

2NN - Gastrointestinal lesions

Bjerke, Martin & Spremic, Mina

2/10/2021

Project 1 - Shrinkage

The goal of the project is to apply the shrinkage methods to a dataset of the groups choice. As one of the main methods of the first part of the course has been lasso, applied in cases when the number of covariates is much larger than number of observations, in addition to there being very few observations. Hence, we have chosen a dataset satisfying these criteria. The dataset in question is a Gastrointestinal Lesions in Regular Colonoscopy data set, which can be accessed at *insert url.*

This dataset had to be labeled all over again, as the txt file including the data, did not have a header. The relabeled dataset is available in the git-repo. In this dataset there are 701 covaraites, which are not neccessarily easy to decipher, both due to the large number and the lack of information on them. The data set is connected to a published paper “quote this in a nice way:Computer-Aided Classification of Gastrointestinal Lesions in Regular Colonoscopy”, which can be accessed here *insert url.*

This is a classification type of problem, where we wish to classify a growth in the bowels, whether it is benign or malignang. The covariates that we have to aid us in this are based on the videos of these growths (lesions). Features are extractes from the videos, and they are the following:

insert a table with the different sets of covariates and their meaning, to the best of our abilities

```
summary(cars)
```

```
##      speed      dist
## Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
## Max.   :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.