

PREDICTING PRE-OWNED CAR PRICES

ANJALI SEBASTIAN

INTRODUCTION

Motivation

Pricing Guide

- Get the best deals ?

Feature Dependent Depreciation

- Color, brand?

Objectives

Develop a model that can reasonably predict price of a used car

Goals

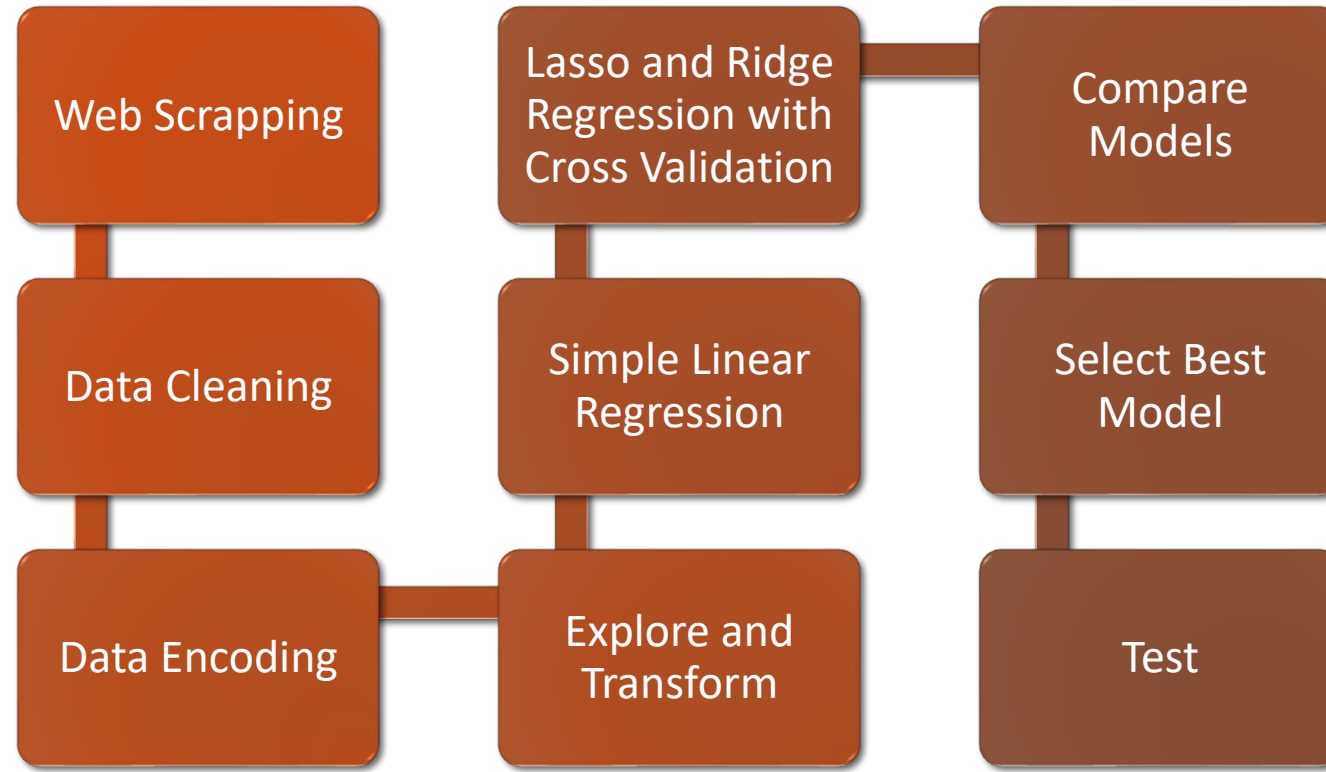
Collect sufficient data and clean it

Understand correlations

Develop a model using supervised linear regression

Test the model - better than a monkey





METHODOLOGY

DATA AND TOOLS

Web Scraping

- Scraped 1700+ pages from carmax.com
- 30+ parameters ... miles, horsepower, curb weight
- Cars around the Seattle and Portland

