

Objective:

In this session, you will learn multiple techniques to reduce dimensionality (Feature Selection) in your data and how to use those features in model building

Key takeaways:

- Implementing Lasso and Ridge Regression
- Generate principle components
 - Components Selection
 - Model building and comparing the results

For each topic, two problems are given. One for demonstration & practice and other as assignment to submit in grader tool:

RegularizationDemo and Practice Problem:

A large child education toy company which sells edutainment tablets and gaming systems both online and in retail stores in the US wanted to analyze the customer data. They are operating from last 15 years and maintaining all transactional information data. The given data 'CustomerData.csv' is a sample of customer level data extracted and processed for the analysis from various set of transactional files. Using this data, they want us to understand the life time value of each customer (LTV). This will enable them to design marketing strategies and customize the product offerings. The objective of activity is building a regression model to predict the customer revenue based on one factor that influences revenue the most.

Task:

Build Linear Regression, Ridge and Lasso Regression on this data and compare the results.

Assignment:

Data: *pbo_bin.csv*

Objective: To identify the factors that are impacting the sales incentivization for the customer executives in a chain of food restaurants.

Variables Description:

- crew_promote - Number of promotions per year for the crew
- mgmt_promote - Number of promotions per year for the management
- YTD_Crew_Annualized_Turnover - This is the % of crew members terminated each year, either for cause, or by their own resignation.
- Friendliness - How friendly the Tendys team are with the customers
- P2_Training_Store_Cnt - Numer of training programs offered to the employees
- CPR - Complaints, Problems and Revists score
- CEI_Score - Customer experience score
- Comp_Last_Year - Comparable last year Sales
- Comp_Last_Year_Trans - Comparable last year Transactions
- Audited_FC_Var_Str_Goal - Actual food cost versus theoretical food cost, or effectively food waste.
- Labor_Var - Actual labor hours used minus labor guide hours.
- Sos_DP2_Cars - No of cars served per min DP2(Lunch (10:30 a.m. - 2:00 p.m.))
- Sos_DP4_Cars - No of cars served per min DP4 (Dinner (5:00 p.m. - 8:00 p.m.)
- Sos_DP6_Cars - No of cars served per min DP6(Late Night (10:00 p.m. - 4:00 a.m.)
- dp2_trans_per_min - No of transactions per min DP2

- dp4_trans_per_min - No of transactions per min DP4(Dinner (5:00 p.m. - 8:00 p.m.))
- Freestyle_flag - The Way coke is delivered to the customer
- IA_Flag IA means - Image activation. This flag indicates whether a particular restaurant is renovated to improve the overall look and enhance customer experience
- Pbo - This variable obtained after deducting the profit after marketing expenses and other operating expenses at each restaurant level.

Task:

Build lasso and ridge regression model using the steps and code given in demonstration problem

Principle Component Analysis

Demo and Practice Problem:

Data and details: "housing_data.csv" & "Data Description.txt"

Task:

- a) Consider numeric variables only
- b) Split data into train and test
- c) Compute principle components on train
- d) Identify number of components
- e) Build regression model using all original attributes and compute error metrics on train and test
- f) Build regression model using principle components and compute error metrics on train and test

Assignment:

Data and details: "OnlineNewsPopularity.csv"

<https://archive.ics.uci.edu/ml/datasets/online+news+popularity>

Task:

Repeat all steps and code given in demonstration problem and submit the code with comments and observations