Real Estate Price Prediction with Elastic-net Regression

Input Data file: data/ason1.csv

## Basic summary statistics

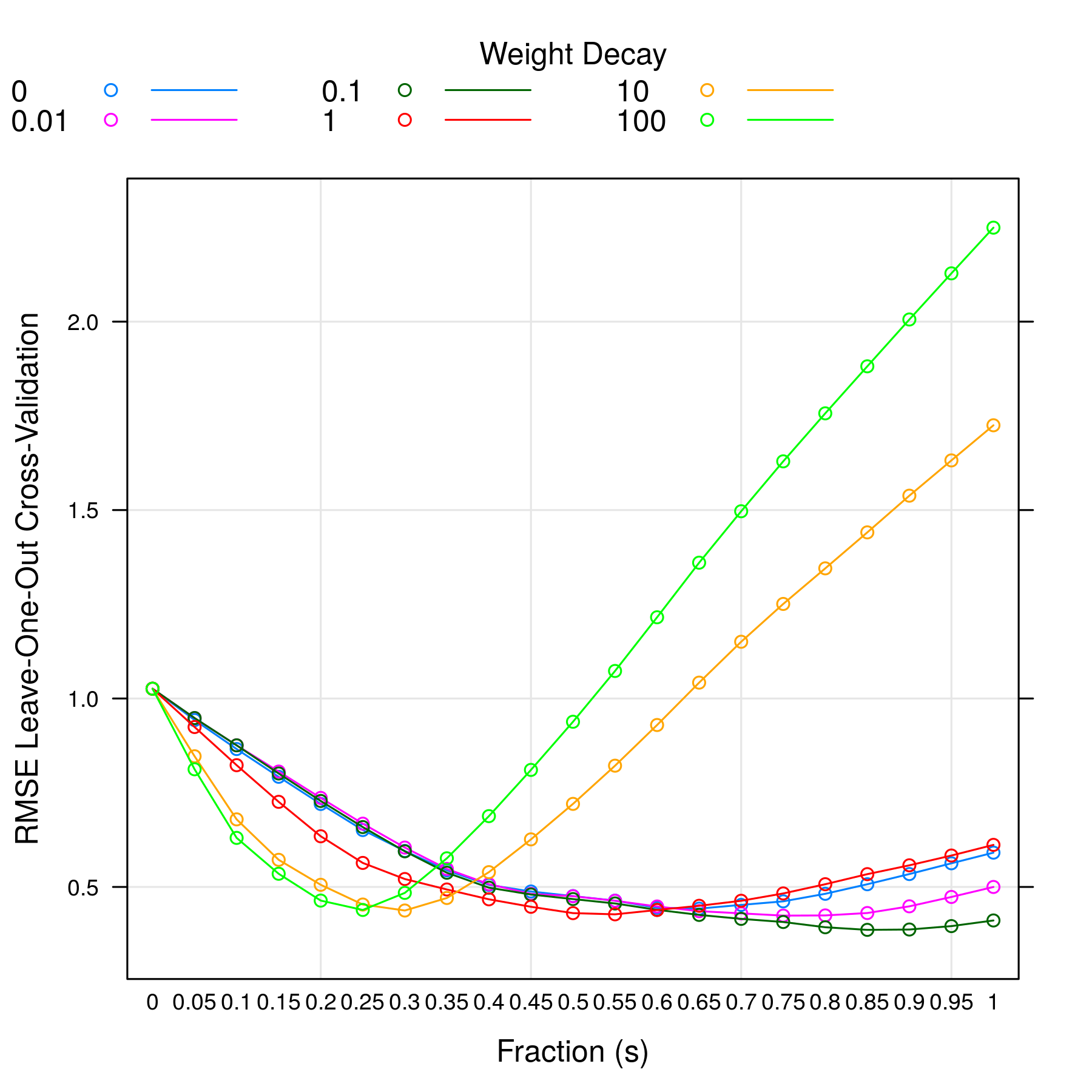
|  | **N** | **Mean** | **Std Dev** | **Min** | **Max** |
| --- | --- | --- | --- | --- | --- |
| SellingPrice | 020 | 236307 | 125759 | 122750 | 450000 |
| SellingDate | 020 | 40473 | 1122 | 39006 | 42308 |
| SiteAreaSF | 020 | 18618 | 12944 | 6120 | 50094 |
| GrossBldgAreaSF | 020 | 4472 | 2045 | 2000 | 8400 |
| OfficeSF | 020 | 939 | 598 | 200 | 2400 |
| MezzanineYes1 | 020 | 000 | 000 | 000 | 001 |
| YearBuilt | 020 | 1993 | 018 | 1948 | 2010 |
| OverheadDoors | 020 | 002 | 001 | 001 | 006 |
| LandToBldgArea | 020 | 003 | 001 | 001 | 007 |
| PavedParkingSF | 020 | 5453 | 4405 | 1800 | 17088 |
| FenceYes1 | 020 | 000 | 000 | 000 | 001 |

NOTE - No summary statistics are provided for categorical variables.

