Tutorial 4 Regression

Lecture

Regression Analysis

From this week onwards, we will be exploring different types of supervised learning methods. Generally, a predictive task can be either a classification or regression problem. In a classification problem, the goal is to categorise data into predefined classes or categories. For example, classifying emails as "spam" or "not spam" involves predicting which of the **discrete categories** an email belongs to based on its features. In contrast, a regression problem involves predicting a **continuous numerical value**. For instance, forecasting the price of a house based on features like location, size, and number of rooms requires estimating a continuous output. While classification yields discrete outcomes, regression provides a continuous range of values, and the choice between the two depends on the nature of the prediction task at hand.

We will start with the simplest approach: assuming a linear relationship between predictors and response. This week, we focus on regression problems, and next week, we will explore models for classification problems.

Textbook Reading

Prescribed reading

Author	Title	Relevant chapters
James et al	An Introduction to Statistical Learning with Applications in R	ch 3
Faraway	Linear model with R	entire book

Tutorial

There is so much to learn about linear and generalized linear models; attempting to thoroughly understand the art of linear modelling in just one or two lectures is ludicrous. But this is unfortunately the modern approach to 'Data Science': knowing how to program it without understanding the ins and outs of the model. Only by truly understanding linear or generalised linear model can we grasp why they are often not desirable for the problem at hand, yet still frequently chosen as contenders.

In a group of two, discuss the following topics

- What is the model structure of a linear model?

- What are the parameters of a linear model?

- How do we estimate the parameters?

- Why is there a variance associated with each parameter?

- Explain a residual in layman's terms

- What are assumptions of a linear model?

- Explain variance-bias trade-off

- Describe how forward and backward model selection work?

- Assume a predictor has four categories; explain why R only produces coefficients for three of those categories.

Independendent Learning

Labs

Exercises

- Question 1, Chapter 3.7

- Question 3, Chapter 3.7

- Question 9, Chapter 3.7 - (exclude 9(f))

- Question 10, Chapter 3.7