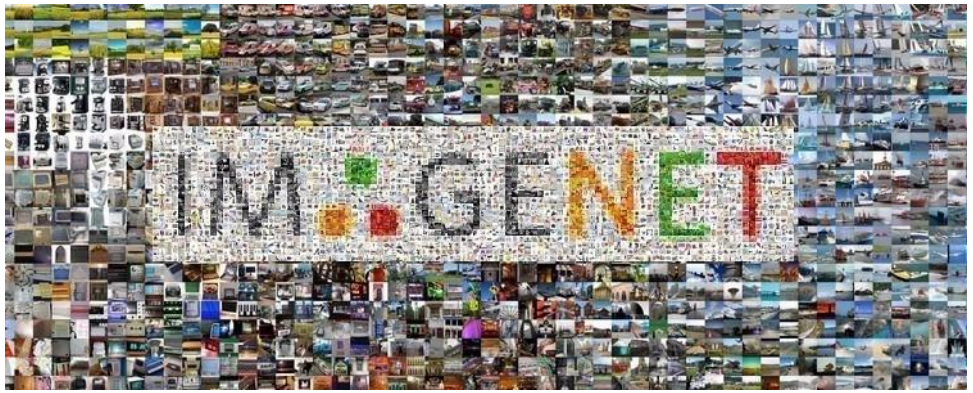
March 2022

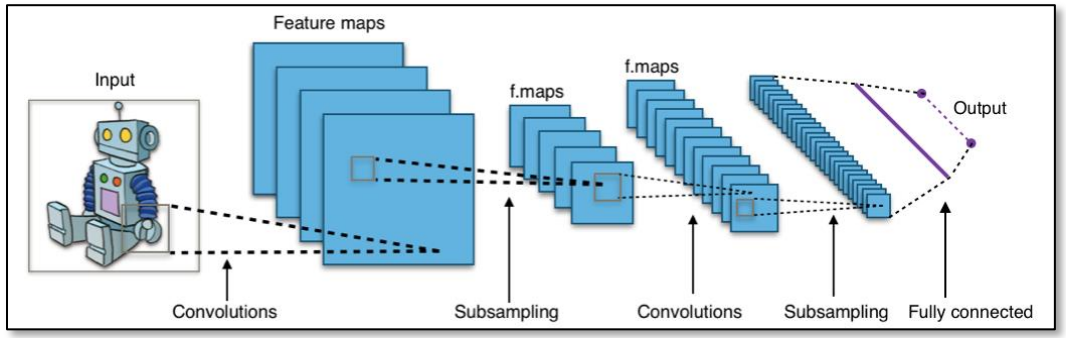
Recent interest in machine learning

Developments in ML have been spurred by:

- **Computational resources**
  (and computation as a service)
  2010: Jaguar at 1.75 x 10^{15} FLOPS
  2021: Fugaka at 442 x 10^{15} FLOPS
- **Massive data repositories**
  e.g., ImageNet containing
  >14 million annotated images
- **Deep learning**
  ability to analyse inputs consisting
  of thousands (millions) of features,
  e.g., 512x512 pixels > 260k features

AlphaGo





Recent interest in machine learning

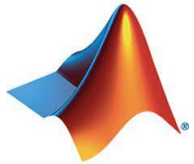
... and developments in ML have spurred:

- Industry interest
- Data awareness
- Development of effective data ecosystems

Overwhelmingly large toolbox

Support vector machine

Gaussian Mixture Model



Actor-critic

Markov Chain Monte Carlo

Partial least squares regression

PyTorch

Dynamic Bayesian network

PyMC3

SARSA

Long-short term memory neural network

Independent component analysis

pandas

Kernel PCA

DBSCAN

Linear Discriminant Analysis

scikits learn

K-nearest neighbours

Random forests

NumPy

Gaussian process regression

Spectral clustering

Ridge regression

Bayesian linear regression

Variational autoencoder

LASSO

Granger Causality

Probabilistic matrix factorization

Q-learning

Goal of this workshop

Provide exposure to fundamentals of machine learning methods to enable effective independent learning