Cleaning Data

- Cleaning and Identifying Outliers

Encoding Data

- Exterior color - light, dark, prime
- Brands - luxury

Chrome Web
Driver

Selenium

Pandas Lib

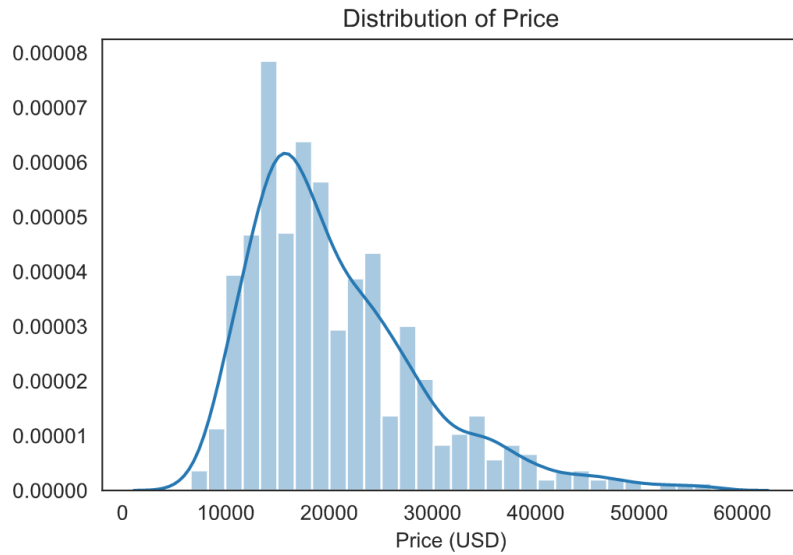
Statsmodel

Sklearn

Matplotlib

Seaborn

~~Beautiful Soup~~



FEATURE ENGINEERING

Log Transform Price

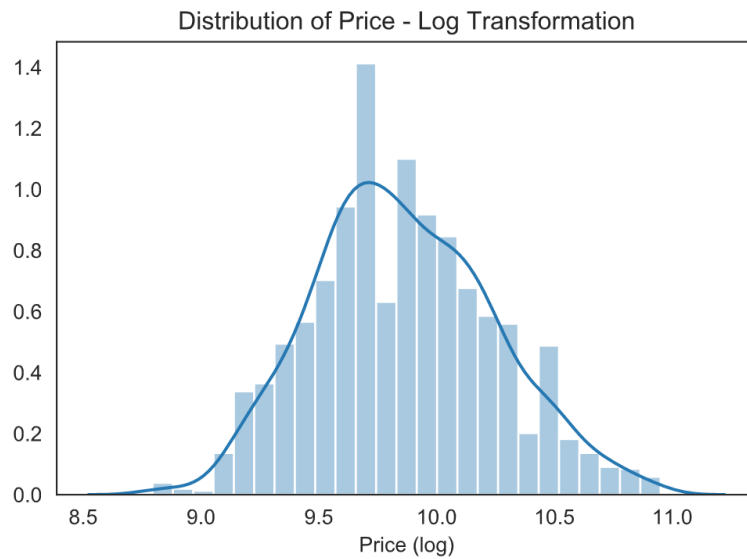
- Heteroskedastic

High correlation between features

- Engine Size vs #cylinders > 0.9

Linear Regression – use P-values to remove features

