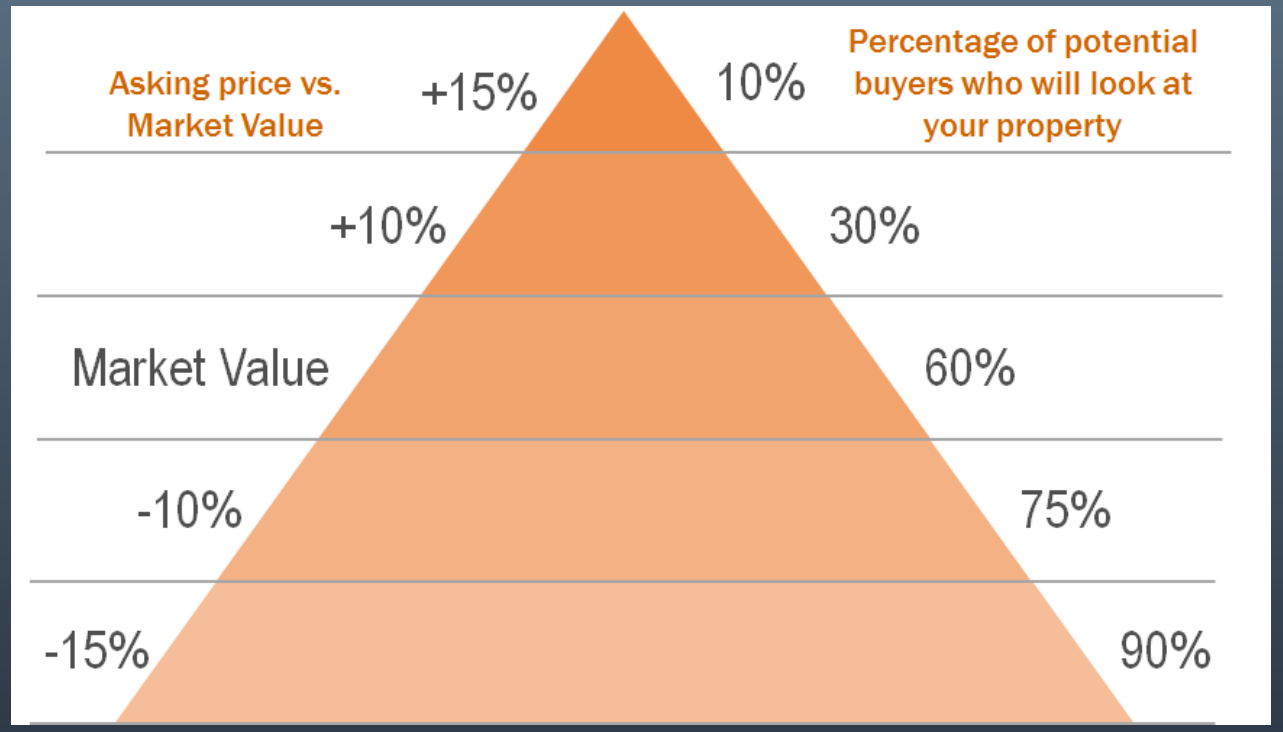
Machine Learning Approach to Estimating Housing Market Value

by Charles Ramey

Background

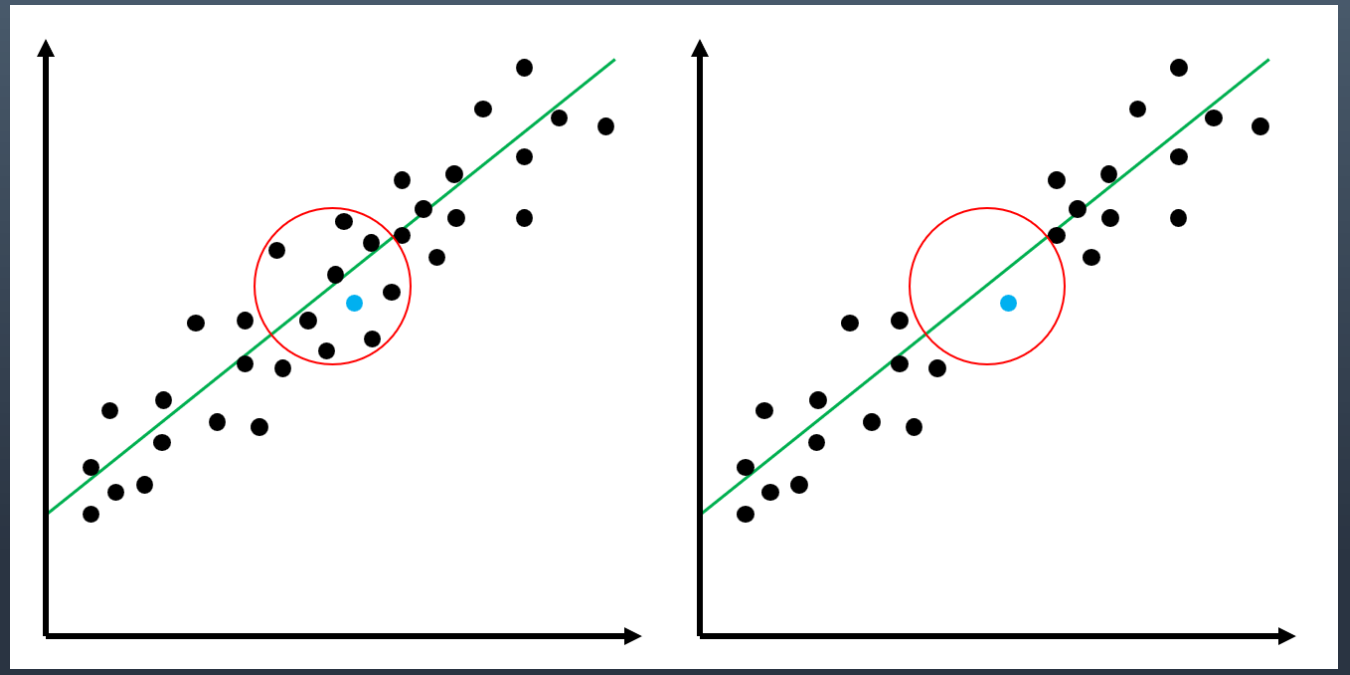
- Choosing the initial asking price is tough
- Various strategies for selling a home:
  - 10-15% above market value
  - 10-15% below market value
- Market Value?
  - Comparative Market Analysis (CMA)
  - Comparable Sales