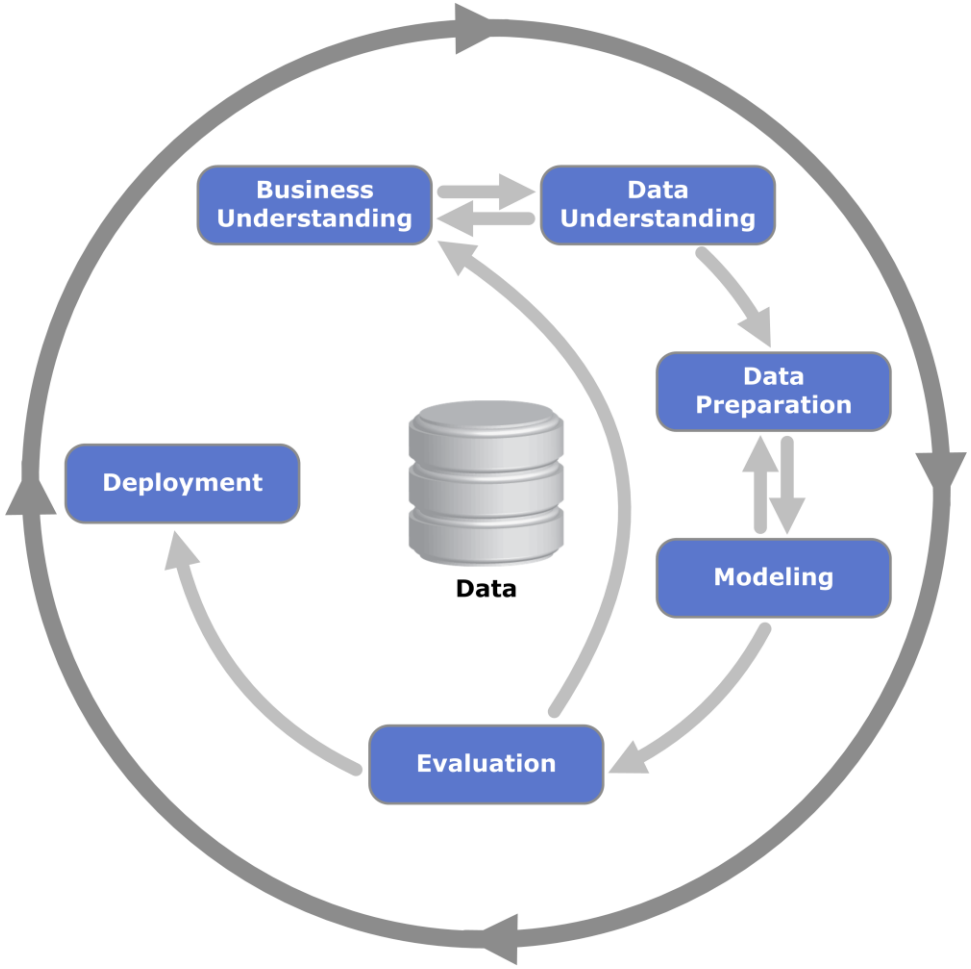
- The bias-variance trade-off
- Estimating model accuracy
- Improving model performance using
  - Regularisation
  - Dimensionality reduction (feature extraction)

Goal of this workshop

Provide exposure to fundamentals of machine learning methods to enable effective independent learning

- Focus on linear regression for supervised learning
- Introduce elementary unsupervised learning through principal component analysis
- Provide a brief overview of state-of-the-art methods in ML

Goal of this workshop



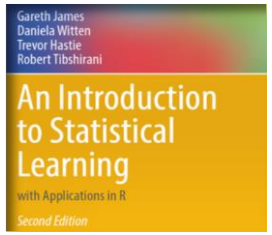
CRISP-DM: Cross-industry standard process for data mining

- Effective deployment of data-based methods requires investment in each phase of the data mining process
- Hands-on workshop will focus on modelling, but remaining factors must not be neglected

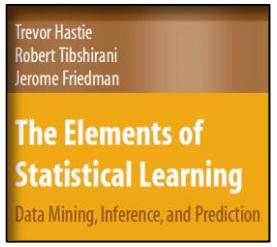
Goal of this workshop

- NOT a zero-to-Deep Learning course
  (although we can discuss the benefits of Deep Learning as questions come up)
- NOT a tutorial on ML / Data Science software or libraries
  (although we can mention useful and popular ones)
- NOT an all-encompassing overview of ML
  (although we hope that the workshop enables you to engage with the field)
- **IT IS a workshop:**
  please ask questions and help us learn together!

Resources

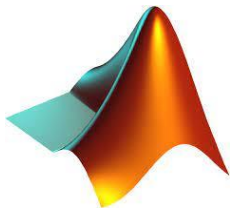


- An Introduction to Statistical Learning (2^{nd} ed. 2021)
  James, Witten, Hastie, Tibshirani
  Free: https://www.statlearning.com/



- Elements of Statistical Learning (2^{nd} ed. 2009)
  Hastie, Tibshirani, Friedman
  Free: https://web.stanford.edu/~hastie/ElemStatLearn/

- Machine Learning: A First Course for Engineers and Scientists (in press)
  Lindholm, Wahlström, Lindsten, Schön
  Free: https://smlbook.org/



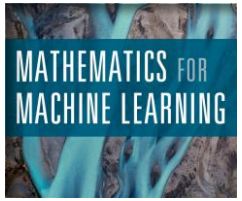
- Mathworks Self-Paced courses
  Free (w/ license): https://matlabacademy.mathworks.com/#ai

