

Building meaningful machine learning models for disease prediction

Dr. Shirin Glander

Dep. of Genetic Epidemiology
Institute of Human Genetics
University of Münster

shirin.glander@wwu.de

<https://shiring.github.io>
<https://github.com/ShirinG>

31.03.2017

Table of contents

- 1 Machine Learning (ML) in disease modeling
- 2 A quick recap of ML basics
 - Supervised vs Unsupervised algorithms
 - Classification vs Regression
 - Features
 - Training, validation and test data
 - Cross-validation
 - Missing data
 - Grid Search
- 3 How to build ML models in R
 - Get to know your data
 - Classification
 - Tree based models
 - Regression
 - Linear Models
 - Deep learning with neural network
- 4 Evaluating ML model performance

Machine Learning (ML) in disease modeling



*“Machine Learning for Detection and Diagnosis of Disease” (2006). In: *Annual Review of Biomedical Engineering* 8.1. PMID: 16834566, S. 537–565.

A quick recap of ML basics

How to build ML models in R

Session setup

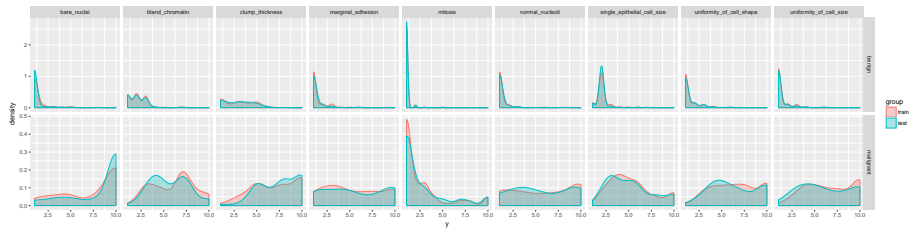
Code will be available on [my website](#) and on [Github](#)

