Machine Learning Project Report

Lena Loye, Anas Yusuf

December 2021

1 Introduction

In this competition we get measurements from different weather stations in Switzerland. The goal was to predict whether there is some precipitation (rain, snow etc.) on the next day in Pully. This problem is a typical binary classification task. We first need to explore our data set, and then we will approach this problem using both linear and non-linear methods. To compare the different approaches throughout the project we mainly used the accuracy and AUC.

2 Process and results

2.1 Exploration

A preliminary exploration of the data allow us to get a first ensemble view of the main characteristics of the data set. To predict if there will have precipitation on the next day, we were provided 529 predictors and 3176 observations.

2.2 Linear Methods

2.3 Non-Linear Methods

2.3.1 K-Nearest Neighbor

2.3.2 Trees

2.3.3 Neural Networks

3 Conclusions