Machine Learning Algorithms

This is an introduction.

It provides an overview of some common ML algorithms.

Clustering   Two common algorithms are

k-means clustering

hierarchical clustering

Regression   The most common algorithm is

linear regression

Normally, you apply linear regression to
a linear-linear plot of two variables x-y

You can also combine it with feature engineering and use

x-log(y) plots

or even

log(x)-log(y) plots

to discover linear relationships between those variables

and transform that to a relationship between x and y

Classification The common algorithms are

logistic regression – yes, it does classification

k-Nearest Neighbour

decision trees — but CART trees also do regression

random forest

support vector machine

artificial neural network

Most importantly, you must know how to use

the Python `scikit-learn` library

to build and evaluate models,

as shown in Example 2.

You should know the following:

- ▶ What is a model

- ▶ What kind of models does ML build

- ▶ Where does feature engineering fit in ML

- ▶ What is unsupervised machine learning

- ▶ What is supervised machine learning
  label, class
  binary classifier
  multi-class classifier
  multi-label classifier

- ▶ What is regression, classification, prediction

You do **not** need to know:

- how the ML algorithms work

- how to handle imbalanced data

- how to generate data

- the theory or statistics behind the ML algorithms

- technical details of the issues, such as

  over-fitting,
  independence (additivity, homoscedasticity),
  regularization

- semi-supervised learning

- reinforcement learning

- deep learning