Project 1: Data analysis

Group 11 [1]

[1] Department of Mathematics

October 1, 2021

Preliminary report

The aim of this project is to study how certain features of the data set has an effect on on the symptoms of Covid-19. The first part of the project is divided in two parts. Firstly, we will study the effect of genes on the symptoms. Secondly, we will estimate the efficacy of vaccines and the probability of vaccination side-effects. We will use Bayesian inference and/or Lasso for feature selection. For the Lasso case, we will fit a Logistic regression model before performing feature selection using Lasso cost function. By tuning the hyper-parameter in Lasso, we acquire the most accurate model.

For estimating the efficacy of vaccines, we will look at "Covid-Positive" samples with vaccination and whether or not they have severe symptoms. We will also compare the efficacy of the three different vaccines. To study the side-effects, we will look at those samples that has been vaccinated and are not "Covid-Positive". Further, we will study how the different vaccines effects the patients and compare them.

The second part of the project is to model the treatment of the patients who have symptoms. These patients are treated with specific treatments (actions), and we want to analyze which treatment gives the best results for alleviating symptoms and preventing death. We want to calculate the probabilities of negative outcomes (severe symptoms) given a specific action. We will use these probabilities to compare the efficacy of the different actions. We want compare the group that got no treatment with the groups that had either action "1" or action "2", or both of the actions.