# Lasso and Ridge Regression



**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name:RAKESH SETHU NP Batch ID: 11052022\_7.30PM**

**Topic: Lasso and Ridge Regression**



**Hints:**

1. **Business Problem**
   1. **What is the business objective?**
   2. **Are there any constraints?**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1 Make a table as shown above and provide information about the features such as its data type and its relevance to the model building. And if not relevant, provide reasons and a description of the feature.**

1. **Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

**3.2 Outlier Treatment.**

1. **Exploratory Data Analysis (EDA):**
   1. **Summary.**
   2. **Univariate analysis.**
   3. **Bivariate analysis.**
2. **Model Building**
   1. **Build the model on the scaled data (try multiple options).**
   2. **Perform Lasso and Ridge Regression.**
   3. **Train and test the model and compare RMSE values. Tabulate R-Squared and RMSE values for different models in the documentation and provide an explanation.**
   4. **Briefly explain the model output in the documentation.**
3. **Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**

**Problem Statements:**

1. Officeworks is a leading retail store in Australia, with numerous outlets around the country. The manager would like to improve the customer experience by providing them online predictive prices for their laptops if they want to sell them. To improve this experience the manager would like us to build a model which is sustainable and accurate enough. Apply Lasso and Ridge Regression model on the dataset and predict the price, given other attributes. Tabulate R squared, RMSE, and correlation values.

**Objective:** Maximize the accuracy in estimating relationship between 2 or more independent variable and one dependent variable.

**Constraints:** Limited to linear Relationships.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| X | Index Values | Nominal | Irrelevant, Doesn’t Provides useful information. |
| price | Price of gadget | Continuous, Ratio | Relevant, Provides useful information. |
| speed | Gadget speed | Continuous, Ratio | Relevant, Provides useful information. |
| hd | Scale of HD | Discrete, count | Relevant, Provides useful information. |
| ram | RAM of device | Discrete, count | Relevant, Provides useful information. |
| screen | Screen size | Discrete, count | Relevant, Provides useful information. |
| cd | CD supported | Categorical, Ordinal | Relevant, Provides useful information. |
| multi | Multimedia supported | Categorical, Ordinal | Relevant, Provides useful information. |
| premium | Premium device | Categorical, Ordinal | Relevant, Provides useful information. |
| ads | ads | Discrete, count | Relevant, Provides useful information. |
| trend | Trend | Discrete, count | Relevant, Provides useful information. |









1. An online car sales platform would like to improve its customer base and their experience by providing them an easy way to buy and sell cars. For this, they would like to have an automated model which can predict the price of the car once the user inputs the required factors. Help the business achieve the objective by applying Lasso and Ridge Regression on it. Please use the below columns for the analysis: Price, Age\_08\_04, KM, HP, cc, Doors, Gears, Quarterly\_Tax, Weight.

**Objective:** Maximize the accuracy in estimating relationship between 2 or more independent variable and one dependent variable

**Constraints:** Limited to linear Relationships

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| Price | Price of the car | Continuous, Ratio | It provides useful information |
| Age\_08\_04 | Age of a person | Continuous, Ratio | It is not useful |
| KM | Total Travelled Distance | Continuous, Ratio | It provides useful information |
| HP | Horse Power | Continuous, Ratio | It is a useful information |
| Cc | Cubic Capacity | Continuous, Ratio | Useful information |
| Doors | Number of doors in car | Discrete, Count | It is useful information |
| Gears | Number of gears | Discrete, Count | It provides useful information |
| Quarterly Tax | Insurance Tax for car | Continuous, Ratio | It provides useful information |
| Weight | Total Weight of car | Continuous, Ratio | It provides useful information |

 .

1. Data of various countries and the factors affecting their life expectancy has been recorded over the past few decades. An analytics firm would like to know how it varies country wise and what factors are influential. Use your skills to analyze the data and build a Lasso and Ridge Regression model and summarize the output. Snapshot of the dataset is given below.

**Objective:** Maximize the accuracy in estimating relationship between 2 or more independent variable and one dependent variable

**Constraints:** Limited to linear Relationships.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of feature** | **Description** | **Type** | **Relevance** |
| Country | Name of country | Discrete, Nominal | Not useful |
| Year | Year | Continuous, Ratio | It is useful |
| Status | Status of country | Discrete, Nominal | Not useful |
| Life\_expectancy | Average Life | Continuous, Ratio | It is useful |
| Adult\_Mortality | Average Adults | Continuous, Ratio | It is useful |
| infant\_deaths | Death Rate | Continuous, Ratio | It is useful |
| Alcohol | Alcohol percentage | Continuous, Ratio | It is useful |
| percentage\_expenditure | percentage expenditure | Continuous, Ratio | It is useful |
| Hepatitis\_B | Disease | Continuous, Ratio | It is useful |
| Measles | Disease | Continuous, Ratio | It is useful |
| BMI | Body mass index | Continuous, Ratio | It is useful |
| under\_five\_deaths | Death rate | Discrete, Count | It is useful |
| Polio | Disease | Continuous, Ratio | It is useful |
| Total\_expenditure | Total expense | Continuous, Ratio | It is useful |
| Diphtheria | Disease | Continuous, Ratio | It is useful |
| HIV\_AIDS | Disease | Continuous, Ratio | It is useful |
| GDP | Gross Democratic Product | Continuous, Ratio | It is useful |
| Population | Population Rate | Discrete, Count | It is useful |
| Thinness | State of body | Continuous, Ratio | It is useful |
| thinness\_yr | State of body | Continuous, Ratio | It is useful |
| Income\_composition | Income Rate | Continuous, Ratio | It is useful |
| Schooling | Education at school | Continuous, Ratio | It is useful |



A screenshot of a cell phone

Description automatically generated