Transforming features vs. target

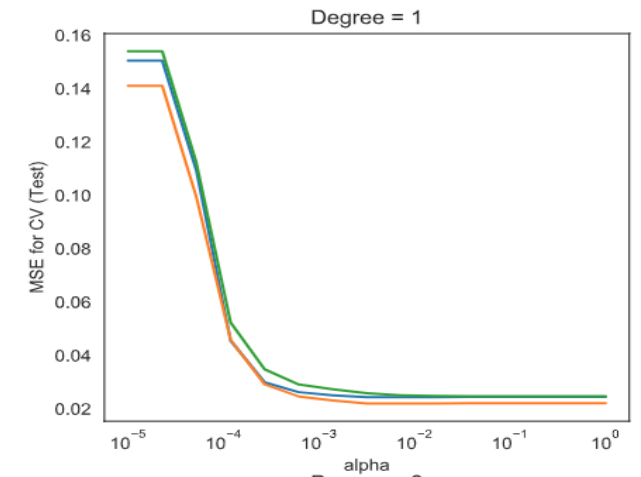


COMPARING MODELS

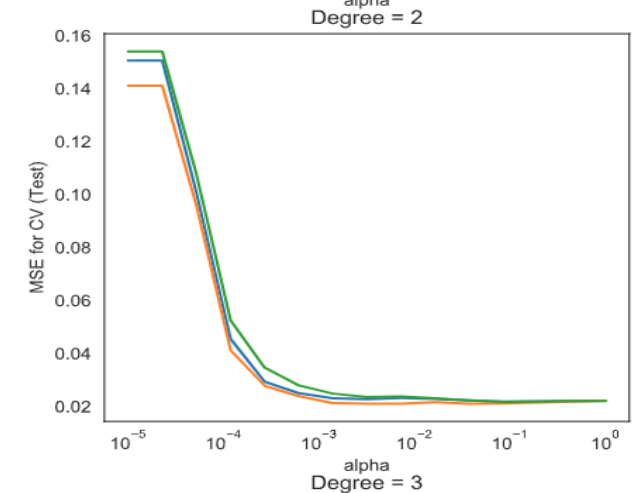
Type	Polynomial Degree	Alpha	MSE-Train	R ² - Train
Ridge	1	1.0	0.02267	0.84693
Ridge	2	1.0	0.01636	0.88949
Ridge	3	1.0	0.01196	0.91925
Lasso	1	0.00026	0.02271	0.84664
Lasso	2	0.00011	0.01657	0.88812
Lasso	3	0.00026	0.01459	0.90144

Use LassoCV and RidgeCV for running 3-fold Cross-validation
MSE_PATH_ to get MSE for the test
Plotted MSE Test for assessing fit
Why Ridge ?

