University of Koblenz - AG Softlang

Data Science

Assignment 6: The Linear Model

To be submitted until the 22.12.2021 (2 p.m)

December 8, 2021

Question 6.1: Exploring

The supplementary material contains the full data set on the *!Kung people*. You are supposed to create a linear model and this time we are also working with the children. Hence, do not filter the data set. The linear model should relate the age (predictor) with the height of a person (outcome).

In this first part of the assignment, plot the two variables in a scatter plot (age on horizontal, height on vertical). You will notice that the data entries seems to be far away from a straight line. Provide us with the plot as PDF.

Question 6.2: Lines

However, stick to the basic linear modeling strategy that has been introduced in the lecture, and model the height as a normal distributed variable with a mean (mu, μ) depending on the age. For the model, you need the parameter alpha, beta and sigma. Fit the model using *ULAM* from the R *rethinking*[1] package, or code it in *Stan*[2] and run it from R or Python.

Plot the data in a scatter plot (as in the previous question 6.1) with 100 regression lines taken from the posterior as overlay. You will notice, this model is a mess! Anyways, please provide the code for fitting the model, and the plot as PDF.

Question 6.3: Curves

There are some options of improving this fit a little. One option is to do a *polynomial regression*. In this case, the linear model adds new predictors variables, i.e., squares or cubes of predictor variable age (i.e., age^2 and age^3). This implies that you need additional parameters to model the linear combination of the variables. You may introduce new parameters named *beta1*, *beta2* and *beta3*. Fit the model adding age^2 (and optionally also the age^3) to the linear part of the model and provide the same artifacts (plot and code) as in the previous assignment. (BTW: The fit will also not be very good.)

[1] http://github.com/rmcelreath/rethinking
[2] http://mc-stan.org