REGULARIZATION

Ridge and Lasso

WHAT WILL YOU LEARN/REVIEW

- The **basic idea behind regularization**
- The difference between the **penalty terms for Lasso and Ridge** regression models
- How the **target function for Lasso** regularized regression models differs from the MSE function of an unregularized model
- How to create a **workflow for a Lasso** regularized regression using the *R* `tidymodels` framework
- How the **target function for Ridge** regularized regression model differs from the MSE function of an unregularized model
- How to create a **workflow for a Ridge** regularized model using the *R* `tidymodels` framework

LOADING THE LIBRARIES, DATA, AND SPLITTING IN TRAINING/TESTING DATA:

- Code

```
      Price Sqft
1    153503 1240
2    199500 1750
3    234950 1720
4    246000 2120
5    355000 1240
6    385000 2090
7    365000  910
8    349000 1690
9    474950 2030
10   450000 1540
11   465000 2020
12   445000 1630
13   568000 2110
14   660000 2470
15   530000 1260
16   600000 2090
17  1150000 3830
```

THE MODEL

$$\widehat{Price}_i = \beta_1 Sqft_i + \beta_2 Sqft_i^2 + \beta_3 Sqft_i^3 + \beta_4 Sqft_i^4 + \beta_5 Sqft_i^5 + \beta_0$$

UNREGULARIZED MODEL MINIMIZES THE MSE BY CHOOSING THE OPTIMAL βs

$$MSE = \frac{1}{20}\sum_{i=1}^{20}\left(\widehat{Price}_i - Price_i\right)^2$$

with:

$$\widehat{Price}_i = \beta_1 Sqft_i + \beta_2 Sqft_i^2 + \beta_3 Sqft_i^3 + \beta_4 Sqft_i^4 + \beta_5 Sqft_i^5 + \beta_0$$

RUNNING THE UNREGULARIZED MODEL

- Code

```
# A tibble: 6 × 5
  term           estimate  std.error statistic       p.value
  <chr>             <dbl>      <dbl>     <dbl>         <dbl>
1 (Intercept)     509945.     36463.    14.0   0.00000000128
2 Sqft           8853783.  10515448.     0.842 0.414
3 Sqft2        -50947114.  54352075.    -0.937 0.364
4 Sqft3        112589222. 111217647.     1.01  0.329
5 Sqft4       -106894260. 101985738.    -1.05  0.312
6 Sqft5         36592435.  34688741.     1.05  0.309
```

ASSESSING PREDICTION QUALITY (TRAINING DATA)

- Code

```
# A tibble: 3 × 3
  .metric .estimator  .estimate
  <chr>   <chr>           <dbl>
1 rmse    standard   136432.
2 rsq     standard        0.715
3 mae     standard   104047.
```

ASSESSING PREDICTION QUALITY (TRAINING DATA)

- Code

```
# A tibble: 3 × 3
  .metric .estimator     .estimate
  <chr>   <chr>              <dbl>
1 rmse    standard   99940240.
2 rsq     standard          0.0215
3 mae     standard    1719470.
```

REGULARIZATION

RIDGE

$$Target = \frac{1}{20}\sum_{i=1}^{20}\left(\widehat{Price}_i - Price_i\right)^2 + \lambda Penalty$$

$$\text{with:} \quad \widehat{Price}_i = \beta_1 Sqft_i + \beta_2 Sqft_i^2 + \beta_3 Sqft_i^3 + \beta_4 Sqft_i^4 + \beta_5 Sqft_i^5 + \beta_0$$

$$\text{with:} \quad Penalty = \sum_{j=1}^{5}\beta_j^2$$

Two Goals: Minimize MSE and Minimize Penalty (small or zero βs)

$Target$ value still only depends on data.

Note, reducing a large or a small β parameter by the same amount has the same impact on the *penalty*.

RUNNING THE RIDGE MODEL

- Code

```
# A tibble: 6 × 3
  term        estimate penalty
  <chr>          <dbl>   <dbl>
1 (Intercept)  509945. 1000000
2 Sqft          25790. 1000000
3 Sqft2         23133. 1000000
4 Sqft3         19885. 1000000
5 Sqft4         16968. 1000000
6 Sqft5         14570. 1000000
```

ASSESSING PREDICTION QUALITY RIDGE MODEL (TRAINING DATA)

- Code

```
# A tibble: 3 × 3
  .metric .estimator  .estimate
  <chr>   <chr>           <dbl>
1 rmse    standard   201534.
2 rsq     standard        0.479
3 mae     standard   152902.
```

ASSESSING PREDICTION QUALITY RIDGE MODEL (TESTINGG DATA)

- Code

```
# A tibble: 3 × 3
  .metric .estimator  .estimate
  <chr>   <chr>           <dbl>
1 rmse    standard   330485.
2 rsq     standard        0.237
3 mae     standard   186431.
```

REGULARIZATION

LASSO

$$T^{arget} = \frac{1}{20}\sum_{i=1}^{20}\left(\widehat{Price}_i - Price_i\right)^2 + \lambda P^{enalty}$$

$$\text{with:} \quad \widehat{Price}_i = \beta_1 Sqft_i + \beta_2 Sqft_i^2 + \beta_3 Sqft_i^3 + \beta_4 Sqft_i^4 + \beta_5 Sqft_i^5 + \beta_0$$

$$\text{with:} \quad P^{enalty} = \sum_{j=1}^{5}|\beta_j|$$

Two Goals: Minimize MSE and Minimize Penalty (small or zero βs)

T^{arget} value still only depends on data.

Note, reducing a large or a small β parameter by the same amount has the same impact on the *penalty*.

RUNNING THE LASSO MODEL

- Code

```
# A tibble: 6 × 3
  term        estimate penalty
  <chr>          <dbl>   <dbl>
1 (Intercept)  509945.     500
2 Sqft        -460508.     500
3 Sqft2       1171967.     500
4 Sqft3             0      500
5 Sqft4             0      500
6 Sqft5       -560318.     500
```

ASSESSING PREDICTION QUALITY LASSO MODEL (TRAINING DATA)

- Code

```
# A tibble: 3 × 3
  .metric .estimator  .estimate
  <chr>   <chr>           <dbl>
1 rmse    standard   144976.
2 rsq     standard        0.679
3 mae     standard   110007.
```

ASSESSING PREDICTION QUALITY LASSO MODEL (TESTING DATA)

- Code

```
# A tibble: 3 × 3
  .metric .estimator    .estimate
  <chr>   <chr>             <dbl>
1 rmse    standard   4723086.
2 rsq     standard         0.0296
3 mae     standard    303118.
```