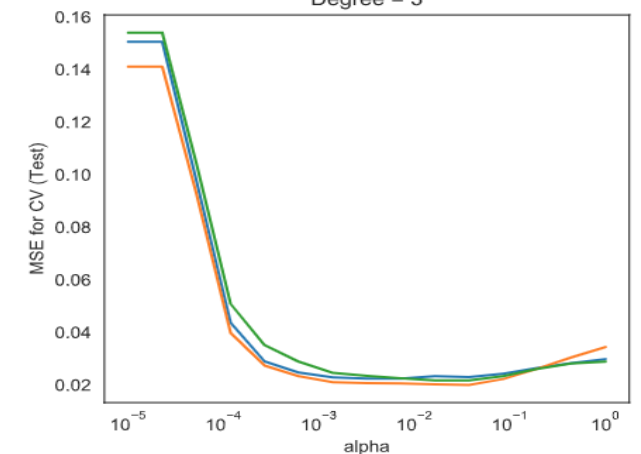
Underfitting

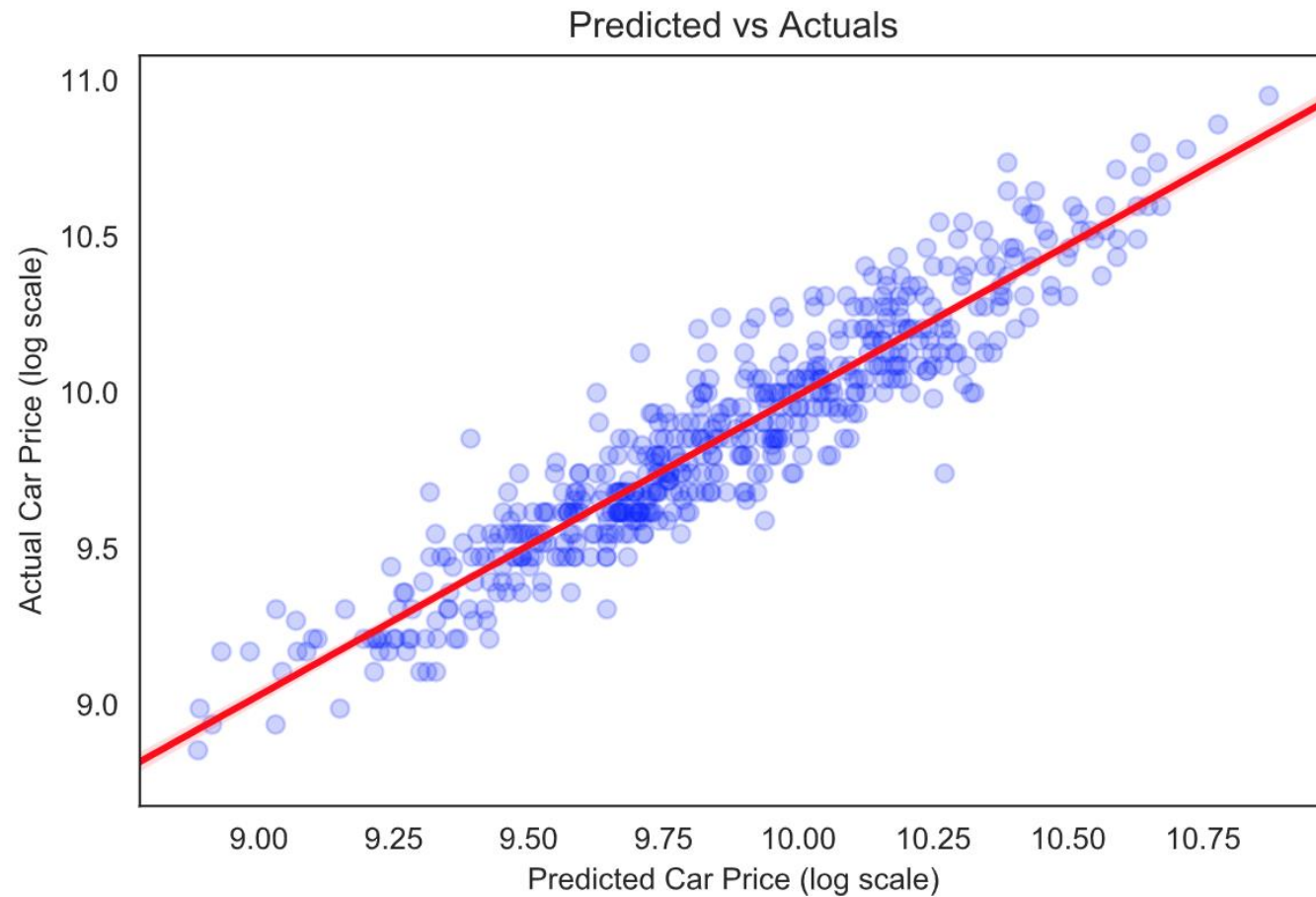


Good fit



Overfitting



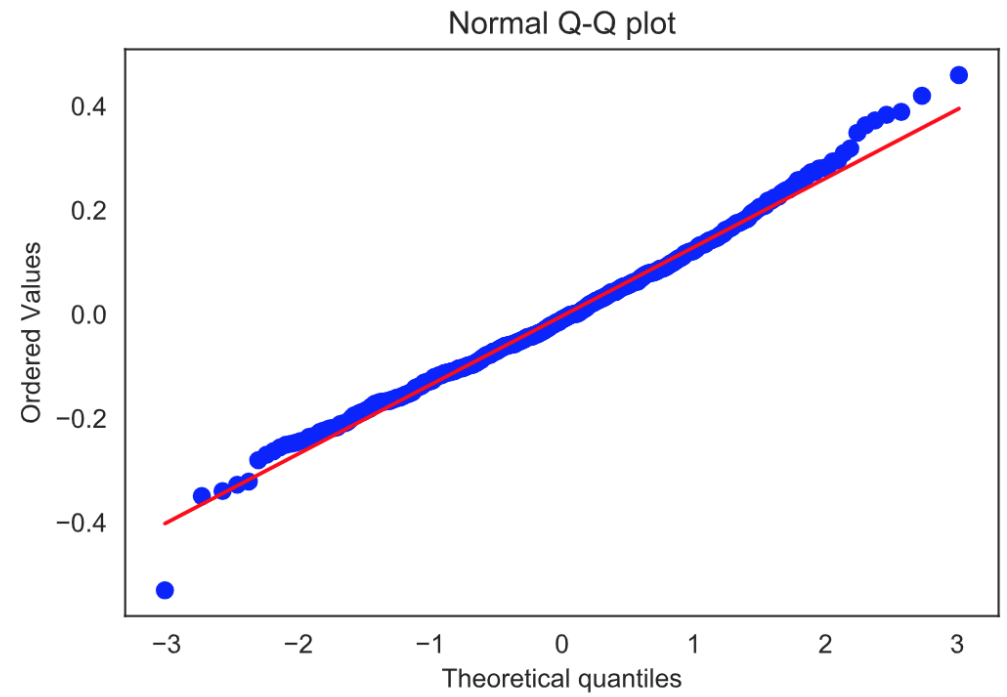
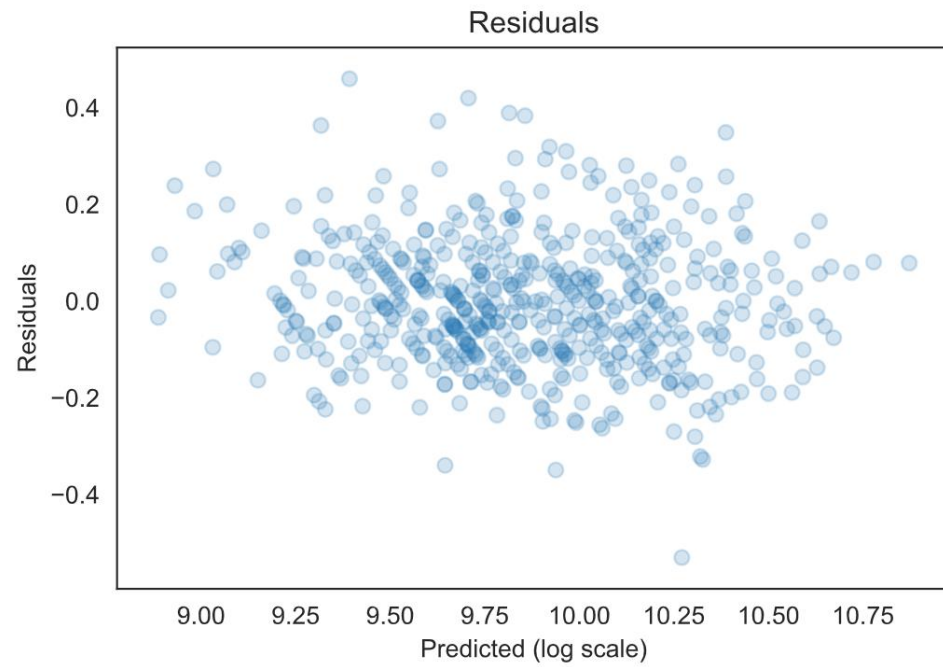


RESULTS

R^2 Test = 0.876

MSE Test = 0.017

$MSE = e^{\sqrt{0.017}} = 1.12 \sim 1.1\%$



RESULTS

CONCLUSION & FUTURE WORK

Are we better than a monkey ?

- Definitely

Can we use this model as a price guide ?

- Case extreme outlier, residual - GMC Yukon SLT, 2017

Can we identify feature dependent depreciation?

