

# Understanding Machine Learning

## Description

Machine Learning is a well understood process. We typically start with some existing data and pass it through an algorithm. The algorithm 'learns' from that specific data and produces a 'data model'. This model has learnt from the data and now encapsulates information derived from the raw data. We then have to test the model (to see how good it is) and try to incrementally improve it. Finally, we evaluate the finished model and deploy it.

## Prerequisites

- An understanding of data
- A good logical mind
- We do not expect people to have a background in mathematics

## Outline

### Introduction

- Definition of Machine Learning (ML)
- Origins of ML
- Rule deduction (Expert Systems) vs induction (ML)
- Why do we want machines to learn?
- Supervised vs. unsupervised learning
- Case studies
- Regression as a classic example of ML

### Data collection and preparation

- Data selection
- Data sampling
- Data volume reduction
- Removing ambiguities
- Normalisation
- Discretisation
- Cleansing
- Missing values
- Outliers
- Data and dimensional reduction
- Data understanding
- Generalisation of hierarchies

### **Introduction to ML in Python**

- Introduction to Python
- ML with Python

### **Creating or choosing an algorithm**

- Examples of creating algorithms
- The use of data mining algorithms
- Classes and examples of data mining/Machine Learning algorithms
- Decision trees
- Clustering
- Segmentation
- Association
- Classification
- Sequence analysis
- Neural nets
- History
- Layers
- Weights
- Back propagation
- Deep Learning
- KNN
- SVM

### **Training and test data**

- Selecting the training and testing data
- Ratio of training to test data
- How to make an unbiased selection

### **Testing and confusion matrices**

- Type 1, 2 and 3 errors
- False positives vs False negatives
- PCC
- Classification models
- Confusion matrices

### **ROC curves**

- Measuring efficiency
- ROC space and ROC curves

### **Efficiency, Overfitting, Bias and Variance**

- More about efficiency
- Overfitting
- Bias and Variance

### **Combining data models**

- Ensemble
- Boosting
- Gradient boosting
- Case study of combining models
- Summary