## Correlations Between Predictors

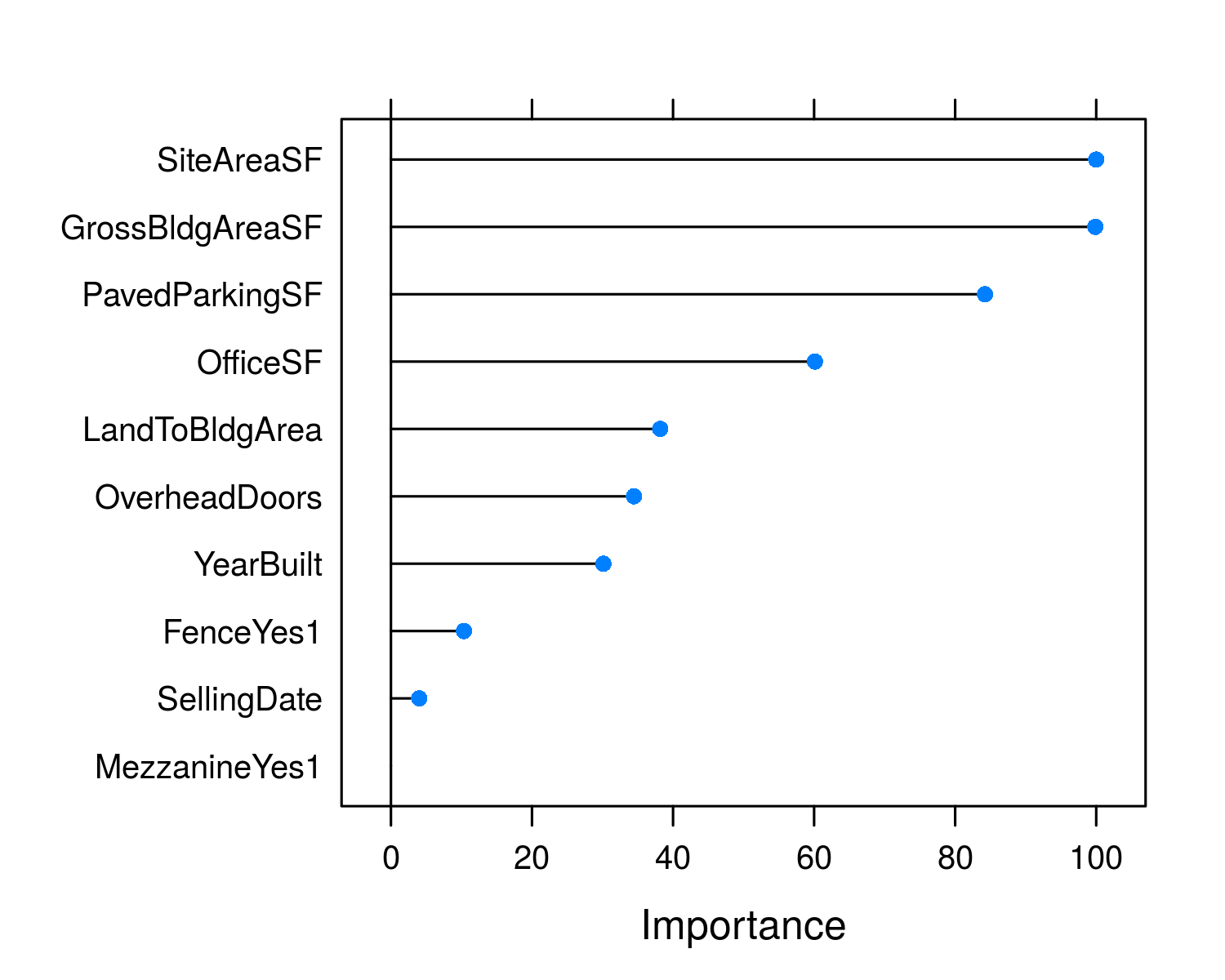


## Tuning Parameter Selection Using LOOCV



From above plot, lambda =0.10 and s =0.85 gives the minimum RMSE model.

## Variable Importance



## Standardized Model Coefficients

| **Variable** | **Estimate** |
| --- | --- |
| SellingDate | 0.185 |
| SiteAreaSF | 0.362 |
| GrossBldgAreaSF | 0.338 |
| OfficeSF | 0.165 |
| YearBuilt | 0.177 |
| OverheadDoors | 0.000 |
| LandToBldgArea | 0.085 |
| PavedParkingSF | 0.110 |
| MezzanineYes1 | 0.000 |
| FenceYes1 | -0.143 |

NOTE std. errors are calculated using bootstrapping which is the only way to determine coef. errors for a penalized regression. But the errors should be only used for reference. It is yet unclear how meaningful the std. errors are in penalized regression.

## Model Prediction

| **Predicted Value** | **Prediction Error** | **R2** |
| --- | --- | --- |
| 416816.820 | 0.386 | 0.845 |