- Future Work

Original Price as a feature

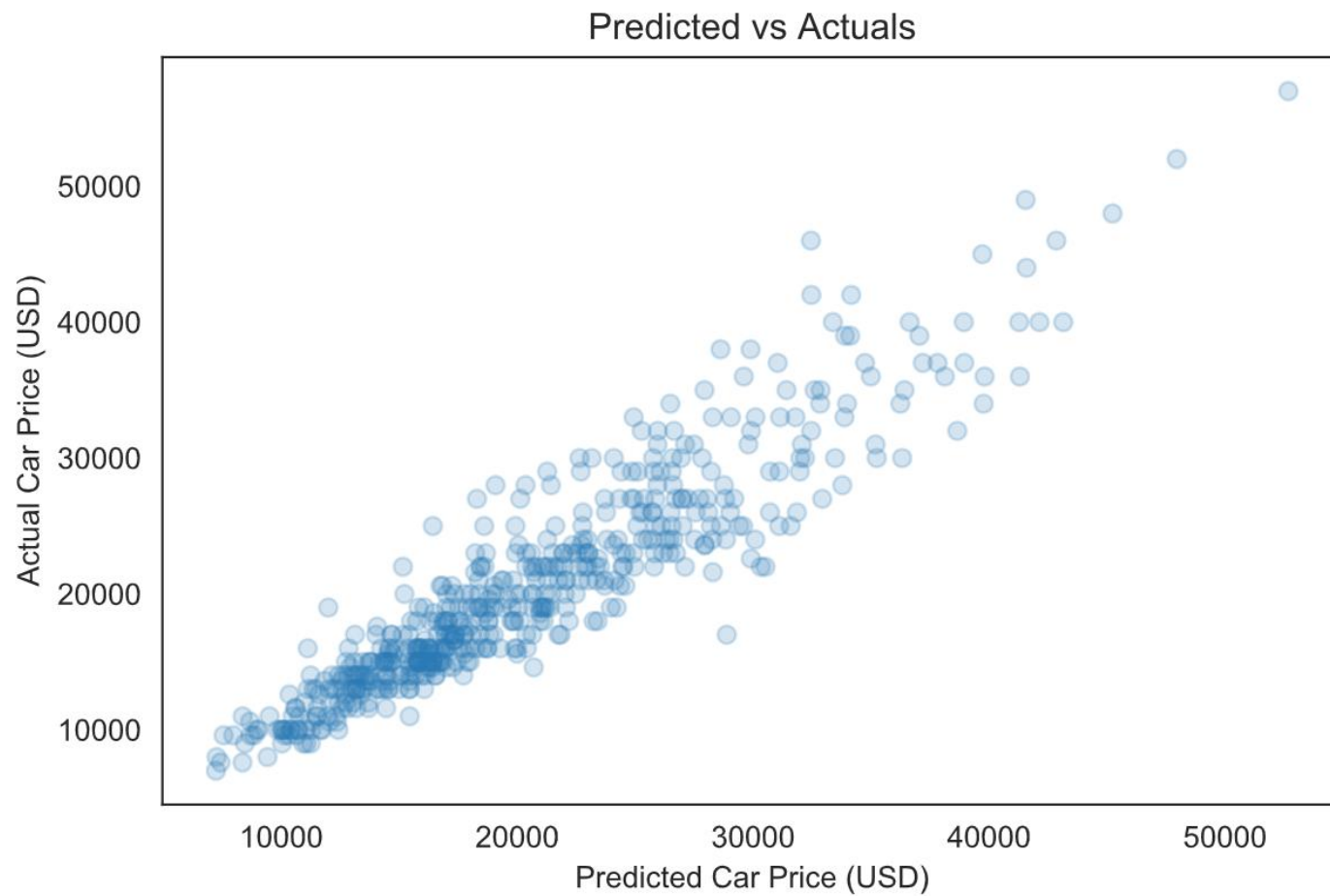
More data points

Have models specific