Deep and Statistical Learning WS 21-22

Report (15 Seiten) (Sebi/Tobi)

This repo holds all the code we wrote for the deep learning seminar WS 21/22. While the main analysis was done in R, the image analysis was done in python. The scripts and files for the image analysis can be found in the folder "image analysis" inside "data". Since we worked with google drive for scraping the images and analysing them, the scripts do not run in this repo. However the basic folder structure was copied to this repo, when it was possible. So copying the content of the folder "image analysis" to a drive account and then openening the colab scripts should set up the enviroment. For the multi detection dnn, one must first download the repository dircribed in the script for the multi detection model and in the paper.

Abstract of the final paper:

Price prediction is one of the classical fields for the application and evaluation of modern innovative methods of statistical learning. Even if now better established, Deep Neural Networks (DNN) are still very promising and empirical research on their usage for regression tasks is relevant. As a part of the statistics department's empirical seminar "Statistical and Deep Learning" of the University of G¨ottingen, we use Inside Airbnb's public available data of the city of Berlin to use DNN and further statistical models for prediction of the price of individual listings. Firstly, we shortly describe the available data and then the process of cleaning and preparing the data for modeling. Thereafter, we present how we analyzed images of the flats to generate further variables. This includes the detection of multiple objects on the the images using convolutional neural networks (CNN) as well as the analysis of the color temperature (CT) and perceived brightness (PB) of the images. Consequently, we describe our DNN and further modeling techniques used for price prediction and the analysis of variable importance. Finally, we present, interpret and discuss the attained results using classical performance measures. Further architectural considerations and insights for DNN are presented on the appendix.