

Introduction to Linear Regression - Introduction

Introduction

In this section, you're going to learn about one of the most basic machine learning models, linear regression! Many of the ideas you learn in this section will be foundational knowledge for more complex machine learning models.

Statistical Learning Theory

We'll start this section by exploring Statistical Learning Theory and how dependent and independent variables relate to it. Statistical Learning Theory provides an important framework for understanding machine learning.

Linear Regression

In this section, we'll introduce our first machine learning model - linear regression. It's really just a fancy way of saying "(straight) line of best fit", but it will introduce a number of concepts that will be important as you continue to learn about more sophisticated models.

Coefficient of Determination

We're then going to introduce the idea of "R squared" as the coefficient of determination to quantify how well a particular line fits a particular data set.

A Complete Regression

From there we look at calculating a complete linear regression, just using code. We'll cover some of the assumptions that must be held for a "least squares regression", introduce Ordinary Least Squares in Statsmodels and introduce some tools for diagnosing your linear regression such as Q-Q plots, the Jarque-Bera test for normal distribution of residuals and the Goldfield-Quandt test for heteroscedasticity. We then look at the interpretation of significance and p-value and finish up by doing a regression model of the Boston Housing data set.

Summary

Congratulations! You've made it through much of the introductory data and we've finally got enough context to take a look at our first machine learning model, while broadening our experience of both coding and math so we'll be able to introduce more sophisticated machine learning models as the course progresses.