for a brand, year

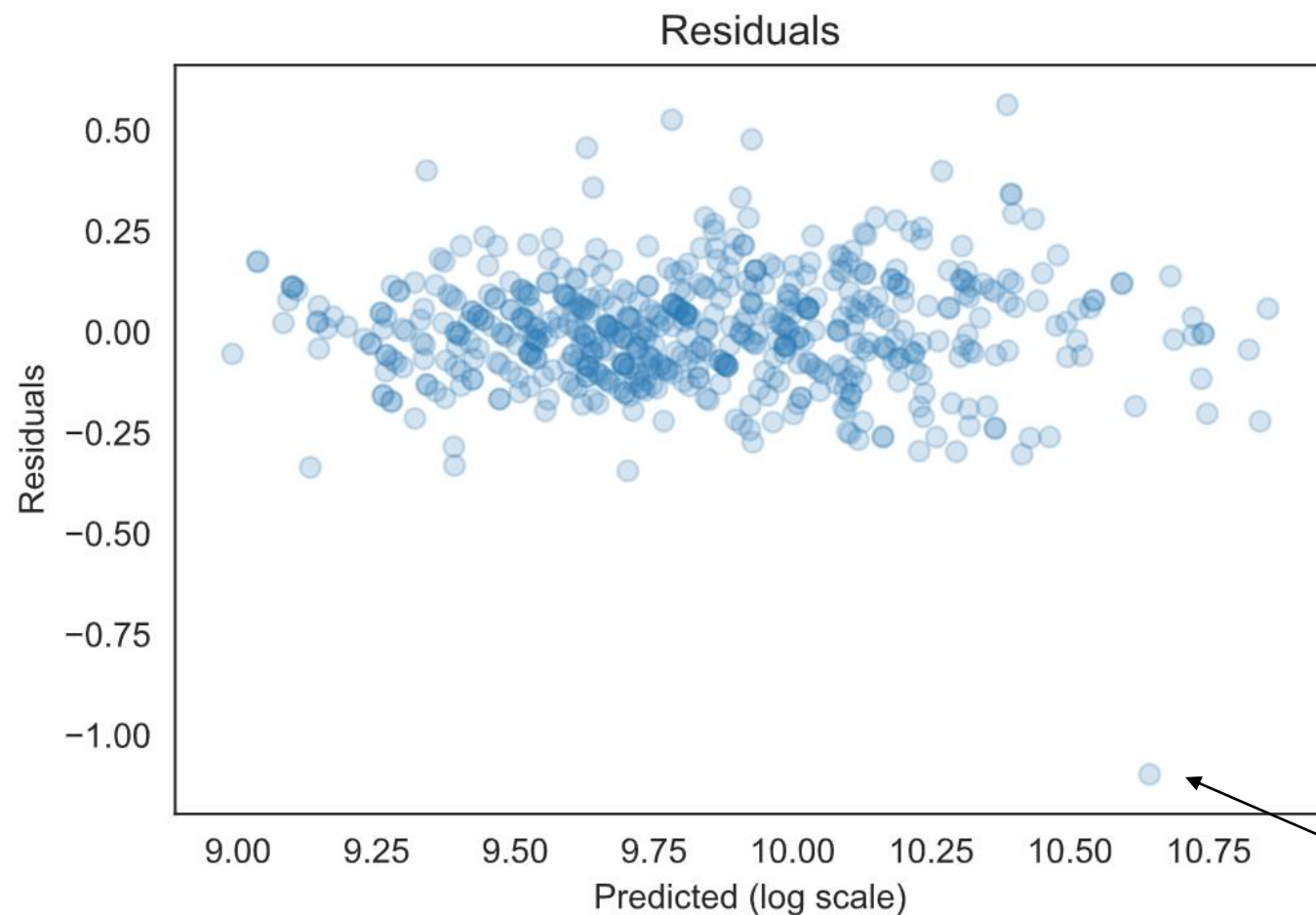
Unsupervised learning to cluster brands and specific models