Breast cancer Wisconsin dataset

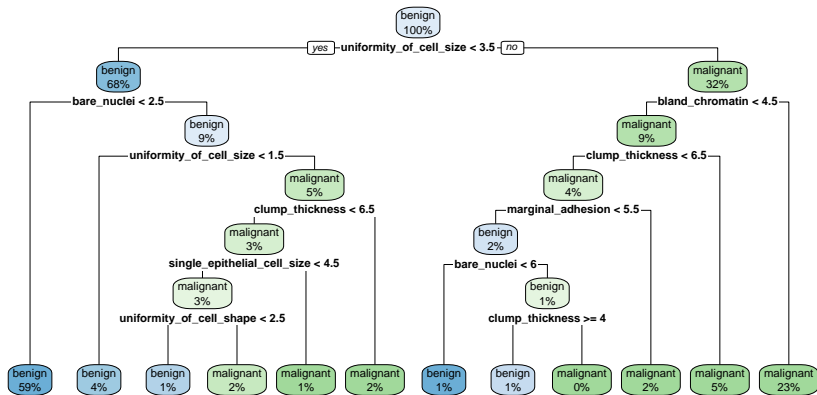
caret package

h2o package

Distribution

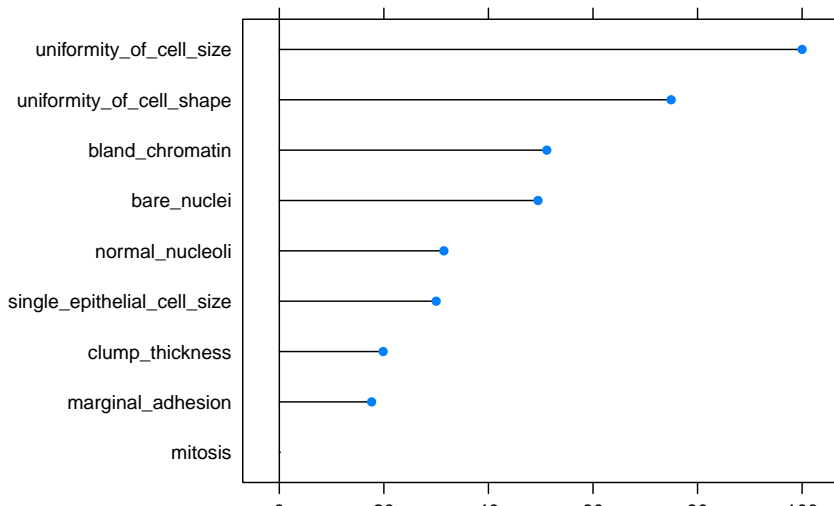


Decision trees

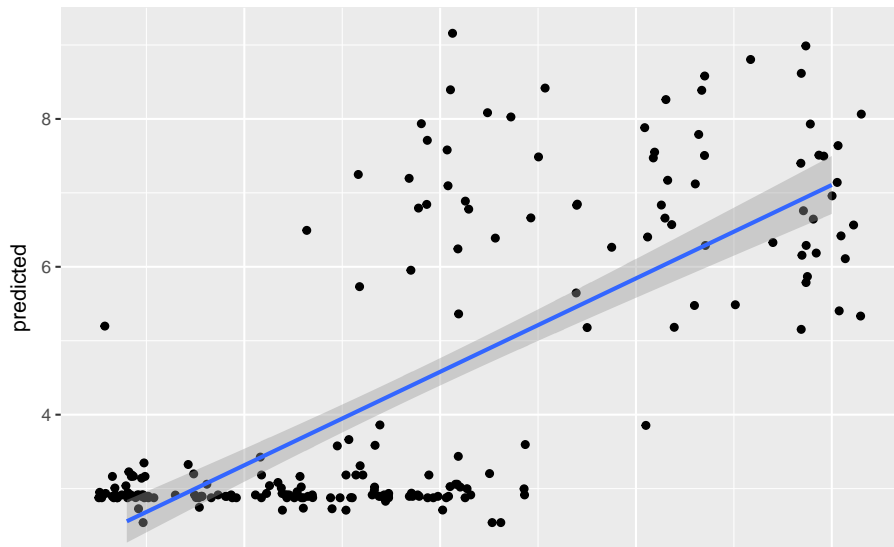


Random Forest

Feature importance

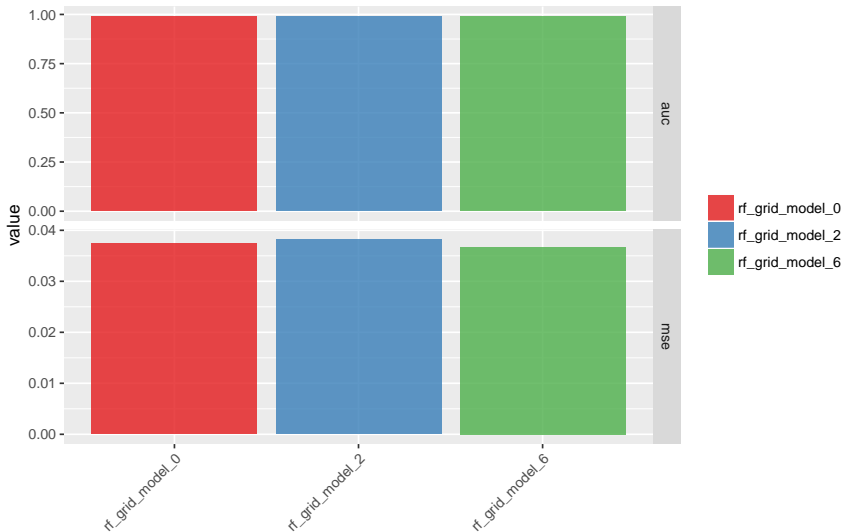


(Generalized) Linear Models



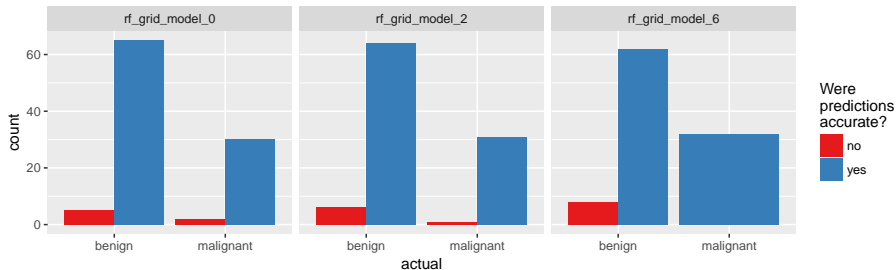
Evaluating ML model performance

AUC and MSE

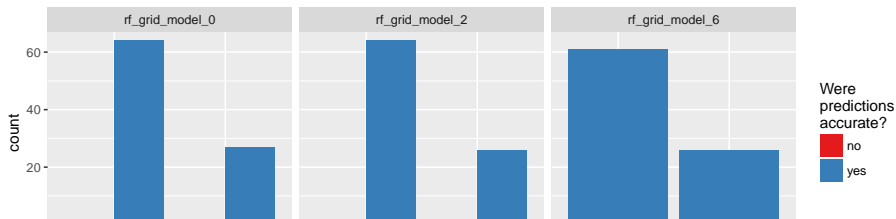


Predictions on test data

Default predictions



Stringent predictions



Thank you for your attention!

Questions?

Slides and code will be available on Github:

https://github.com/ShirinG/Webinar_ML_for_disease

Code will also be on my website: <https://shiring.github.io>

shirin.glander@wwu.de