Resources

- Pattern Recognition and Machine Learning (1st ed., 2006)
  Bishop
  Not free, but beautiful: https://link.springer.com/book/9780387310732

- Reinforcement Learning: and introduction (2nd ed. 2018)
  Sutton, Barto
  Free, focus on reinforcement learning: http://www.incompleteideas.net/book/the-book-2nd.html

- Probabilistic Machine Learning: An Introduction (2022)
  Murphy
  ***Free:** https://probml.github.io/pml-book/book1.html



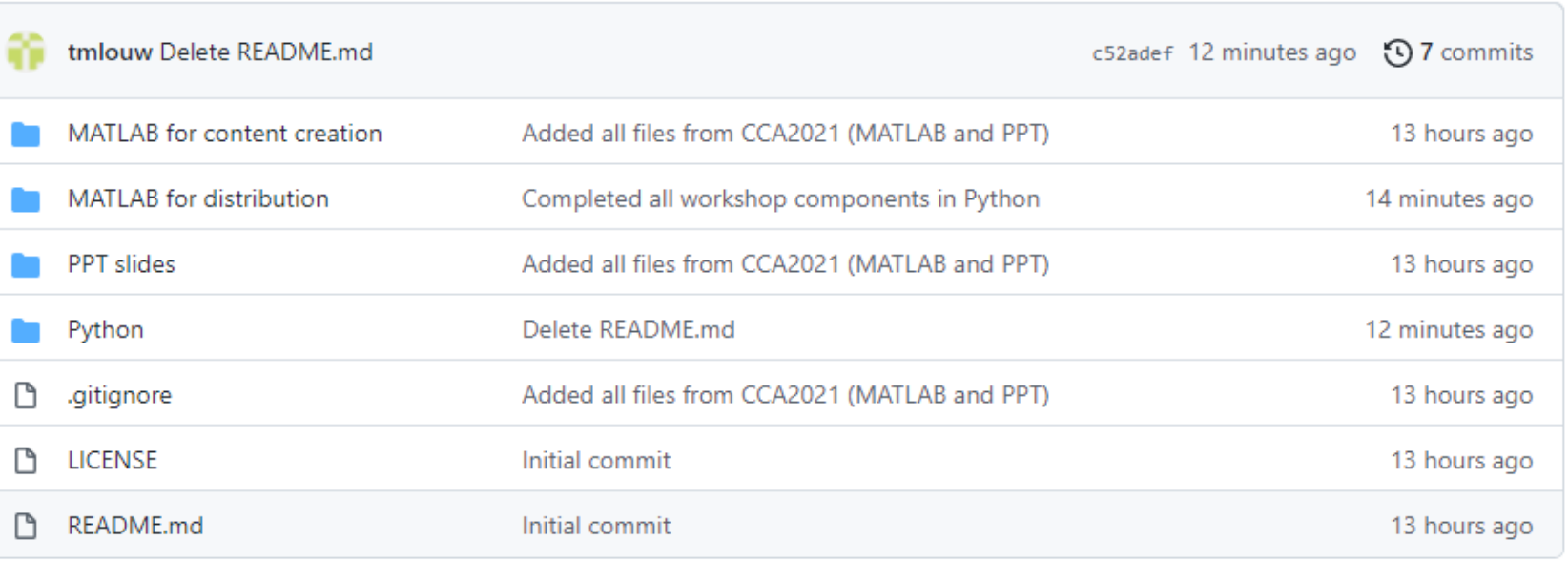
- Mathematics for Machine Learning(2020)
  Deisenroth, Faisal, Ong
  ***Free, introduction to underlying mathematics:** https://mml-book.com/

- Foundations of Machine Learning (2nd ed. 2018)
  Mohri, Rostamizadeh, Talwalkar
  ***Free, introduction to underlying mathematics:** https://cs.nyu.edu/~mohri/mlbook/

* I have not read these books myself, but they are recommended by Lindholm et al. (smlbook.org)

Resources

https://github.com/tmlouw/ML-for-Process-Engineers



tmlouw Delete README.md — c52adef 12 minutes ago — 7 commits

MATLAB for content creation	Added all files from CCA2021 (MATLAB and PPT)	13 hours ago
MATLAB for distribution	Completed all workshop components in Python	14 minutes ago
PPT slides	Added all files from CCA2021 (MATLAB and PPT)	13 hours ago
Python	Delete README.md	12 minutes ago
.gitignore	Added all files from CCA2021 (MATLAB and PPT)	13 hours ago
LICENSE	Initial commit	13 hours ago
README.md	Initial commit	13 hours ago

README.md

ML-for-Process-Engineers

This repository contains the relevant files for a workshop presented to introduce the basic concepts of machine learning in the context of chemical- and minerals processing