

# Introducing Machine Learning

Jan-Philipp Kolb

07 Mai, 2019

# Intro Machine Learning

Categorizing machine learning algorithms is tricky

- ... they can be grouped into generative/discriminative, parametric/non-parametric, supervised/unsupervised, and so on.
- Scikit-Learn's documentation page groups algorithms by their learning mechanism. This produces categories such as: Generalized linear models, Support vector machines, nearest neighbors, decision trees, neural networks, ...

# Machine Learning - Components

- Feature Extraction + Domain knowledge
- Feature Selection
- Choice of Algorithm - e.g. Naive Bayes, Support Vector Machines, Decision Trees, k-Means Clustering, ...
- Training
- Choice of Metrics/Evaluation Criteria
- Testing

# Feature selection

Feature selection is for filtering irrelevant or redundant features from your dataset. The key difference between feature selection and extraction is that feature selection keeps a subset of the original features while feature extraction creates brand new ones.

To be clear, some supervised algorithms already have built-in feature selection, such as Regularized Regression and Random Forests. Typically, we recommend starting with these algorithms if they fit your task.

As a stand-alone task, feature selection can be unsupervised (e.g. Variance Thresholds) or supervised (e.g. Genetic Algorithms). You can also combine multiple methods if needed.

# Supervised vs unsupervised learning

## Supervised Learning

- we have prior knowledge of what the output values for our samples should be.

## Unsupervised Learning

- In unsupervised learning we observe only the features  $X_1, X_2, \dots, X_p$ . We are not interested in prediction, because we do not have an associated response variable  $Y$ .

## Task: Find R-packages

Go to <https://cran.r-project.org/> and search for packages that,...

- can be used for lasso regression

# Task View Machine Learning

CRAN Task View: Machine Learning & Statistical Learning

**Maintainer:** Torsten Hothorn

**Contact:** Torsten.Hothorn at R-project.org

**Version:** 2018-08-05

**URL:** <https://CRAN.R-project.org/view=MachineLearning>

Several add-on packages implement ideas and methods developed at the borderline between computer science and statistics - this field of research is usually referred to as machine learning. The packages can be roughly structured into the following topics:

- *Neural Networks and Deep Learning* : Single-hidden-layer neural network are implemented in package [nnet](#) (shipped with base R). Package [RSNNS](#) offers an interface to the Stuttgart Neural Network Simulator (SNNS). [rnn](#) implements recurrent neural networks. Packages implementing deep learning flavours of neural networks include [deepnet](#) (feed-forward neural network, restricted Boltzmann machine, deep belief network, stacked autoencoders), [RcppDL](#) (denoising autoencoder, stacked denoising autoencoder, restricted Boltzmann machine, deep belief network) and [h2o](#) (feed-forward neural network, deep autoencoders). An interface to [tensorflow](#) is available in [tensorflow](#).

## Install all packages of a task view

```
install.packages("ctv")  
ctv::install.views("MachineLearning")
```



# Prediction vs. Causation in Regression Analysis

# Literature for machine learning



# Introduction to machine learning with R

- Your First Machine Learning Project in R Step-By-Step
- chapter about machine learning in awesome R
- Shiny App for machine learning

# Time measurement

```
start_time <- Sys.time()
ab <- runif(10000000)
end_time <- Sys.time()

end_time - start_time

## Time difference of 1.139065 secs
```

# How many cores are available

```
library(doParallel)  
detectCores()  
  
## [1] 4
```

```
cl <- makeCluster(detectCores())
registerDoParallel(cl)

start_time <- Sys.time()
ab <- runif(10000000)
end_time <- Sys.time()

end_time - start_time

## Time difference of 1.297075 secs

stopCluster(cl)

?parallel::makeCluster
```

# Links

- Presentations on 'Elements of Neural Networks & Deep Learning'
- Understanding the Magic of Neural Networks
- Neural Text Modelling with R package ruimtehol
- Feature Selection using Genetic Algorithms in R
- Lecture slides: Real-World Data Science (Fraud Detection, Customer Churn & Predictive Maintenance)
- Automated Dashboard for Credit Modelling with Decision trees and Random forests in R
- Looking Back at Google's Research Efforts in 2018
- Selecting 'special' photos on your phone
- Open Source AI, ML & Data Science News