Predict car depreciation of a new car

APPENDIX

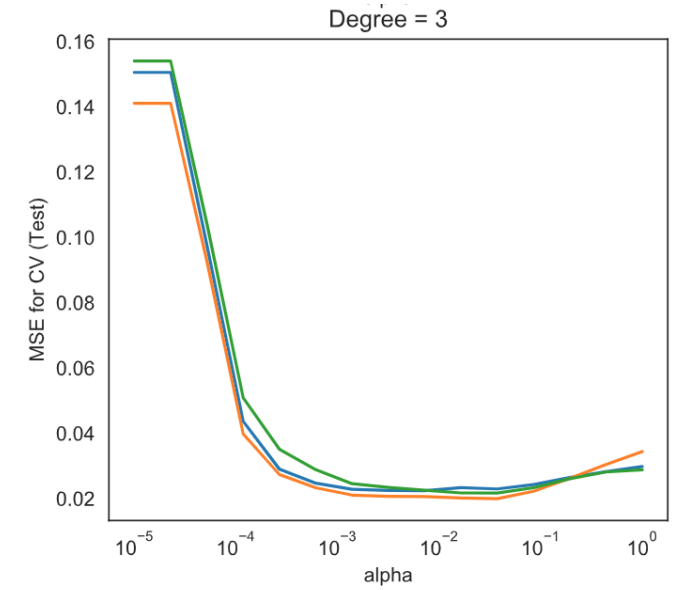
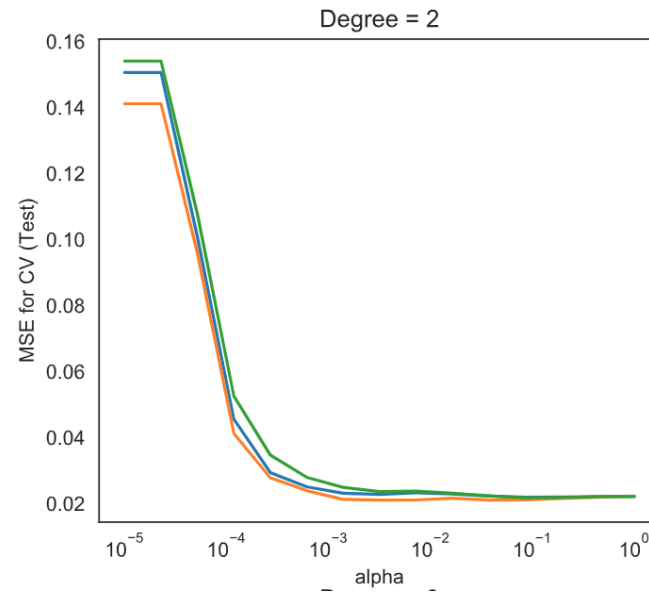
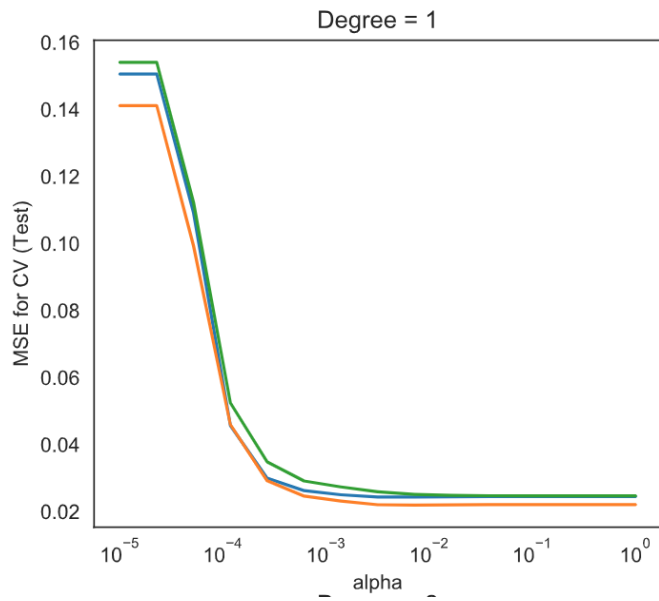


PREDICTED VS.
ACTUALS (RE-
TRANSFORMED)



RESIDUALS WITH OUTLIER

GMC Yukon SLT, 2017



OVER FITTING AND UNDERFITTING