https://palermolistings.com/market-data/the-great-pyramid-of-real-estate

What's the benefit?

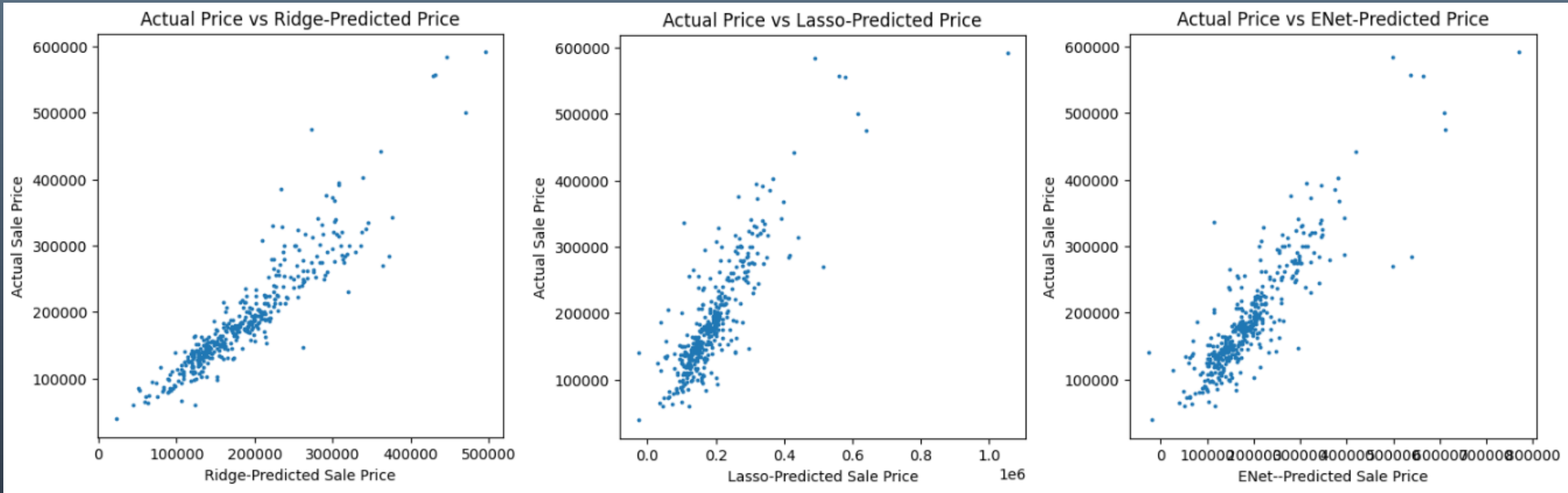
- Limited time and resources
- What if there are no comparable sales?



Problem Statement

Can we generate a machine-learning model to help real estate agents accurately estimate a fair market value for a home?

Iteration 1


Actual Price vs Ridge-Predicted Price
Actual Sale Price
Ridge-Predicted Sale Price
600000
500000
400000
300000
200000
100000
0
100000
200000
300000
400000
500000

Actual Price vs Lasso-Predicted Price
Actual Sale Price
Lasso-Predicted Sale Price
1e6
600000
500000
400000
300000
200000
100000
0.0
0.2
0.4
0.6
0.8
1.0

Actual Price vs ENet-Predicted Price
Actual Sale Price
ENet--Predicted Sale Price
600000
500000
400000
300000
200000
100000
0

Iteration 2


Actual Price vs Ridge-Predicted Price
Actual Sale Price
600000
500000
400000
300000
200000
100000
100000
200000
300000
400000
Ridge-Predicted Sale Price
Actual Price vs Lasso-Predicted Price
Actual Sale Price
600000
500000
400000
300000
200000
100000
0
100000
200000
300000
400000
500000
Lasso-Predicted Sale Price
Actual Price vs ENet-Predicted Price
Actual Sale Price
600000
500000
400000
300000
200000
100000
0
100000
200000
300000
400000
500000
ENet--Predicted Sale Price

The numbers...

Iteration 1

Model	Training Error ($)	Testing Error ($)
Ridge	0.92	0.84
Lasso	0.98 →	0.61
ElasticNet	0.98 →	0.72

Iteration 2

Model	Training Error ($)	Testing Error ($)
Ridge	0.92	0.84
Lasso	0.95 →	0.79
ElasticNet	0.95 →	0.80

How far off are we, really?

Using the Ridge model...

$20,000 - $35,000

With an average home sale value of about $180,000
that's an error between 11 – 19%

Conclusions

- The Ridge model is promising, however…
- The error is too large

Recommendations

- Deeper data cleaning
- Use a different k-value for cross-validation
- Collecting more data to train the model on
- Including sale date