Final Year Project Report

Full Unit - Interim Report

Comparison of Ridge Regression Algorithm against Others When Solving Boston Housing Problem

Fabio Eugenio
dos Santos de Sampaio Doria

A report submitted in part fulfilment of the degree of

BSc (Hons) in Computer Science

Supervisor: Nicolo Colombo

Fabio Eugenio
dos Santos de Sampaio Doria

Department of Computer Science
Royal Holloway, University of London

November 24, 2023

Declaration

This report has been prepared on the basis of my own work. Where other published and unpublished source materials have been used, these have been acknowledged.

Word Count:

Student Name: Fabio Eugenio
dos Santos de Sampaio Doria

Date of Submission:

Signature:

Table of Contents

Abstract

For most people buying a house will be one of their most important and expensive economic decisions that they will take in their lives.[1] Because of this it would be logical to say that being able to accurately predict the prices of said houses would be of extreme value to people. One possible way to make these predictions would be to create a machine learning model that, given a certain amount of features from each house, would be able to create an accurate prediction of their price.[2] There are a wide range of different machine learning algorithms that could be used to solve this problem. However, this project will focus on two: Ridge Regression and Decision Trees.

In this project I plan on implementing both of these algorithms and comparing their performance on one another to see which one is more effective at resolving the Boston housing problem. To measure their accuracy I will use two metrics specifically used for regression problems, mean absolute error (MAE) and root mean squared error (RMSE). MAE returns the average residual of the predictions that each model makes, this is useful as it can be directly compared against the other model for differences.[3] However, RMSE returns the square root of the average residual of the predictions. This highlights larger errors caused by the model which is practical as it helps to differentiate the performance of both algorithms.[4]

The dataset that will be used to judge the different algorithms is called the Boston housing problem datasets. The dataset is comprised of 506 entries, each having a total of 14 features which describe a property inside of the Boston Massachusetts area.[5] Two variations of this dataset will be used, one with full 14 features and another with a lower number 5. Two models will be created, one with each algorithm and each of these models will be trained and tested on both of these datasets. Their performance will then be analysed, compared, and finally conclusions will be drawn from the results in order to define the effectiveness both algorithms.

Project Specification

Your project specification goes here.

Chapter 1: **Introduction**

The project report is a very important part of your project and its preparation and presentation should be of extremely high quality. Remember that a significant portion of the marks for your project are awarded for this report.

The format of the final report is fixed by the template of this document and the Department of Computer Science suggests its usage.

While this may sound like a rather prescriptive approach to report writing, it is introduced for the following reasons:

1. The template allows students to focus on the critical task of producing clear and concise content, instead of being distracted by font settings and paragraph spacing.
2. By providing a comprehensive template the Department benefits from a consistent and professional look to its internal project reports.

The remainder of this document briefly outlines the main components and their usage.

A **final project report** is approximately 15,000 words and must include a word count. It is acceptable to have other material in appendixes. Your **interim report** for the December Review meeting, even if it is a collection of reports, should have a total word count of about 5,000 words. This should summarise the work you have done so far, with sections on the theory you have learnt and the code that you have written.

Also remember that any details of report content and submission rules, as well as other deliverables, are defined in the project booklet.

1.1 How to use this template

The simplest way to get started with your report is to save a copy of this document. First change the values for the initial document definitions such as `studentname` and `reportyear` to match your details. Delete the unneeded sections and start adding your own sections using the styles provided. Before submission, remember to fill in the Declaration section fields.

Chapter 2: **Page Layout & Size**

The page size and margins have been set in this document. These should not be changed or adjusted.

In addition, page headers and footers have been included. They will be automatically filled in, so do not attempt to change their contents.

Chapter 3: **Headings**

Your report will be structured as a collection of numbered sections at different levels of detail. For example, the heading to this section is a first-level heading and has been defined with a particular set of font and spacing characteristics. At the start of a new section, you need to select the appropriate LaTeX command, `\chapter` in this case.

3.1 Second Level Headings

Second level headings, like this one, are created by using the command `\section`.

3.1.1 Third Level Headings

The heading for this subsection is a third level heading, which is obtained by using command `\subsection`. In general, it is unlikely that fourth of fifth level headings will be required in your final report. Indeed it is more likely that if you do find yourself needing them, then your document structure is probably not ideal. So, try to stick to these three levels.

3.2 A Word on Numbering

You will notice that the main section headings in this document are all numbered in a hierarchical fashion. You don't have to worry about the numbering. It is all automatic as it has been built into the heading styles. Each time you create a new heading by selecting the appropriate style, the correct number will be assigned.

Chapter 4: **Presentation Issues**

4.1 Figures, Charts and Tables

Most final reports will contain a mixture of figures and charts along with the main body of text. The figure caption should appear directly after the figure as seen in Figure 4.1 whereas a table caption should appear directly above the table. Figures, charts and tables should always be centered horizontally.

Figure 4.1: Logo of RHUL.

4.2 Source Code

If you wish to print a short excerpt of your source code, ensure that you are using a fixed-width sans-serif font such as the Courier font. By using the `verbatim` environment your code will be properly indented and will appear as follows:

```
static public void main(String[] args) {
  try  {
    UIManager.setLookAndFeel(UIManager.getSystemLookAndFeelClassName());
  }
  catch(Exception e) {
    e.printStackTrace();
  }
  new WelcomeApp();
}
```

Chapter 5: **References**

Use one consistent system for citing works in the body of your report. Several such systems are in common use in textbooks and in conference and journal papers. Ensure that any works you cite are listed in the references section, and vice versa.

Chapter 6: **Project Information and Rules**

The details about how your project will be assessed, as well as the rules you must follow for this final project report, are detailed in the project booklet.

You must read that document and strictly follow it.

Bibliography

[1] L. R. Weinstock, "Introduction to u.s. economy: Housing market," *Congressional Research Service*, Jan. 2023.

[2] P. Herman, "The importance of price prediction," *Future Processing*, Mar. 2023.

[3] P. Schneider and F. Xhafa, "Chapter 3 - anomaly detection: Concepts and methods," in *Anomaly Detection and Complex Event Processing over IoT Data Streams* (P. Schneider and F. Xhafa, eds.), pp. 49–66, Academic Press, 2022.

[4] S. Olumide, "Root mean square error (rmse): What you need to know," *Arize*, Aug 2023.

[5] V. Roman, "Root mean square error (rmse): What you need to know," *Towards Data Science*, Jan. 2019.