

# Homework3

Statistical Learning for Data Science

Due time: April 3, 2023 (Monday) 12:00

## 1 Answer the question

Explain the differences between the KNN classifier and KNN regression methods.

## 2 Answer the question

It is claimed in the text that in the case of simple linear regression of  $Y$  onto  $X$ , the  $R^2$  statistic is equal to the square of the correlation between  $X$  and  $Y$ . Prove that this is the case. For simplicity, you may assume that  $\bar{x} = \bar{y} = 0$ .

## 3 Answer the question

Why do linear regression models assume that the error terms are normally distributed?

## 4 Answer the question

In a linear regression model, what happens if there is collinearity between independent variables? How to handle collinearity?

## 5 Answer the question

What are the commonly used variable selection methods in multiple linear regression? Under what circumstances should it be used?

Paper version should be submitted in class.

The electronic version should be sent in PDF format in the form of "homework3-name-student ID" to 12132147@mail.sustech.edu.cn