

Learning Guide

Machine Learning with Tidymodels (4/23/21)

Carsten Lange

Table of Contents

- [General](#)
- [Part 1: Overlearning](#)
- [Part 2: Using Tidymodels to Clean Data, Create and Assess Machine Learning Models](#)
- [Part3: Lasso, Ridge, and ElasticNet Models](#)
- [Part 4: Cross Validation with Tidymodels](#)

General

Begin: 04/23/2021 End: 04/23/2021

Learning Module Question

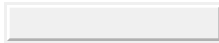
Video, Presentation, and RScripts (coming soon)

Learning Outcomes

Multimedia Resources

- [Video Recording 4/23/2021 \(coming soon\)](#)
- [Presentation Slides 4/23/2021](#)
- [R-Script from Presentation on 4/23/2021 \(coming soon\)](#)

Literature



Part 1: Overlearning

Begin: 04/23/2021 End: 04/23/2021

Learning Module Question

What is Overlearning and how can it be identified?

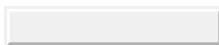
Learning Outcomes

- Review Regression
- Recognize the Problem of Overlearning
- Identify Overlearning with Train and Test Sets

Multimedia Resources

- [RScript for a very basic regression example \(Reading Level\)](#)
A textbook example is used to review basic regression concepts such as OLS coefficients, (squared) error, and estimates.
- [Shiny Manual Regression Tool](#)
Click on the link and you can perform a regression analysis in a browser by manually changing the regression coefficients. The simulation is a R Shiny App.
- [Error Surface for The King County Housing Dataset](#)
This is an interactive Html 3D graph. You can choose values for b_0 and b_1 (intercept and slope) similar as you did in the Shiny app. The resulting Mean Squared Error (in billions) is the height of the error surface.
- [RScript: Opimizer.R](#)
This R-Script shows how to use R's `optimizer()` function to find the parameters b_0 and b_1 for a linear regression with one predictor. Almost every machine learning algorithm uses an optimizer similar to the one demonstrated in this R-script.
- [Finding the Error Surface Minimum Mathematically with the Method of Gradient Decent](#)
The video describes the mathematics to find the Error Surface minimum with the Method of Gradient Decent. It is a little more advanced> Although not relevant for the exam it is well made and understanding the underlying math might help to understand optimizers in machine learning.
- [RScript: Analyzing Daily Minimum Temperatures in Melbourne/Season as Categorical Variable](#)
This R Script shows how to analyze seasonal effects of temperature. It introduces dummies for the seasons (Spring, Summer, Fall). The dataset for the script is provided by Jason Brownlee in a blog titled "7 Time Series Datasets for Machine Learning" (see: <https://machinelearningmastery.com/time-series-datasets-for-machine-learning/>). The download link above loads the R-Script and from there you can load the data.
- [Wage Data for the in Class Exercise](#)
Here you can download data to estimate hourly wages based on several predictors including categorical variables. The data were copied into Excel from the Woolridge dataset (see `library(woolridge)`).
Please use the sheet DataClean for the class exercises.

Literature



Part 2: Using Tidymodels to Clean Data, Create and Assess Machine Learning Models

Begin: 04/23/2021 End: 04/23/2021

Learning Module Question

How can we make machine learning analysis more productive?

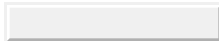
Learning Outcomes

- How to build a OLS model with **Tidymodels**
- **Cleaning and engineering** data with Recipies
- **Building** the model
- **Assessing** predictive quality

Multimedia Resources

- [A Gentle Introduction to Tidymodels by RStudio](#)
A step by step introduction to TidyModels
- [R Tidy Models Video for Machine Learning Part 1](#)
The video is a recording from the class and uses an OLS model to show how tidymodels can be utilized to create recipes, define models, and analyze the error metrics. Part 2 will demonstrate how to tune hyper-parameters of ElasticNetModel.
- [Recipes Currently Availabe for Tidymodels](#)
Here is a list of all currently available tidymodels recipes, most of these recipes are for more advanced users.

Literature



Part3: Lasso, Ridge, and ElasticNet Models

Begin: 04/23/2021 End: 04/23/2021

Learning Module Question

How to Regulate Machine Learning Parameter?

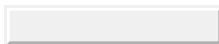
Learning Outcomes

- Understand the Lasso Model
- Understand the Ridge Model
- Understand the ElasticNet Model

Multimedia Resources

- [Tutorial with Code about Ridge Regression](#)
Tutorial about Ridge Regression based on the Cars dataset. Explains Ridge and when to use it. R Code based on glmnet package provided. Well and short written. Good demos.
- [Ridge and Lasso Example](#)
This R Script provides a annotated example for an Ridge and Lasso regression using the glm package. It emphasizes the role of lambda.
- [Dutta, Sujan, Diving Deeper into Linear Regression](#)
Extends Regression to Ridge, Lasso, Orthogonal, etc.
- [Regularization: Ridge, Lasso and Elastic Net \(by M. Oleszak\)](#)
A little more advanced but full of useful information. You will get acquainted with the **bias-variance trade-off problem** in linear regression and how it can be solved with regularization.
- [Ridge Regression R Script](#)
Here you can download the Ridge Regression R script that we used in class.

Literature



Part 4: Cross Validation with Tidymodels

Begin: 04/23/2021 End: 04/23/2021

Learning Module Question

How can hyper-parameters be tuned?

Learning Outcomes

Be able to use Tidymodels to **tune hyper-parameters** with **cross-validation**.

Multimedia Resources

- [RScript for tuning and cross validation](#)
RScript from class
- [R Tidy Models Video for Machine Learning Part 2](#)
This is Part 2 of the video sequence. The video is a recording from the class and uses an ElasticNet to demo cross validation and hyper-parameter tuning with tidymodels. Part 1 shows how to create recipes, define models, and analyze the error metrics based on an OLS model.

Literature

